

周必高,鲁小琴,郑峰,等,2022. 基于改进的 VGG16 模型的副热带高压相似识别及应用评估[J]. 气象,48(12):1608-1616.
Zhou B G, Lu X Q, Zheng F, et al, 2022. Research on the similarity recognition and application evaluation of subtropical high based on modified VGG16 model[J]. Meteor Mon, 48(12):1608-1616(in Chinese).

基于改进的 VGG16 模型的副热带高压 相似识别及应用评估 *

周必高^{1,2} 鲁小琴³ 郑 峰^{1,2} 黄克慧^{1,2} 洪水洁⁴ 谢海华^{1,2} 赵兵科³

1 浙江省温州市气象局,温州 325000

2 浙江省温州市台风监测预报技术重点实验室,温州 325000

3 中国气象局上海台风研究所,上海 200030

4 四创科技有限公司,福州 350000

提 要: 台风预报除常规方法外,查找历史相似作为预报和决策的参考依据是常用手段,但从海量历史台风中检索相似费时费力。提出了一种基于改进的视觉几何组模型 VGG16 的副热带高压(以下简称副高)相似检索方法,进行基于副高相似的历史相似台风查询。通过对 1979—2020 年台风季 19736 个对应时次的副高图像提取、数据增强、模型学习和优化,并以学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)作为副高相似的度量指标,最终建立了改进的 VGG16 模型。试验结果表明,使用该模型可以找出较为相似的历史台风,模型检索得到的排名第一的历史相似台风与目标台风相似度高达 92.55%,该方法可为台风预报业务人员提供了积极参考。同时,该模型相较于传统的人工识别,识别时间较短、检索效率高,可在业务及科研中推广应用。

关键词: 台风,副热带高压,VGG16 模型,LPIPS (learned perceptual image patch similarity),几何图像算法

中图分类号: P456

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2022.042901

Research on the Similarity Recognition and Application Evaluation of Subtropical High Based on Modified VGG16 Model

ZHOU Bigao^{1,2} LU Xiaoqin³ ZHENG Feng^{1,2} HUANG Kehui^{1,2}

HONG Shuijie⁴ XIE Haihua^{1,2} ZHAO Bingke³

1 Wenzhou Meteorological Office of Zhejiang Province, Wenzhou 325000

2 Wenzhou Key Laboratory of Typhoon Monitoring and Forecasting Technology of Zhejiang Province, Wenzhou 325000

3 Shanghai Typhoon Institute, CMA, Shanghai 200030

4 ISTRONG Technology Co., Ltd., Fuzhou 350000

Abstract: Historical similar typhoons are important references for typhoon prediction and decision-making besides conventional methods. However, it is time-consuming and laborious to retrieve similar information from a large number of historical typhoons. In this paper, a similarity retrieval method of subtropical high based on improved visual geometry group model (VGG16) is proposed to query historical similar typhoons based on subtropical high similarity. Through the image extraction, data enhancement, model learning and optimization of 19736 subtropical high corresponding times in the typhoon season from 1979 to 2020, taking

* 中国气象局上海台风研究基金项目(TFJJ202013)和上海市自然科学基金项目(21ZR1477300)共同资助

2022 年 1 月 24 日收稿; 2022 年 7 月 6 日收修定稿

第一作者:周必高,主要从事气象大数据、软件平台、神经网络和台风业务服务技术研究. E-mail:105172878@qq.com

通讯作者:鲁小琴,主要从事台风结构客观分析技术研究及台风多源数据融合应用. E-mail:luxq@typhoon.org.cn

the learned perceptual image patch similarity (LPIPS) as the measurement index of subtropical high similarity, we establish an improved VGG16 model finally. The test results show that greatly similar historical typhoons can be found by using this model. The similarity between the historical similar typhoon ranked first retrieved by this model and the target typhoon is 92.55%, which provides a very meaningful reference for typhoon forecasters. At the same time, compared with the traditional manual recognition, this model takes shorter recognition time and has higher retrieval efficiency, and can be applied in typhoon forecasting and research.

Key words: typhoon, subtropical high, VGG16 model, LPIPS (learned perceptual image patch similarity), geometric image algorithm

引 言

我国地处太平洋西岸,海岸线漫长,从广西到辽宁常年都会受到台风的影响(陈联寿和丁一汇,1979)。尤其在夏季,沿海及内陆受台风影响地区常常会因台风袭击而造成人员伤亡和重大财产损失,因此精准的台风路径、强度及风雨影响预报对保障人民的生命财产安全具有重大的现实意义。台风路径预报的准确性是开展防台减灾服务的基础。气象学家们针对台风路径预报的方法发展了大量统计或动力模型,同时也发现历史相似台风能够为台风路径、强度和风雨预报提供非常有价值的参考信息。已有的历史台风相似检索大多基于人工主观判别进行或者针对台风路径空间相似开展(刘勇等,2006;王远飞等,2006;鲁小琴等,2013;徐高扬和刘姚,2019),但存在以下问题:(1)历史数据中采用人工识别查找历史相似台风。该方法判别时间长,不能满足预报决策的实时性需求。(2)基于空间相似的历史台风自动检索。该方法速度快,但找到的历史相似台风,大部分由于和目标台风所对应的大尺度环境流场存在差异,并且造成的风雨影响及灾害也各异,因而此类相似台风并不一定属于气象意义上的相似台风。在此基础上,预报员们提出能否根据台风环境场的相似去查找历史相似台风,进而为当前台风预报和防台减灾决策提供更为科学且专业的参考。

