

文章编号: 2096-1618(2024)05-0540-06

# 基于 X 波段雷达组网的强对流识别

蓝敬珂<sup>1,2</sup>, 杨玲<sup>1,2</sup>, 师春香<sup>3</sup>, 徐梓欣<sup>1,2</sup>, 练金鑫<sup>1,2</sup>, 余杰<sup>1,2</sup>

(1. 成都信息工程大学电子工程学院, 四川 成都 610225; 2. 中国气象局大气探测重点开放实验室, 四川 成都 610072; 3. 国家气象信息中心气象数据研究室, 北京, 10081)

**摘要:** 冰雹、暴雨等强对流天气对人类活动的危害巨大, 但由于其发展迅速、结构复杂, 对其进行自动识别非常困难。基于成都航空港、龙泉、资阳地区 3 部 X 波段雷达构成的雷达组网平台, 利用雷达组网的高时空分辨率的体扫数据, 在传统的 U-Net 网络基础上进行改进, 提出一种基于 U-Net 语义分割网络的强对流天气识别模型。选取雷达组网中 0.5°、2°、3.5°、5°、6.5°、8° 和 9° 共 7 层仰角基数据中的径向速度和反射率数据, 进行配准、归一化等预处理后形成模型训练所需的数据集。经实验验证, 训练后得到的模型对强对流识别的结果比较好。

**关键词:** U-Net; 强对流识别; 语义分割

**中图分类号:** TN959.4

**文献标志码:** A

**doi:** 10.16836/j.cnki.jcuit.2024.05.004

## 0 引言

强对流天气是一种常发于夏季, 具有强破坏性的天气现象, 其主要表现为冰雹、暴雨和飑线等, 易对生活、生产造成损害<sup>[1]</sup>, 因此强对流天气的预报与预警显得尤为重要。

天气雷达是现代大气探测领域中的常用设备, 由于地球自身曲率限制及地物遮挡和满足净空条件等因素, 通常 S 波段雷达建设的位置都会处于一定的海拔高度上, 这使其本身存在的低空探测盲区及低时空分辨率的劣势更加明显, 并且其他的天气雷达如 C、X 波段的雷达也存在相似的问题, 而这些不足可以在美国大气协同自适应遥感 (CASA) 中心提出的雷达组网基础上得到改善<sup>[2-3]</sup>。此方法关键在于使用多部小体积、低成本和短探测距离的 X 波段天气雷达构建雷达组网平台, 用以弥补单一的 S、X、C 波段雷达在探测时存在的不足。

从 1950 年代起就开始通过气象雷达对对流单体进行研究<sup>[4-6]</sup>。国外科学家通过分析提出一系列经典方法, 改变了使用传统气象雷达资料分析对流单体的局面, 取得一定的成果。Johnson 等<sup>[7]</sup>提出 SCIT (storm cell identification and tracking) 算法, 该方法设定 7 个不同的反射率阈值 (30、35、40、45、50、55、60 dBz), 从而获得不同仰角下符合条件的反射率区域。当各仰角的反射率区域出现重叠时, 只保留最高阈值区域, 然后通过垂直相关法将不同仰角上的区域合成三维单体。虽然 SCIT 算法能正确跟踪大部分对流单体, 但只考虑

了反射率因子强度和空间连续性, 对反射率因子强度较弱对流活跃的单体识别率较低<sup>[8-9]</sup>。因此, 通过单一的反射率因子进行对流单体研究已经不足, 王萍等<sup>[10]</sup>研究发现径向速度等雷达资料能展现对流云团的结构、形态以及对流场分布, 从而促进人们对冰雹、暴雨等强对流天气的特点及演变规律的认识。

虽然通过传统方式能对对流单体进行分析, 但其不仅需要一定的气象背景知识, 并且每次分析过程需要花费较多时间, 因而存在一定的门槛和局限性。随着机器学习发展, 强对流的识别可通过神经网络来实现<sup>[11-14]</sup>。通过利用不同的网络学习不同时刻天气过程的雷达回波图或卫星云图特征, 实现对之后时刻的预测。如 2015 年, Mecikalski 等<sup>[15]</sup>利用逻辑回归和随机森林等方法提出基于数值模式和卫星资料的对流初生临近预报技术。2019 年, 韩丰等<sup>[16]</sup>及郭瀚阳等<sup>[17]</sup>分别用深度学习、循环神经网络显著地提升了雷达回波外推的预报准确率。2021 年, 刘新伟等<sup>[18]</sup>基于机器学习中的 LightGBM (light gradient boosting machine) 算法, 对位于甘肃三地的地面观测数据及 C 波段雷达回波产品进行研究, 构建 LightGBM 模型, 用于识别冰雹、雷暴大风、短时强降水 3 类强对流天气。

