文章编号：2096-1618(2023)06-0637-06

基于残差 Swin Transformer 的天气图像识别技术研究

张卓然[1]， 张 倩[2]， 宋 智[3]， 何 嘉[1]

(1. 成都信息工程大学计算机学院，四川 成都 610225；2. 活跃网络(成都)有限公司，四川 成都 610000；3. 四川省气象探测数据中心，四川 成都 610072)

摘要：人类活动经常受到天气条件的影响，基于图像的自动天气识别在实际应用中具有重要意义。然而现有方法均使用卷积神经网络，未能有效地利用图像的全局信息和像素点之间长距离的依赖关系，且识别的天气类型较少，识别精度较低。为解决这些问题，尝试将视觉 Transformer 应用到天气识别领域，同时提出一种基于残差 Swin Transformer 的模型，并使用先进的优化器 Ranger 来提高天气识别的正确率。该模型在包含 11 种天气现象的公开数据集 WEAPD 上进行验证，实验结果表明，其整体性能优于其他先进的识别网络，识别正确率达到93.6%，可为天气图像识别和天气预报研究提供参考。

关 键 词：天气现象；图像识别；深度学习；Swin Transformer

中图分类号：TP391.4 **文献标志码**：A

doi：10.16836/j.cnki.jcuit.2023.06.003

0 引言

从户外活动到工业生产，天气在人类活动中起着十分重要的作用。如在雾、雪、暴雨、沙尘暴等极端天气条件下，会出现模糊、光滑、湿润的道路情况，导致交通堵塞甚至造成交通事故。通过对天气现象的实时监控并结合交通信息，可以有效地避免这些情况的产生。此外，不同的天气现象也极大地影响着农业生产，准确地对天气现象进行识别，有助于保障农作物的生长。

在传统的天气识别中，主要是依靠各种传感器和人工采集结果来测量温度、湿度和天气状况。由于传感器成本普遍较高，又需要定期进行人工维护，导致不能大范围地布置传感器，从而使部分区域的天气预报不准确。

随着深度学习的发展和摄像头的普及，通过获得天气图像来进行天气识别将成为计算机视觉一个重要的应用。近年来，Lu 等[1]采集并创建了一个基于晴天和阴天的天气数据集，通过分别提取天空、阴影、反射、对比度和模糊 5 个天气特征，提出协作学习框架进行天气分类。Song 等[2]通过提取图片本身的特征并结合 K-NN 算法实现了对晴天、雨天、雾天和雪天的识别。上述方法均是采用机器学习，未能准确地学习天气图像的特征，对天气现象的识别效果不够理想。

自 2012 年 AlexNet[3]在 ImageNet 大规模视觉识别挑战赛取得成功以来，卷积神经网络(CNN)被广泛应用到天气现象识别领域。Lin 等[4]提出一种面向多类别天气识别的区域选择和并发模型的深度学习框架，但该方法性能开销过大、时效性差且平均准确率低。Zhao 等[5]提出一种基于 CNN-RNN 的多类别天气识别方法，但该方法需要大规模的数据集作为支持，且只能在高端 GPU 上进行训练，计算代价非常昂贵。Wang 等[6]提出一种基于轻量级卷积神经网络的天气识别方案，该方案虽然节省模型的内存开销，但降低了天气识别的准确度。Tan 等[7]通过一种三通道卷积神经网络对常见的 6 种天气现象进行有效识别，但只考虑了少数的天气类别。此外，Xiao 等[8]在 VGG16 的基础上提出一种针对 11 种天气现象的网络 MeteCNN，然而在个别类别识别率较低，没有达到很好的效果。

上述方法虽然运用了不同的 CNN 模型，但都忽视了图像的全局信息，随着网络的加深会产生梯度消失的问题，同时感受野在一定程度上受局限。2021 年研究人员提出一个适用于图像分类的 Transformer 模型 ViT[9]，并在 ImageNet 数据集获得了出色的结果。相比于 CNN 模型的平移不变性、局部敏感性和特征之间依赖关系差，Transformer 模型的自注意力(self-attention，SA)不受局部相互作用的限制，可以根据不同的任务目标学习最合适的归纳偏置[10]。