台风的移动受到各种复杂因素的影响,但实践表明牵引运动是最主要的因素,因此大型气压场的分布及与此相应的大范围深厚的基本气流是台风路径预报的重要根据(陈联寿和丁一汇,1979)。其中,副热带高压(以下简称副高)就是影响台风移动路径的一个非常重要的大型环流系统,其周期变化、形状特征及与西风槽的相互作用等,很大程度上引导着

未来台风走向及可能带来的风雨影响。因此以副高相似来查询得到的历史相似台风,可能可为当前台风的路径及风雨影响预报提供较为精准的辅助资讯。

科学家们已经开展了很多有关气象要素场的相似研究,大多基于格点数据进行距离或相关分析(李开乐,1986;罗阳,1996;罗阳和赵伟,2008;丁晨晨等,2019;黄小玉等,2019;宋文彬等,2020),并未针对某个特定的环流系统开展。事实上影响台风移动的副高具有特定的形状和相对的地理空间分布(如距离大陆远近等),因此具有图像特征识别功能的神经网络技术对实现大区域环境场中副高的识别和提取,并进行相似判定具有非常大的优势。且神经网络技术已在各行各业开展了大量的研究和应用,如人脸识别(吴素雯和战荫伟,2017)、字符识别(葛明涛等,2014)和语音识别(Xue et al,2014)等。气象领域中,黄小燕和金龙(2013)在综合考虑台风移动路径的气候持续因子和数值预报产品动力预报因子的基础上,建立了基于主成分分析的人工智能台风路径预报模型,预报精度高于气候持续法;叶小岭等(2014)基于主成分分析及 RBF 神经网络开展了浙江省台风灾害损失预测;代刊等(2013)基于自组织神经网络算法分析了华北区域夏季天气气候学特征;陈文兵等(2019)利用 G-CNN 模型开展了浓雾天气形势识别。以上研究表明基于一般神经网络技术在气象领域进行天气特征识别(如浓雾)或预报预测(如台风路径和强度、灾害等)已取得了一些成效,但采用神经网络技术对台风大尺度环境场中的副高进行识别和判定,成果鲜见。

本文拟基于卷积神经网络(convolutional neural network,CNN)的视觉几何组模型(visual geometry group,VGG16)(Lecun et al,1998;Simonyan and Zisserman,2016)来识别台风大尺度环境场中

的副高,并以学习感知图像块相似度(learned perceptual image patch similarity, LPIPS)(Wang et al, 2004)作为图像相似度度量指标,建立副高相似的识别算法,进而开展对应历史相似台风查询,以期台风路径预报提供更高效而精准的参考信息。

1 数据及预处理

1.1 数据说明

本文中大尺度环境场主要用 500 hPa 位势高度来表征,资料来自 ERA5 再分析数据集(<https://cds.climate.copernicus.eu/>)。资料集覆盖时间范围为 1979—2020 年,时间分辨率为 6 h(逐日 00、06、12 和 18 时,世界时,下同);空间范围为 $90^{\circ}\text{S}\sim 90^{\circ}\text{N}$ 、 $180^{\circ}\text{E}\sim 180^{\circ}\text{W}$,空间分辨率为 $1^{\circ}\times 1^{\circ}$ 。为节约计算资源,只截取本研究关注的西北太平洋区域($10^{\circ}\text{S}\sim 75^{\circ}\text{N}$ 、 $75^{\circ}\text{E}\sim 165^{\circ}\text{W}$),用于开展对历史副高的识别和相似检索分析。

台风路径数据(Ying et al, 2014; Lu et al, 2021)来自中国气象局热带气旋资料中心的台风最佳路径数据集(https://tcdata.typhoon.org.cn/zjljsjj_zlhq.html),包含 1979—2020 年发生在西北太平洋及南海地区的逐台风记录时间、中心经度和纬度、中心最大风速和最低气压,时间分辨率为 6 h(逐日 00、06、12 和 18 时)。

1.2 数据预处理

为了便于后续算法模型对副高的识别,对 ERA5 再分析数据进行预处理:(1)数据提取。读取 500 hPa

位势高度层网格数据,并提取大于 5880 gpm 的格点数值用来表征副高。(2)数据清洗。采用图像形态学算法中的腐蚀和膨胀运算等技术,对部分孤立点、边缘处裂痕等异常数据进行剔除或弥补,对提取得到的副高图像进行边缘光滑,以提高副高相似分析的准确性。(3)近似多边形提取。采用 Canny 边缘检测技术,提取副高边缘上的离散点并连接形成多边形,将其作为最终副高轮廓。图 1 为 2019 年 8 月 5 日 00 时台风环境场和预处理提取到的副高对照图,可见该时刻台风中心位于副高西南侧,在副高的影响下呈西北行趋势;而预处理提取到了完整的副高形态和分布。

2 神经网络模型及相似度度量指标

2.1 CNN 基本结构

CNN 由输入层、隐含层和输出层组成(图 2),隐含层中又包含卷积层、池化层和全连接层(Zeiler and Fergus, 2014)。其中,在卷积层中,利用设定大小的卷积核和指定步长在输入图像(原图或特征图)上进行滑动并做卷积计算,该层的作用是提取输入图像的主要关键特征;池化层的作用是对提取的特征进行抽样,在减轻网络模型过拟合程度(Matsugu et al, 2003)的同时可以减小后续运算量。卷积层和池化层合作,实现对输入图像的特征提取并抽样降维,能提取出副高形态中最具有代表性的关键特征;卷积层和池化层一般交替出现在网络中,一个卷积层加一个池化层被称为一个特征提取过程。最后,全连接层负责把提取的特征图连接起来,通过分类