综上, 本研究基于成都航空港、成都龙泉和资阳 3 部 X 波段天气雷达构成的组网平台, 选取雷达资料中 0.5°、2°、3.5°、5°、6.5°、8° 和 9° 共 7 层仰角的径向速度和反射率数据进行处理后作为输入模型所需的数据集, 并将改进的 U-Net 网络作为本研究所用深度学习模型, 利用预处理好的数据进行训练, 实现对强对流天气的识别。

## 1 数据集处理

### 1.1 数据资料

选取的雷达组网平台由位于龙泉、航空港和资阳的3部X波段雷达构成,3部雷达均为西安华腾微波公司生产的HTXY-A型多普勒天气雷达,具有重量轻、价格低、机动性高和可靠性高的优点。在进行扫描时主要进行协同模式的扫描,其扫描覆盖的公共部分为 $103.8^{\circ}\text{E} \sim 104.8^{\circ}\text{E}$ ,  $29.8^{\circ}\text{N} \sim 31^{\circ}\text{N}$ 。其中协同模式由RHI扫描和快速体扫组成。为克服传统单一雷达探测时的低时空分辨率缺点,同时考虑到3部雷达构成的雷达组网覆盖范围存在的地形遮挡等因素,在快速体扫时仅进行 $0.5^{\circ}$ 、 $2^{\circ}$ 、 $3.5^{\circ}$ 、 $5^{\circ}$ 、 $6.5^{\circ}$ 、 $8^{\circ}$ 和 $9^{\circ}$ 这7层仰角的 $360^{\circ}$ 扫描,使扫描时间从传统的6 min降至2 min。

模型训练所选取数据为2021年7月含有强对流和较大面积层状云天气时间段的雷达资料数据,包含龙泉、航空港和资阳的3部X波段雷达的7层仰角基数据。

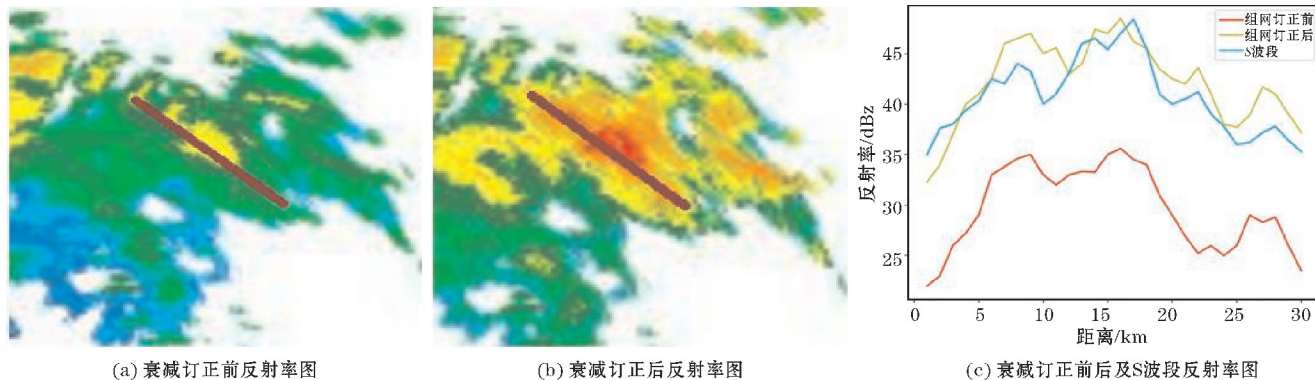


图1 衰减订正对比图

3部雷达的径向速度数据则是作为模型识别强对流区域的辅助信息参与模型训练过程,因此其处理方式直接为直接将每部雷达的7层径向速度数据在维度上进行堆叠,即21层径向速度数据,为方便后续的网络训练,将径向速度和组合反射率也在维度上进行一次叠加,得到的一组网络输入数据,即尺寸 $601 \times 601 \times 28$ 的网格数据。考虑到网络训练所需时间以及网格数据有效值所在位置等因素,将每一层数据的尺寸等比例采样为 $512 \times 512$ 的尺寸,得到 $512 \times 512 \times 28$ 的网格数据并将其保存为npz文件作为最终训练网络的输入数据。