针对以上问题，本文尝试将视觉 Transformer 应用到天气现象识别领域，选取一种利用滑动窗口操作、具有层级设计的 Swin Transformer 网络结构[11]作为基础网络，并在其基础上加入残差结构和先进的优化器来增强模型的学习能力。实验结果表明，本文方法与其

收稿日期：2022-11-20

基金项目：四川省科技厅资助项目(2021005)；四川省重点实验室科技发展基金资助项目(2018-青年-11)

通信作者：何嘉. E-mail：hejia@cuit.edu.cn

他识别方法相比具有更高的识别精度。

1 网络结构

本文提出一种基于残差结构的 Swin-Transformer 网络模型用于天气图像的识别，该模型由图像块分割层、线性嵌入层、图像块合并层、残差窗口自注意力块（residual swin transformer block，RSTB）、归一化层（layer norm，LN）、自适应平均池化层、全连接层组成。网络结构图如图 1（a）所示。

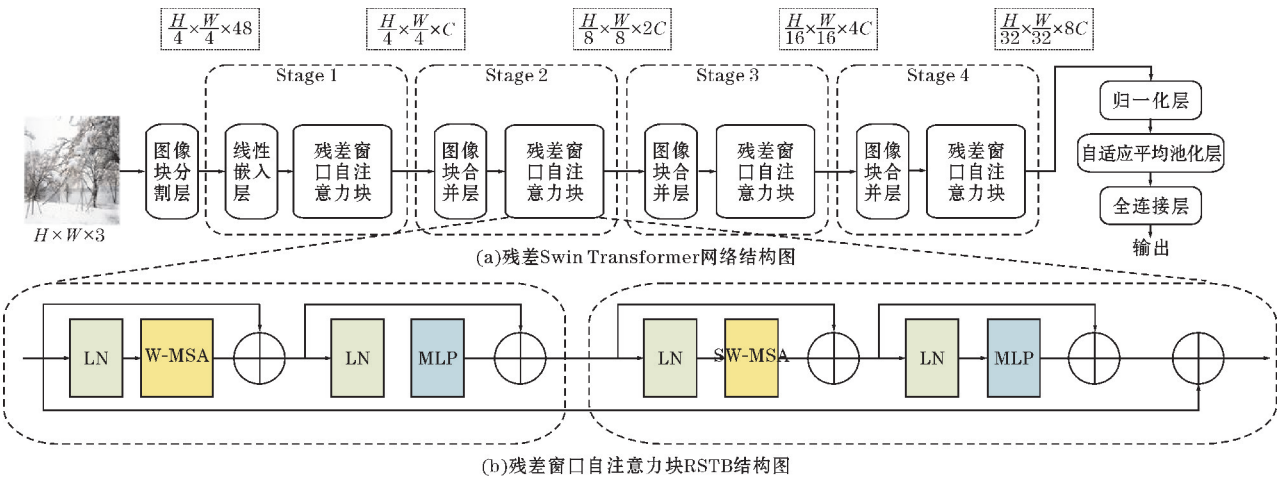

图 1 基于残差 Swin Transformer 的天气图像识别网络结构

模型首先通过图像块分割层将输入的 RGB 天气图片划分为多个不重叠的 patch，每个 patch 的大小为 4×4，特征维度为 4×4×3 = 48，经过 4 个阶段构建不同大小的特征图。其中只有第 1 阶段通过线性嵌入层将特征投影到任意维度 C，后 3 个阶段通过图像块合并层进行下采样，使特征图的宽和高缩小一半，通道数扩大 1 倍。接着，使用 RSTB 模块进行特征变换，这个模块由一个残差结构连接不同数目的窗口自注意力块（swin transformer block，STB）模块构成。最后，经过一个归一化层、一个自适应平均池化层和一个全连接层后输出天气图像的类别。