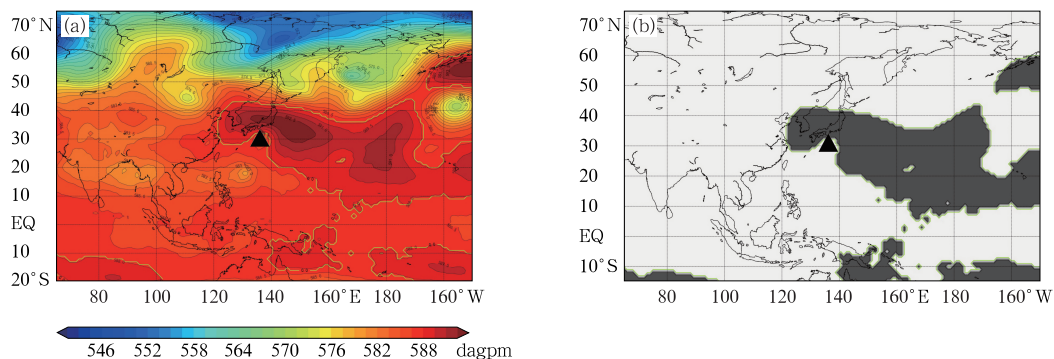


图 1 2019 年 8 月 5 日 00 时(UTC)(a)台风环境场和(b)预处理提取的副高
(黑色三角形:台风中心位置)

Fig. 1 Comparison of (a) environmental field of typhoon and (b) extracted subtropical high at 0000 UTC 5 August 2019
(black triangle: typhoon center)

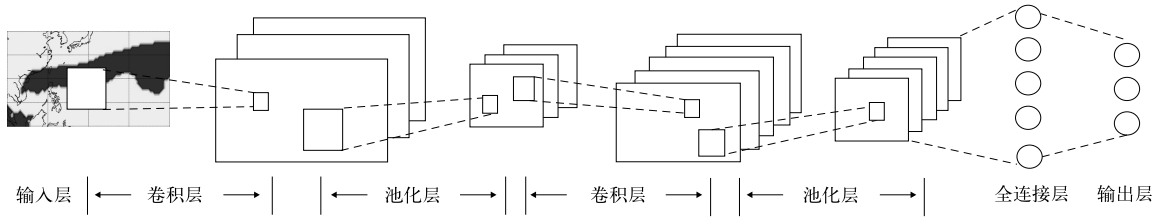


图 2 卷积神经网络结构示意图

Fig. 2 Structure diagram of convolutional neural network

器得到最终的分类估算结果(郑泽宇等,2018)。

2.2 VGG16 模型

VGG16 是 CNN 的一个经典模型,该模型具有 5 个卷积组(13 个卷积层)和 3 个全连接层(图 3),在使用同等大小卷积核的条件下,提升了网络的深度和效果(Simonyan and Zisserman, 2016)。其网络结构非常一致,全部使用 3×3 卷积和 2×2 池化。该模型相比于其他卷积神经网络模型,主要采用连续的 3×3 的小卷积核代替较大卷积核。对于给定的特征图,连续采用小卷积核堆积的策略要优于大的卷积核,因为较高深度的多层非线性层能使卷积神经网络学习更复杂的模态,并且需要的参数较少,模型的计算代价更小,一定程度上提升了神经网络的效果。

2.3 LPIPS 度量指标

LPIPS 也被称为“感知损失”,用于度量两张图像之间的相似程度。本研究当中主要使用 LPIPS 来度量两幅副高图像的相似程度。该指标在视觉相似性判断上更接近人类感知(Zhang et al, 2018)。其计算公式如下:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h, w} \| \mathbf{w}_l \odot (\hat{\mathbf{y}}_{hw}^l - \hat{\mathbf{y}}_{0hw}^l) \|_2^2 \quad (1)$$

式中: x, x_0 为要比较的两幅副高图像; d 是 x 与 x_0 之间的 LPIPS 值; l 是网络的层数; h 和 w 是特征层的两个不同维度; 从第 l 层提取特征并在通道维度中进行单元标准化得到 $\hat{\mathbf{y}}_{hw}^l - \hat{\mathbf{y}}_{0hw}^l$; H_l 和 W_l 分别是第 l 层特征的高和宽; \odot 为矢量相乘运算; $\| \cdot \|_2^2$ 为 2-范数的平方。LPIPS 的值越低表示两张图像差异越小,越相似;反之,则差异越大。

为了便于比较输出图像相似度,副高 x 与 x_0 的图片相似度 P 定义为:

$$P(x, x_0) = [1 - d(x, x_0)] \times 100 \quad (2)$$

3 模型的建立

3.1 改进的 VGG16 副高识别模型构建

VGG16 模型是基于大图像数据库(ImageNet)训练而成的,具备较强的深度特征学习能力的参数和权重网,特别是对卷积层输入的图像轮廓、边缘、曲线等特征提取具有较强的能力。因此,为减少网