### 1.3 组合反射率及标签处理

因本研究使用U-Net模型对强对流天气进行识

### 1.2 网络输入数据处理

雷达基数据中包含仰角、方位角、雷达反射率因子、径向速度、雷达经纬度等信息,而本研究所需数据为反射率和径向速度数据,需对基数据进行解析并提取出所需部分。因雷达组网数据需对3部雷达处于相邻时间段的数据进行组合拼图,所以需将3部雷达的数据进行时间匹配,考虑到局部强对流天气生消时间短以及3部雷达扫描时间等因素,将时间匹配阈值设置为2 min,并在匹配时以航空港雷达采集数据时间为基准时间。

考虑到X波段雷达反射率因子在降水区域存在较大衰减,因而需对反射率数据进行衰减订正。由于本研究重点为强对流识别算法,因而在此仅简单介绍衰减订正步骤,详细步骤可见文献[19]。其简要步骤如下:与时间匹配时相同,选取航空港区域的雷达作为基准雷达,在共同观测区域选取各部雷达反射率最大值为初始值,采用李兆明等<sup>[19]</sup>网络化雷达反射率因子衰减订正算法对初始值进行订正,由此得到订正后的组合反射率。如图1所示,经过衰减订正后,组合反射率值已接近同区域S波段雷达反射率值,因此订正后所得组合反射率数据质量有一定保证。

别,所以在训练前需要对组合反射率图进行标记,组合反射率图个例如图2所示,包含经纬度和色标等信息。

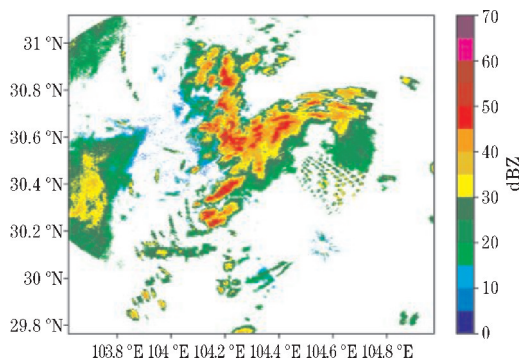


图2 组合反射率个例图

在制作数据集时主要进行两类天气现象的标记,即强对流和层状云天气的标记,标记时主要依据雷达获取的径向速度、反射率、云高及水平、垂直梯度等特征判断在组合反射率图中哪些区域存在强对流或层状云天气现象并完成对应标记。标记时使用 labelme 软件对组合反射率图进行处理,通过 labelme 软件对图片进行标记之后,软件会生成与标记图片同名的 json 文件,通过代码可将 json 文件转为 png 的标签图,即输入所对应的输出标签图,对应标签图如图 3 所示,标签图中红色部分是强对流区域,绿色部分是层状云区域,黑色部分为背景,最终经过标记得到 488 张标签图。



图3 标签结果图

将对组合反射率标记生成的标签图与预处理阶段生成的 npy 网格文件看作一组输入输出数据对,由此得到 488 个数据对的数据集,将整个数据集按 8:2 划分为训练集和验证集,其构成如表 1 所示,用于进行后续的训练和验证等过程。

表1 数据集组成表

类别	类别个例数目
强对流	141
层状云	347

## 2 模型及评价指标选取

### 2.1 模型简要介绍

U-Net 是 2015 年被提出的一种具有高效、易构建特点的语义分割模型,常用于生物医学图像的分类任务中。在该类任务中图片数据往往较少,与本研究中所使用的数据情况类似,且在组合反射率图中感兴趣区域只占整个图片的一小部分,这与 U-Net 网络应用于医学图像分类的思想贴近,同时研究为像素级的分