1.1 窗口自注意力块

窗口自注意力块由 1 个多头自注意力机制（multi-head self-attention，MSA）和 1 层多层感知机（multi-layer perceptron，MLP）组成，在 MSA 和 MLP 之间需要应用一个 LN 层，之后应用一个残差连接。MSA 拥有两种不同的结构：窗口多头自注意力机制（windows multi-head self-attention，W-MSA）；移动窗口多头自注意力机制（shifted windows multi-head self-attention，SW-MSA）。

如图 2（a）所示，MSA 在计算过程中需要对特征图中的每个像素进行运算，这种操作增加了整个模型的计算量。W-MSA 模块将特征图按照大小为 $M\times M$ 将大小为 $H\times W$ 的图像划分成多个窗口，然后单独对每个窗口进行自注意力计算，从而减少模型的计算量。MSA 和 W-MSA 的计算复杂度如下：

$$\Omega(\mathrm{MSA})=4HWC^2+2(HW)^2C$$

$$\Omega(\mathrm{W\text{-}MSA})=4HWC^2+2M^2HWC$$

由于单一使用 W-MSA 会出现不同窗口之间缺乏信息交互，这限制了模型的能力。因此，需要在 W-MSA 之后引入 SW-MSA 进行跨窗口连接，同时可以保持非重叠窗口的高效计算，SW-MSA 的跨窗口连接重组如图 2 所示。首先，将 W-MSA 的 4 个窗口通过滑动窗口变为 SW-MSA 的 9 个窗口，且这 9 个窗口大小不一致。为方便后续计算需要将 SW-MSA 的 9 个窗口重新整合成与原 W-MSA 窗口大小一致的 4 个窗口，再使用掩膜 Mask 隔绝不同区域的信息，最后将窗口还原。

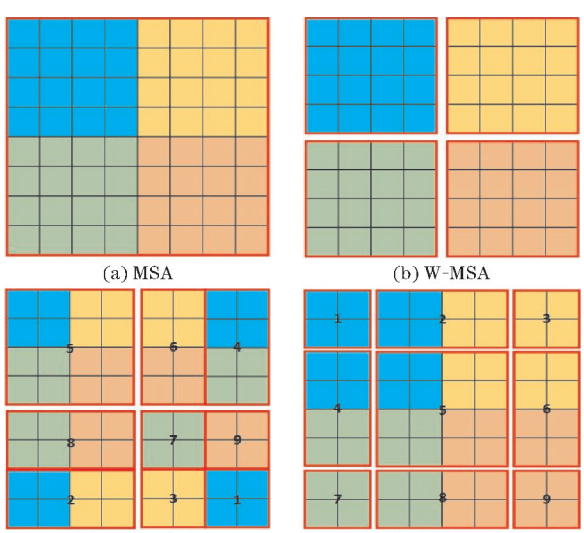

图 2 移动窗口重组示意图

故 STB 模块必须成对使用,第 1 个模块使用 W-MSA,第 2 个模块使用 SW-MSA,2 个连续的 STB 模块如图 2(b)所示。

1.2　残差窗口自注意力块

在天气识别任务中,为使网络获取更多有用信息,就需要对特征进行充分的提取和利用。在之前的研究中,天气识别任务通常使用卷积神经网络进行识别,而卷积神经网络通常使用残差连接[12]和稠密连接[13]提高网络性能,减少因模块堆叠、网络过深而造成的模型退化问题。由式(1)可知,W-MSA 计算复杂度与通道数成正比关系,而稠密连接使用过多会极大地增加通道数。故本文将在 STB 模块堆叠之间引入残差连接,并将改进后的堆叠 STB 模块称为 RSTB,其结构如图 1(b)所示。