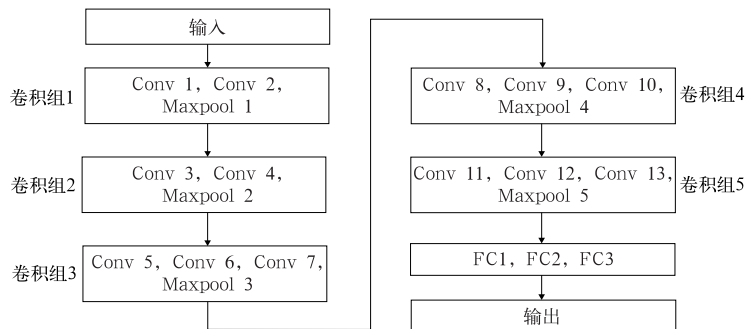


图 3 VGG16 网络架构

Fig. 3 Network architecture of VGG16

络训练时间和提高网络训练效率,将已训练好的 VGG16 模型作为预训练模型,采用迁移学习的方式(关胤,2019),引入副高识别模型进行卷积层、池化层再训练。VGG16 模型的输入和网络架构参数设置改进如下文所述,模型训练流程如图 4 所示。

3.1.1 训练样本

将输入的三通道彩色图改变为单通道灰度图,输入尺寸由 $224 \times 224 \times 3$ 改为 $86 \times 44 \times 1$ 。图像灰度化并框定小区域之后可以大幅减少计算量,同时可重点突出副高在大尺度环境场中的空间位置和大小。设置数据集中用于学习和测试的样本比例分别为 80% 和 20%。

3.1.2 数据增强

为弥补 CNN 训练需要海量的数据样本而当前历史副高图库中的样本数量较少的缺陷,采用数据增强的方法,对已有的数据样本进行翻转、剪裁、高斯噪声等处理,并生成新的数据样本来补充原图库样本资料集,由此扩大了训练集和测试集的样本数量,同时提高了模型的泛化能力,并减少过度拟合。

3.1.3 网络架构参数设置

设定本模型中使用 3 个卷积组,网络架构的轻巧化有利于快速提取副高图像的特征数据;每个卷积组设定只含 2 个卷积层和 1 个池化层,但为充分提取副高细微图像特征,在个别卷积组中增加卷积核的数目;同时只进行图像特征提取,即去掉分类

器,只包括卷积层、池化层和全连接层。

3.2 模型训练

模型训练采用 Tensorflow 深度学习框架,处理器使用 Intel 酷睿 i7—9700,机器内存为 16 GB,优化器使用 Adam,批大小为 30,学习率为 4×10^{-3} ,损失函数采用 Cross Entropy Loss 计算。

针对 1979—2020 年台风季 19 736 个对应时次的副高图像样本,设定 8 : 2 的样本划分比例确定训练集和测试集,对模型进行学习训练(图 5)。由图 5a 可见,当迭代次数到达 20 次,该模型的训练集和测试集准确率都明显增加,迭代达 80 次之后,该模型的训练精准度和损失函数值趋于平衡,其训练集和测试集准确度峰值分别为 89.7% 和 86.5%。从图 5b 中可以看出,迭代 80~100 次后,损失函数值达到最小,表明模型已收敛并稳定。保存此时模型结构参数及网络权重为最终副高相似识别模型。

4 模型的试验及检验

4.1 模型的试验

以 0309 号台风为例(以下称为目标台风),选 2003 年 8 月 4 日 06 时副高图像(图 6a)作为检索目标,利用改进的 VGG6 模型进行相似检索试验。

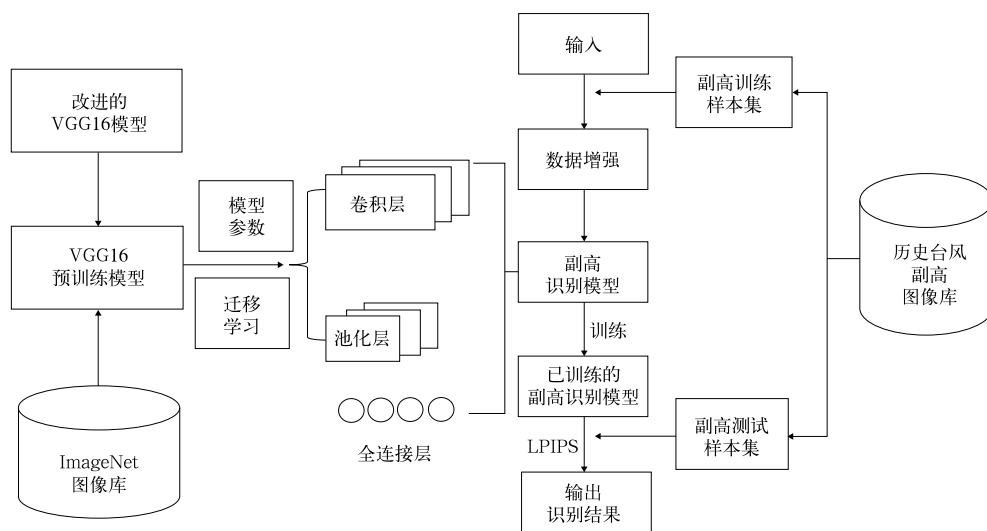


图 4 副高相似识别模型训练流程图

Fig. 4 The training flow chart of similarity recognition model of subtropical high