类,与 U-Net 应用于语义分割的场景相似,因此在研究中采用 U-Net 网络作为研究所用深度学习模型。模型结构可以分为 3 部分,第一部分主要为主干特征的提取,即编码器部分;第二部分为加强特征提取,即解码器部分;最后一部分为预测部分。传统的 U-Net 网络如图 4 所示,其主干提取部分在结构上由 5 个下采样模块组成,每个下采样模块由 2 个  $3 \times 3$  卷积层和  $2 \times 2$  的 maxpooling 组成,通过第一部分的下采样操作后可以获得 5 个有效特征层,在第二部分则对前一部分得到的特征层进行上采样,在 U-Net 网络上采样是通过反卷积实现的,可表示为

$$X = C^T Y$$

其中  $X$  与  $C$  分别为输入和参数矩阵,  $Y$  为输出。

经过上采样之后得到的特征层再与前一部分的特征层进行特征融合,将得到一个含有所有特征的有效特征层,最后在预测部分对得到的有效特征层进行特征点分类。

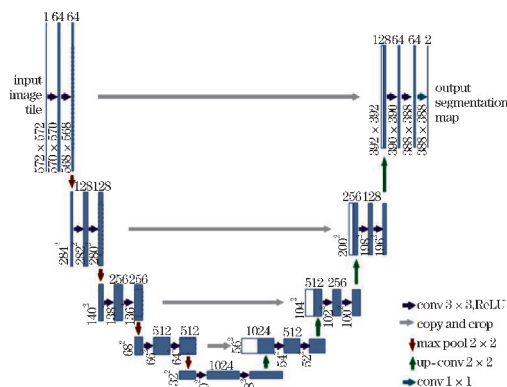


图4 传统 U-Net 网络结构图

虽然传统的 U-Net 与其他分类网络相比具有许多优点,但仍然存在不足之处。如卷积层的步距为 1,且在卷积过程中没有进行 padding 操作,因此通过卷积之后高和宽都会减少,这会造成在模型输出时,其输出尺寸与输入不同,即输出的分割图不是原始输入图片大小,而是输入图片中心部分区域的结果,这些不足对模型最终的识别效果产生影响。因此根据以上缺点进行修改:即在网络的卷积层中加入 padding,保证每次卷积之后输入的高和宽不会发生改变,同时在卷积和 relu 激活函数中间加入 batch normalization 模块,之后在解码器部分就不用先将编码器部分的结果进行裁剪再拼接,而是直接进行拼接,保证在输出部分的大小与输入部分大小一致,由此可以解决以上所提及的不足。修改后的 U-Net 网络如图 5 所示。



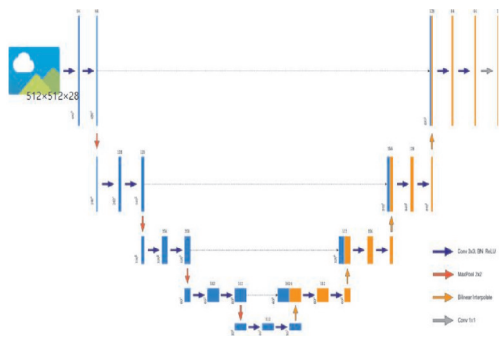


图 5 改进 U-Net 网络图

在使用改进后的模型进行训练时,输入数据的尺寸为 512×512×28,输入后先进行 2 次通道数为 64 的 3×3 卷积,再进行 2×2 的最大池化,实现一次下采样操作,得到尺寸为 256×256×64 的特征层。后面 3 次下采样仅对前一次下采样模块中的卷积通道数进行加倍处理,根据类似的步骤依次再进行 3 次下采样后,得到 3 个尺寸分别为 128×128×128、64×64×256、32×32×512 的特征层。最后一次下采样的卷积通道数与上一次相同,但去掉最大池化的操作,得到尺寸 32×32×512 的特征层。获得 5 个特征层之后,在解码器部分,首先对编码器输出的最后一个特征层进行上采样操作,再将其与前一个特征层进行拼接,并连续进行 2 次卷积处理,这构成一组完整的上采样过程。经过 4 次上采样操作之后,得到一个 512×512×64 的特征层,最后利用 1×1 的卷积进行通道调整,将输出通道数目调整为分类数目。