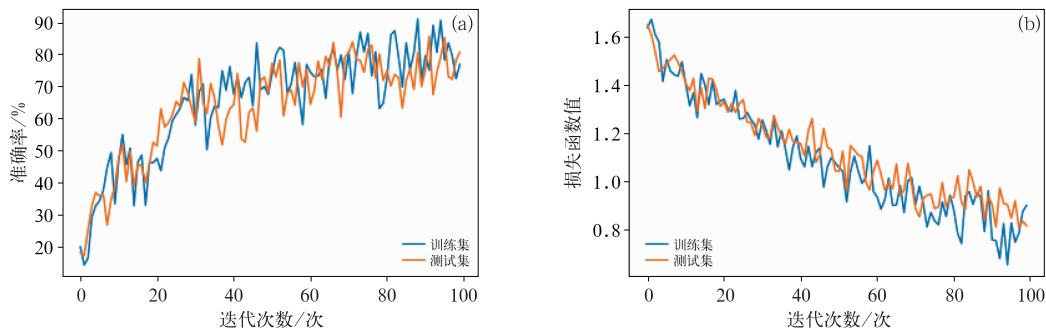


图 5 1979—2020 年台风季样本训练过程

(a)准确率,(b)损失函数

Fig. 5 Training process of samples in typhoon season from 1979 to 2020

(a) accuracy, (b) loss function

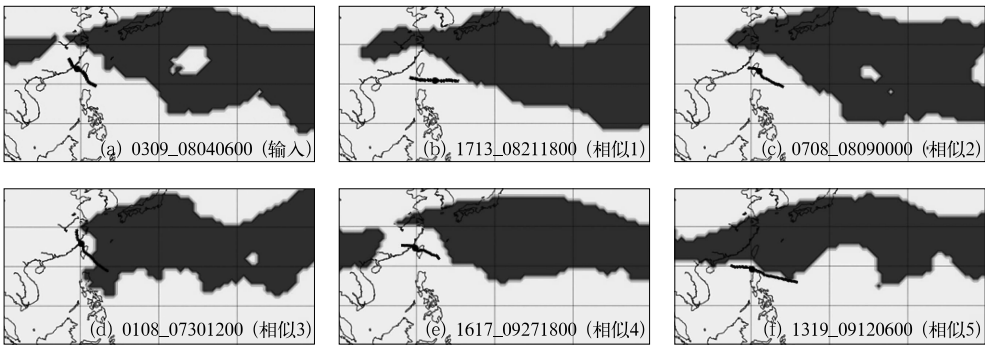


图 6 (a)2003 年 8 月 4 日 06 时 0309 号台风及其对应副高,(b~f)相似检索得到的历史相似台风及其对应副高

(b)1713 号,(c)0708 号,(d)0108 号,(e)1617 号,(f)1319 号

(黑点:台风中心位置,黑线:前后 36 h 台风路径)

Fig. 6 (a) Typhoon No. 0309 and its corresponding subtropical high at 06:00 UTC 4 August 2003, (b—f) historical similar typhoons and the corresponding subtropical high obtained by similar retrieval

(b) No. 1713, (c) No. 0708, (d) No. 0108, (e) No. 1617, (f) No. 1319

(black dot: location of typhoon center, black line: typhoon track before and after 36 h)

4.1.1 相似副高图像比较

利用本模型检索得到的历史相似台风排名前五位的分别为 1713 号(图 6b)、0708 号(图 6c)、0108 号(图 6d)、1617 号(图 6e)和 1319 号台风(图 6f)。图 6 显示,相似台风对应时刻的副高形状(西伸型)、副高分布范围大小(南北跨度及东西长度)及伸至大陆以西时的略微南掉都非常一致。就视觉感知情况而言,其中 1713 号台风的对应时刻副高最为相似(图 6b),0108 号台风的对应时刻副高(图 6d)相似度略差。

4.1.2 LPIPS 相似度值的比较

计算检索得到的排名前五位台风对应时刻副高相似度 LPIPS 值(表 1)。发现 1713 号台风的对应 LPIPS 值最高,同视觉感知情况一致;但是 LPIPS 值排在前五位的台风就视觉感知而言其相似程度排

序会有不同,这与人眼识别和客观判据的差异有关。

表 1 前五位相似台风对应副高间的 LPIPS 值

Table 1 The LPIPS value of corresponding subtropical high of top 5 similar typhoons

相似度 排序	相似台风 编号	对应副高相似时间 /年-月-日-时	LPIPS/%
/	0309	2003-8-4-06	/
1	1713	2017-8-21-18	92.55
2	0708	2007-8-9-00	92.25
3	0108	2001-7-30-12	92.00
4	1617	2016-9-27-18	85.55
5	1319	2013-9-12-06	84.50

综上所述可见,利用改进的 VGG16 模型识别检索到的历史相似台风对应的副高,从视觉特征上看与目标台风的副高大小、位置、形状是一致的,且对应的 LPIPS 值的大小与视觉特征的相似度也基

本匹配。

4.2 模型效果检验

为了检验 VGG16 模型检索到的历史相似副高与目标副高的相似程度,以目标台风对应的 2003 年 8 月 4 日 06 时副高的相似检索结果为例,针对目标台风和历史相似台风对应时刻的副高区域分别提取近似多边形,分别定义为 A 和 B ,通过几何图像算法分别分析其大小(面积)、重心(位置)和形状(轮廓)等几何特征差异(周必高等,2021),从几何学的角度综合比较其相似程度。

4.2.1 大小相似度

采用面积作为参数来衡量副高大小相似程度。获取 A 和 B 的面积 S_A 和 S_B ,使用式(3)计算面积相似度:

$$\text{SimArea}(A, B) = 1 - \text{abs}(S_A - S_B) / \text{Max}(S_A, S_B) \quad (3)$$

式中: S_A 、 S_B 分别为 A 和 B 的面积, $\text{abs}(S_A - S_B)$ 表示求取 S_A 、 S_B 差的绝对值, $\text{Max}(S_A, S_B)$ 为取 S_A 、 S_B 最大值, $\text{SimArea}(A, B)$ 为 A 和 B 间的面积相似程度值。以目标副高的相似检索结果为例,面积大小相似度排名前五位如表 2 所示。

表 2 同表 1, 但为面积相似度

相似度排序	相似台风编号	对应副高相似时间/年-月-日-时	面积/像素	面积相似度/%
/	0309	2003-8-4-06	1275.5	/
1	1713	2017-8-21-18	1255.0	98.41
2	0708	2007-8-9-00	1366.5	93.34
3	0108	2001-7-30-12	1163.5	91.92
4	1319	2013-9-12-06	1097.5	87.75
5	1617	2016-9-27-18	1032.5	83.99

4.2.2 位置相似度

提取 A 和 B 的质心坐标 $A(X_a, Y_a)$ 、 $B(X_b, Y_b)$ 作为副高中心位置,其中质心定义为副高多边形的几何中心;通过欧氏距离计算两质心间的距离,并将距离归一化到 $[0, 1]$ 范围之内,作为位置相似度衡量指标。

欧氏距离计算公式如下所示:

$$\text{Dis}(A, B) = \sqrt{(X_a - X_b)^2 + (Y_a - Y_b)^2} \quad (4)$$

位置归一化算法如下所示:

$$\text{SimDis}(A, B) = 1 / [\text{Dis}(A, B) / C + 1] \quad (5)$$

式中: X_a 、 Y_a 和 X_b 、 Y_b 分别为副高质心 A 和 B 的坐标, $\text{Dis}(A, B)$ 为 A 和 B 质心间的距离; C 为 A 和 B 的任意边界点间距离的最大值, $\text{SimDis}(A, B)$ 为归

一化后的位置相似度值。位置相似度排名前五位如表 3 所示。

表 3 同表 1, 但为位置相似度

相似度排序	相似台风编号	对应副高相似时间/年-月-日-时	位置坐标	位置相似度/%
/	0309	2003-8-4-06	44, 26	/
1	1713	2017-8-21-18	44, 26	100
2	0708	2007-8-9-00	43, 26	95.23
3	0108	2001-7-30-12	42, 27	89.84
4	1617	2016-9-27-18	41, 28	84.73
5	1319	2013-9-12-06	37, 27	73.80

4.2.3 轮廓相似度

基于射线的几何匹配方法,以副高近似多边形的几何中心为原点,从原点发出射线与 x 轴的夹角为 Q_i ,将 Q_i 作为几何形状描述函数的参数, $Q_i \in [0, 2\pi]$ 。射线分别相交于近似多边形 A 和 B 的轮廓边界于点 P_{Ai} 和 P_{Bi} ,分别计算交点 P_{Ai} 和 P_{Bi} 到原点的距离: $L_A(Q_i) = |P_{Ai}O|$ 和 $L_B(Q_i) = |P_{Bi}O|$ 。则 A 和 B 在 Q_i 方向上的相似度为:

$$\text{Simangle}(Q_i) = 1 - \frac{|L_A(Q_i) - L_B(Q_i)|}{\text{Max}[L_A(Q_i), L_B(Q_i)]} \quad (6)$$

式中: $\text{Max}[L_A(Q_i), L_B(Q_i)]$ 为取这两个距离中的最大值, $\text{Simangle}(Q_i)$ 则为 A 和 B 在 Q_i 方向上的相似程度值。

旋转射线角度与近似多边形 A 、 B 连续相交,则 A 和 B 的形状轮廓相似度计算公式如下:

$$\text{SimShape}(A, B) = \frac{1}{2\pi} \sum_{Q=0}^{2\pi} \text{Simangle}(Q_i) = \frac{1}{2\pi} - \frac{A(M) |L_A(Q_i) - L_B(Q_i)|}{\text{Max}[L_A(Q_i), L_B(Q_i)]} \quad (7)$$

式中: M 表示射线数量, $A(M) = 2e^{-1/M}$,用来调节差异距离大小, $|L_A(Q_i) - L_B(Q_i)|$ 为 A 和 B 在 Q_i 方向的差距, $\text{SimShape}(A, B)$ 即为 A 和 B 的轮廓相似度。以目标副高的相似检索结果为例,轮廓相似度排名前五位如表 4 所示。

表 4 同表 1, 但为轮廓相似度

相似度排序	相似台风中央台编号	对应副高相似时间/年-月-日-时	轮廓相似度/%
/	0309	2003-8-4-06	/
1	0708	2007-8-9-00	57.25
2	1617	2016-9-27-18	56.02
3	1713	2017-8-21-18	53.93
4	0108	2001-7-30-12	47.73
5	1319	2013-9-12-06	31.05

4.2.4 综合相似度

利用 A 和 B 之间的面积、位置和轮廓相似度进行综合相似度计算,计算公式如下:

$$SimLand(A, B) = aSimArea(A, B) + bSimDis(A, B) + cSimShape(A, B) \quad (8)$$

式中: $SimLand(A, B)$ 为计算得到的 A 和 B 的综合相似度。 a 、 b 和 c 分别表示对应维度的权重值,根据业务应用经验设定 a 、 b 和 c 权重值分别为 0.5、0.3 和 0.2,综合相似度排名前五位如表 5 所示。

表 5 同表 1, 但为综合相似度

Table 5 Same as Table 1, but for comprehensive similarity

相似度 排序	相似台风 中央台编号	对应副高相似时间 /年-月-日-时	综合 相似度/%
/	0309	2003-8-4-06	/
1	1713	2017-8-21-18	89.99
2	0708	2007-8-9-00	86.69
3	0108	2001-7-30-12	82.49
4	1617	2016-9-27-18	78.62
5	1319	2013-9-12-06	72.25