在所使用的网络中,采用 relu 作为卷积层的激活函数,使用 Focal loss 作为研究所用的损失函数在模型训练过程中对模型进行微调,其定义如下:

$$FL(p_i) = -\alpha_i(1-p_i)\gamma\lg(p_i)$$

其中  $\alpha$  权重帮助处理不同类别的不均衡,  $(1-p_i)$  为调节因子,  $\gamma \geq 0$  为可调节聚焦参数,其范围为  $[0, 5]$ 。  $\gamma$  控制曲线的形状,其值越大,易分类样本的 loss 就越小,可以使模型的注意力投向那些难以分类的样本,优化器 Adam 的初始学习率设为 0.0001,并在训练过程中使用学习率衰减的方法,网络的输入为 28 层 512×512 的 npy 文件(7 层组合反射率数据和 3 部雷达分别在各仰角的径向速度数据),输出为 512×512 的标签图。在进行了 125 次迭代后 loss 降到当前模型最低值,将此时得到的训练权重保存到 best\_epoch\_weights.pth 文件中,在预测阶段,加载训练得到的权重对待预测的输入数据进行预测得到对应的预测标签图,同时将预测标签图与对应的组合反射率图片进行叠加可以得到最终的分类结果。

2.2 评价指标

评价参数是网络训练中必需的元素之一,通过评

价参数可以评判分割的优劣,主要采用平均像素准确率(mean pixel accuracy, MPA)、均交并比(mean Intersection over union, MIoU) 两种指标对 U-Net 网络进行评估。指标的计算基于混淆矩阵,如表 2 所示。

表 2 混淆矩阵

真实情况	预测结果	
	正例	反例
正例	TP(真正例)	FN(假反例)
反例	FP(假正例)	TN(真反例)

MIoU 的计算可描述为预测区域和实际区域交集与并集的比值,即假设网络为  $k+1$  类网络,  $i$  与  $j$  分别表示真值和预测值,  $p_{ij}$  表示本属于类  $i$  但被预测为类  $j$  的像素量,即 FN,同理得  $p_{ii}$  为 TP,  $p_{ji}$  为 FP,因此其计算公式如下:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN+FP+TP}$$

MPA 为对 PA 的改进,是先对各个类别计算 PA,然后对所有类的 PA 求平均,其计算式如下:

$$MPA = \frac{1}{k+1} \sum_{i=0}^k \frac{TP+TN}{FN+TP+FP+TN}$$

在对指标进行计算时,利用前面提及的划分比例为 20% 的验证集数据,即 98 个个例进行指标计算,最后使用改进前后的模型,计算所得的指标结果如表 3 所示。各类指标的分析在结论部分给出。

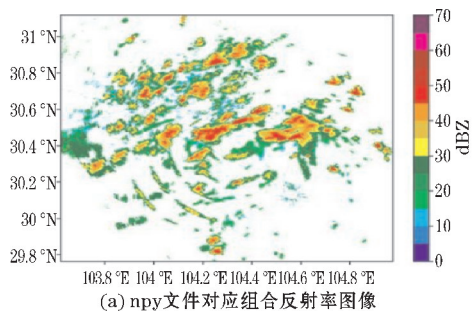
表 3 语义分割指标结果 单位: %

分类类别	MIoU	MPA
强对流改进前	69.63	74.36
强对流	80.19	86
层状云改进前	73.14	77.82
层状云	87	90.32

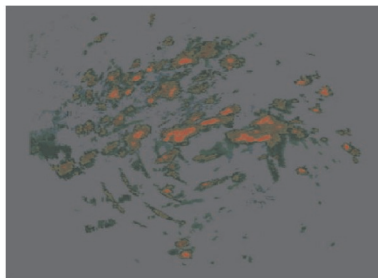
3 数据效果分析

研究主要注重于强对流区域的识别且为了验证所得模型具有普适性,在预测阶段中,选取时间区别于训练过程的强对流个例数据输入网络进行识别分类。网络输出时会同时输出前景和背景两部分,再将输出结果叠加到对应的组合反射率图上,由此得到最终的模型预测结果。在识别结果中,红色高亮的部分为模型预测的强对流部分,绿色高亮部分为网络预测的层状云部分,黑色阴影部分为网络预测的背景部分,对所选择的 3 个个例的预测结果如图 6~8 所示。

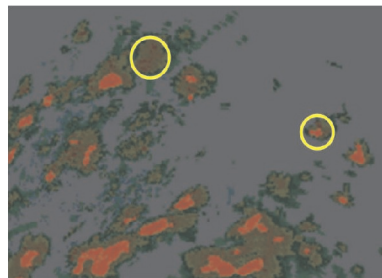




(a) npy文件对应组合反射率图像

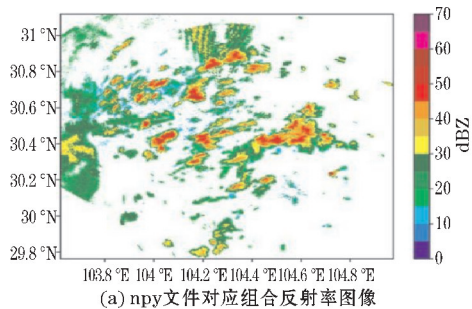


(b) 识别结果图

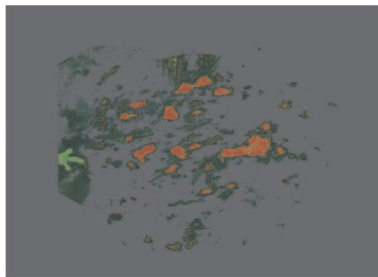


(c) 识别结果局部放大图

图6 0709\_0631个例识别结果



(a) npy文件对应组合反射率图像

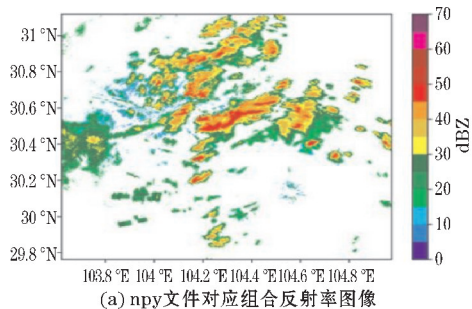


(b) 识别结果图



(c) 识别结果局部放大图

图7 0709\_015个例识别结果



(a) npy文件对应组合反射率图像



(b) 识别结果图



(c) 识别结果局部放大图

图8 0709\_152个例识别结果

从图6~8可以看出,模型的预测结果对大部分强对流区域的识别能达到一定的准确率,但对个别强对流区域的识别还是存在着一些不足,如出现在小范围密集分布的强对流情况时,预测得到的结果会将附近的一些非强对流区域划分到强对流类别中,或是对一些面积较小强对流区域的划分时,出现未将其识别为强对流区域的情况(如局部放大图中黄色圆圈所示),从而产生一些误差。而对层状云的识别分类效果则明显优于强对流的识别效果。其主要原因是在成都地区发生强对流天气的概率较小,因此在所选取的时间段内的强对流个例数目会明显少于层状云,存在样本不对等的现象。同时,在所标记的数据集中强对流现象出现时在雷达回波图上的面积与层状云出现时相比大多都较小,对其进行标记出现误差的概率也会高于层状云,即在对强对流区域进行标记时不能完全贴合其形状进行标记,这也会对识别的效果造成影响。

## 4 结束语

使用3部雷达构成的组网雷达组合反射率和径向速度数据对修改后的U-Net网络进行训练,经过对网

络调参和不断调整迭代次数后得到最佳模型,并使用训练得到的模型进行预测,得到MIoU为80.19%、MPA为86%的结果。但本研究中仍然存在一些不足,可从以下几方面改进:

(1)解决样本中强对流天气和层状云天气数量不对等的问题。如对强对流个例的数据进行数据增强操作,即根据已有的强对流个例生成新的训练样本,使所构建的数据集中的强对流个例比例得到提高,进而可以提高网络对强对流特征学习的深度。

(2)在利用labelme进行数据集制作的过程中,尽可能让标记的点更加贴合层状云和强对流云的边缘,同时在边缘处标记的更为精细一点,有利于提高网络训练过程中,对强对流和层状云的识别效果。