综上所述,通过对目标副高和相似副高的几何图像学大小、中心位置和轮廓各维度的比对,发现排名前五位的相似台风的面积如表 2 所述,相似度排序和 VGG16 模型检索出的结果完全一致。而相似副高的质心位置如表 3 所示,相似度排序和 VGG16 模型得出的结果也基本一致。轮廓相似度排序与模型检索稍有差异,一方面因为对副高本身外围边界的描述算法不尽完善;另外副高形态变化快,导致检索得到的相似副高与目标副高轮廓会略有差异。总而言之,基于综合的几何图像学比较了 VGG16 模型检索得到的相似副高与目标副高的相似度高度一致。表明改进后的 VGG16 模型在识别副高及查找历史相似副高和对应相似台风的分析中是非常有效的。

5 结论与讨论

本文针对影响台风路径移动的一个非常重要的大型环流系统——副高的相似度判别,提出了一种基于 CNN 技术的改进 VGG16 模型。利用 1979—2020 年台风季历史样本对模型进行训练和测试,使用 LPIPS 作为相似度量标准,进行历史副高相似识别。试验表明,通过该方法可以找到历史相似副高所对应的历史相似台风,解决了针对海量历史台风数据进行人工相似识别的难题;同时通过几何图像学相似检验其检索效果,结论基本一致。改进的

VGG16 模型检索效率高,可为预报员开展台风预报和防台减灾决策提供快速且更科学的参考。

但影响台风移动的副高在时刻发生变化,当前时次相似,下一个时次可能发生较大的改变,故需要综合考虑某个时间段内平均环境场的相似,对台风未来的移动走向趋势引导才具有更重要的预报决策参考意义。另外,本文仅考虑了位势高度值为 5880 gpm 的副高主体对台风路径移动的影响,实际上有时副高主体的西北侧会向外伸出高压坝或高压脊(位势高度值不等于 5880 gpm),其对台风移动也有很好的指示意义。因此,下一步将结合高压坝和高压脊,开展对确定时段内副高区域平均场的相似或者 24、48、72 h 预报场的相似检索,以期给出综合性更高,更为合理和科学的历史相似台风,为预报决策提供支撑。

参考文献

- 陈联寿,丁一汇,1979. 西太平洋台风概论[M]. 北京:科学出版社.
Chen L S, Ding Y H, 1979. Tropical Cyclones in the Northwestern Pacific [M]. Beijing: Science Press (in Chinese).
- 陈文兵,刘小明,王宏斌,等,2019. G-CNN 模型在浓雾天气形势识别中的应用[J]. 计算机工程与应用, 55(8): 124-131. Chen W B, Liu X M, Wang H B, et al, 2019. Application of G-CNN model in foggy weather situation recognition[J]. Comput Eng Appl, 55(8): 124-131 (in Chinese).
- 代刊,金荣花,胡婷,2013. 基于自组织神经网络算法的华北区域夏季天气气候学特征[J]. 气象, 39(7): 801-812. Dai K, Jin R H, Hu T, 2013. Using self-organizing maps to investigate summer synoptic climatology in North China Area[J]. Meteor Mon, 39(7): 801-812 (in Chinese).
- 丁晨晨,任福民,邱文玉,等,2019. 基于路径相似的登陆热带气旋降水之动力-统计集合预报模型[J]. 气象, 45(1): 29-37. Ding C C, Ren F M, Qiu W Y, et al, 2019. A dynamical statistical ensemble forecast model for landfalling tropical cyclone precipitation based on track similarity[J]. Meteor Mon, 45(1): 29-37 (in Chinese).
- 葛明涛,王小丽,潘立武,2014. 基于多重卷积神经网络的大模式联机手写文字识别[J]. 现代电子技术, 37(20): 19-21, 26. Ge M T, Wang X L, Pan L W, 2014. Large pattern online handwriting character recognition based on multi-convolution neural network [J]. Mod Electr Technol, 37(20): 19-21, 26 (in Chinese).
- 关胤,2019. 基于残差网络迁移学习的花卉识别系统[J]. 计算机工程与应用, 55(1): 174-179. Guan Y, 2019. Flower species recognition system based on residual network transfer learning [J]. Comput Eng Appl, 55(1): 174-179 (in Chinese).
- 黄小燕,金龙,2013. 基于主成分分析的人工智能台风路径预报模型[J]. 大气科学, 37(5): 1154-1164. Huang X Y, Jin L, 2013. An artificial intelligence prediction model based on principal component analysis for typhoon tracks[J]. Chin J Atmos Sci, 37(5): 1154-1164 (in Chinese).
- 黄小玉,张焱,叶成志,等,2019. 地面积雪天气现象人工智能观测识