(3)选择能够训练更深的网络,使在训练过程中网络获得更多的数据特征,以便在之后的预测环节得到更好的结果。

## 参考文献:

[1] 姚蓉,唐佳,兰明才,等.一次混合强对流天气的环境条件及演变特征分析[J].暴雨灾害,2017,

- 36(3):217–226.
- [2] Junyent, Francesc, Chandrasekar, et al. The CASA Integrated Project 1 Networked Radar System[J]. Journal of Atmospheric and Oceanic Technology, 2010, 27(1):61–78.
- [3] 马舒庆,陈洪滨,王国荣,等. 阵列天气雷达设计与初步实现[J]. 应用气象学报, 2019, 30(1):1–12.
- [4] Best, William H, Jr. Radars over the Hump: Recollections of the First Weather Radar Network[J]. Bulletin of the American Meteorological Society, 1973, 54(3):205–208.
- [5] Wang Yanting, Chandrasekar. Quantitative Precipitation Estimation in the CASA X-band Dual-Polarization Radar Network[J]. Journal of Atmospheric and Oceanic Technology, 2010, 27(10):1665–1676.
- [6] Barbary M, Zong P. Optimisation for stealth target detection based on stratospheric balloon-borne netted radar system[J]. IET Radar, Sonar & Navigation, 2015, 9(7):802–816.
- [7] Johnson J T, MacKeen P L, Witt A, et al. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm[J]. Weather and Forecasting, 1998, 13(2):263–276.
- [8] 段艺萍,徐亚娟,支树林. 改进的 SCIT 算法在人工防雹作业中的应用分析[J]. 气象与减灾研究, 2014, 37(4):42–49.
- [9] 张栋. 修正 SCIT 算法在冰雹识别、跟踪、预警中的应用[J]. 内蒙古气象, 2020(1):40–44.
- [10] 王萍,牛智勇. 基于多普勒天气雷达数据的中层径向辐合自动识别及其与强对流天气的相关性研究[J]. 物理学报, 2014, 63(1):432–444.
- [11] 南刚强,陈明轩,秦睿等. 基于雷达组合反射率拼图和深度学习的尺度对流系统识别、追踪与分类方法[J]. 气象学报, 2021, 79(6):1002–1021.
- [12] Jergensen E G, McGovern A, Lagerquist R, et al. Classifying Convective Storms Using Machine Learning[J]. Weather and Forecasting, 2020, 35(2):537–559.
- [13] Zi-Xin Xu, Ling Yang, Bo Lei, et al. An Automatic Framework of Region-of-Interest Detection and Classification for Networked X-Band Weather Radar System[J]. Earth and Space Science, 2020, 7(4):1–11.
- [14] Jingjing G, Haihong T. Cramer-Rao lower bounds of target positioning estimate in netted radar system[J]. Digital Signal Processing, 2021, 118.
- [15] Mecikalski J R, Williams J K, Jewett C P, et al. Probabilistic 0–1 hour convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data[J]. Journal of Applied Meteorology and Climatology, 2015, 54(5):1039–1059.
- [16] 韩丰,龙明盛,李月安等. 循环神经网络在雷达临近预报中的应用[J]. 应用气象学报, 2019, 30(1):61–69.
- [17] 郭瀚阳,陈明轩,韩雷等. 基于深度学习的强对流高分辨率临近预报试验[J]. 气象学报, 2019, 77(4):715–727.
- [18] 刘新伟,黄武斌,蒋盈沙等. 基于 LightGBM 算法的强对流天气分类识别研究[J]. 高原气象, 2021, 40(4):909–918.
- [19] 李兆明,陈洪滨,段树等. 网络化多普勒天气雷达反射率因子的衰减订正[J]. 气象科学, 2015, 35(5):593–598.

## Identification of Severe Convection based on X-band Netted Radar

LAN Jingke<sup>1,2</sup>, YANG Ling<sup>1,2</sup>, SHI Chunxiang<sup>3</sup>, XU Zixin<sup>1,2</sup>, LIAN Jinxin<sup>1,2</sup>, YU Jie<sup>1,2</sup>

(1. College of Electronic Engineering, Chengdu University of Information Technology, Chengdu 610225, China; 2. Key Laboratory of Atmospheric Sounding, China Meteorological Administration, Chengdu 610072, China; 3. Meteorological Data Research Office, National Meteorological Information Center, China Meteorological Administration, Beijing, 10081, China)

**Abstract:** Severe convective weather, such as hail and rainstorms, does great harm to human activities, but it is very difficult to identify them automatically because of their rapid development and complex structure. Based on the radar network platform composed of three X-band radars in Chengdu Airport, Longquan, and Ziyang, this study improved the traditional U-Net network by using the bulk sweep data of the netted radar with high spatial and temporal resolution, so as to propose a strong convective weather recognition model based on U-Net semantic segmentation network. In this study, the radial velocity and reflectance data from seven layers of elevation base data of  $0.5^\circ$ ,  $2^\circ$ ,  $3.5^\circ$ ,  $5^\circ$ ,  $6.5^\circ$ ,  $8^\circ$  and  $9^\circ$  in the radar network were selected, and the data sets needed for model training were formed after pre-processing such as registration and normalization. The experimental results show that the trained model has a good result for strong convection recognition.

**Keywords:** U-Net; strong convection identification; semantic segmentation