- 别研究[J]. 气象, 45(9): 1189-1198. Huang X Y, Zhang Y, Ye C Z, et al, 2019. Research on artificial intelligence observation and identification of snow cover weather phenomenon on surface[J]. Meteor Mon, 45(9): 1189-1198 (in Chinese).
- 李开乐, 1986. 相似密度及其使用技术[J]. 气象学报, 44(2): 174-183. Li K L, 1986. A new similarity parameter and its application[J]. Acta Meteor Sin, 44(2): 174-183 (in Chinese).
- 刘勇, 吴必文, 王东勇, 2006. 一种台风路径相似检索的算法研究[J]. 气象, 32(7): 18-24. Liu Y, Wu B W, Wang D Y, 2006. Algorithm study on similarity retrieval of typhoon tracks[J]. Meteor Mon, 32(7): 18-24 (in Chinese).
- 罗阳, 1996. 一种新的相似性度量-高分辨相似系数[J]. 空军气象学院学报, 17(1): 23-32. Luo Y, 1996. A new similarity measure-high resolution similarity coefficient[J]. J Air Force Meteor College, 17(1): 23-32 (in Chinese).
- 罗阳, 赵伟, 2008. 相似性度量研究及最优相似系数[C]//中国气象学会 2008 年年会天气预报准确率与公共气象服务分会场论文集. 北京: 中国气象学会: 373-392. Luo Y, Zhao W, 2008. Research on similarity measurement and optimal similarity coefficient [C]//Collected Papers on Weather Forecast Accuracy and Public Meteorological Service Branch of the 2008 Annual Meeting of the Chinese Meteorological Society. Beijing: China Meteorological Society: 373-392 (in Chinese).
- 鲁小琴, 余晖, 赵兵科, 2013. 热带气旋环境场相似检索方法的对比分析[J]. 气象, 39(12): 1609-1615. Lu X Q, Yu H, Zhao B K, 2013. Study on similarity retrieval method for ambient field of tropical cyclones[J]. Meteor Mon, 39(12): 1609-1615 (in Chinese).
- 宋文彬, 郑卫江, 薛峰, 2020. 相似台风智能检索与可视化分析研究[J]. 气象, 46(10): 1330-1339. Song W B, Zheng W J, Xue F, 2020. Research of intelligent retrieval and visualization of similar typhoons[J]. Meteor Mon, 46(10): 1330-1339 (in Chinese).
- 王远飞, 陆涛, 朱海燕, 等, 2006. 基于 GIS 的热带气旋相似路径检索系统研究[J]. 测绘科学, 31(5): 124-125, 128. Wang Y F, Lu T, Zhu H Y, et al, 2006. GIS-based tropical cyclone similar path query and analysis[J]. Sci Survey Mapp, 31(5): 124-125, 128 (in Chinese).
- 吴素雯, 战荫伟, 2017. 基于选择性搜索和卷积神经网络的人脸检测[J]. 计算机应用研究, 34(9): 2854-2857, 2876. Wu S W, Zhan Y W, 2017. Face detection based on selective search and gabor optimizing convolutional neural network[J]. Appl Res Comput, 34(9): 2854-2857, 2876 (in Chinese).
- 徐高扬, 刘姚, 2019. LSTM 网络在台风路径预测中的应用[J]. 计算机与现代化, 28(5): 64-68, 73. Xu G Y, Liu Y, 2019. Application of LSTM in typhoon path prediction[J]. Comput Modernizat, 28(5): 64-68, 73 (in Chinese).
- 叶小岭, 梁伟, 邓华, 2014. 基于主成分分析及 RBF 神经网络的浙江省台风灾害损失预测[J]. 科技通报, 30(9): 13-19. Ye X L, Liang W, Deng H, 2014. Forecasting of typhoon disaster loss of Zhejiang Province based on principal component analysis and RBF neural network model[J]. Bull Sci Technol, 30(9): 13-19 (in Chinese).
- 郑泽宇, 梁博文, 顾思宇, 2018. TensorFlow: 实战 Google 深度学习框架(第 2 版) [M]. 北京: 电子工业出版社. Zheng Z Y, Liang B W, Gu S Y, 2018. TensorFlow: Practice in Google Deep Learning Framework[M]. 2nd ed. Beijing: Electronic Industry Press (in Chinese).
- 周必高, 郑峰, 朱忠勇, 等, 2021. 基于几何图像处理和神经网络的副热带高压相似方法分析: 中国, 113421294A [P]. Zhou B G, Zheng F, Zhu Z Y, et al, 2021. Analysis of subtropical high similarity method based on geometric image processing and neural network. The patent number: CN, 113421294A [P] (in Chinese).
- Lecun Y, Bottou L, Bengio Y, et al, 1998. Gradient-based learning applied to document recognition[J]. Proc IEEE, 86(11): 2278-2324.
- Lu X Q, Yu H, Ying M, et al, 2021. Western North Pacific tropical cyclone database created by the China Meteorological Administration[J]. Adv Atmos Sci, 38(4): 690-699.
- Matsugu M, Mori K, Mitari Y, et al, 2003. Subject independent facial expression recognition with robust face detection using a convolutional neural network[J]. Neural Networks, 16(5/6): 555-559.
- Simonyan K, Zisserman A, 2016. Very deep convolutional networks for large-scale image recognition[J/OL]. (2015-04-10) [2022-01-24] <http://doi.org/10.48550/arxiv.1409.1556>.
- Wang Z, Bovik A C, Sheikh H R, et al, 2004. Image quality assessment: from error visibility to structural similarity [J]. IEEE Trans Image Process, 13(4): 600-612.
- Xue S F, Abdel-Hamid O, Jiang H, et al, 2014. Fast adaptation of deep neural network based on discriminant codes for speech recognition[J]. IEEE/ACM Trans Audio Speech Language Process, 22(12): 1713-1725.
- Ying M, Zhang W, Yu H, et al, 2014. An overview of the China Meteorological Administration tropical cyclone database[J]. J Atmos Oceanic Technol, 31(2): 287-301.
- Zeiler M D, Fergus R, 2014. Visualizing and understanding convolutional networks[M]// Fleet D, Pajdla T, Schiele B, et al. Computer Vision-ECCV 2014. Cham: Springer.
- Zhang R, Isola P, Efros A A, et al, 2018. The unreasonable effectiveness of deep features as a perceptual metric[C]//The 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City: IEEE.

(本文责编: 何晓欢)