2　损失函数和优化器

2.1　交叉熵损失函数

模型训练中需要使用损失函数来衡量模型的效果,进而改进模型参数的质量。本文使用的损失函数是交叉熵(CrossEntropy),主要用于度量两个概率分布之间的差异性,值总是大于等于 0。其值越接近于 0,就代表两个分布越相似。其数学公式:

$$\text{CrossEntropy}=-\frac{1}{N}\sum_{i}\sum_{c=1}^{M}y_{ic}\lg(p_{ic})$$

式中:N 代表样本总数,M 代表类别的数量;y_{ic} 代表符号函数(0 或 1),如果样本 i 的真实类别等于 c 则取 1,否则取 0;p_{ic} 代表样本 i 属于类别 c 的预测概率。

2.2　Ranger 优化器

Ranger 优化器[14]结合了 RAdam[15]和 Lookahead[16]两种优化器。RAdam 为优化器的初期训练提供最好的基础,用一个动态整流器调整 Adam 的自适应动量,并有效地提供一种基于当前数据集的自动训练预热机制,以确保训练迈出扎实的第一步。Lookahead 的灵感来自对深度神经网络损失曲面的理解,并为在整个训练过程中进行健壮和稳定的探索提供突破。本文选择的 Ranger 优化器将两者结合起来,这样可以获得更高的精度。

3　实验数据集

3.1　数据集介绍

模型训练和测试所使用的数据集来自文献[8]所开源的天气现象数据集。该数据集包含 6877 张具有代表性和独特的天气现象图像,每张图像的大小不定,被分为 11 种常见的天气现象,如图 3 所示。其中,数据集由露水(700 张)、雾/霾(855 张)、霜(475 张)、雨凇(639 张)、冰雹(592 张)、闪电(378 张)、雨(527 张)、彩虹(238 张)、雾凇(1160 张)、沙尘暴(692 张)和雪(621 张)组成。

为方便算法的训练和验证,在训练前先把图像顺序打乱,然后按 8∶1∶1 的比例分为训练集、验证集和测试集,它们之间不会出现重复的图像。

图 3　数据集展示

3.2　数据增强

由于视觉 Transformer 网络十分依赖数据量,为抑制模型学习过程中因数据量过少而造成的过拟合问题,本文算法在模型训练过程中对训练图像进行一系列的数据增强处理。包括:将天气图片调整为256×256后对图像进行随机裁剪到 224×224、水平翻转、水平移动和 Cutout 方法,原始图像经过数据增强后的效果如图 4 所示。采用这种方法图像会在形态上发生变化,模型在学习过程中不会使用任意两张相同的图像,这有利于抑制过拟合,提高模型的鲁棒性和泛化能力。

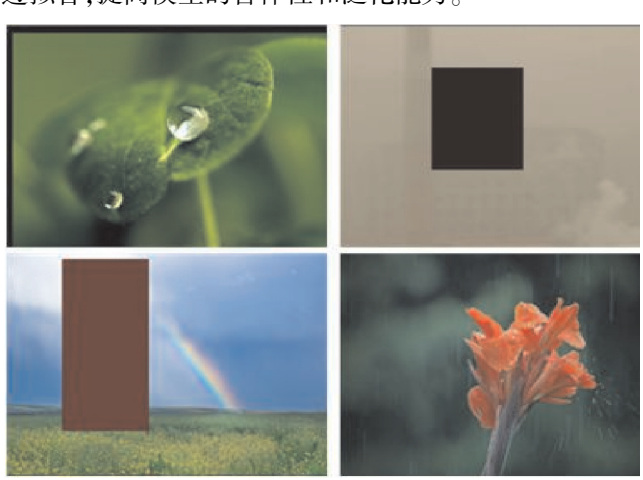

图 4　数据增强展示

4 实验

4.1 实验环境

实验在 DELL Precision 7920 塔式工作站开展,配置为:CPU: Intel(R) Xeon(R) Silver 4210,125 GB内存,GPU: NVIDIA TITAN RTX, 24 GB 显存, Ubuntu 16.04 LTS 64 位操作系统,Python3.8,PyTorch1.10.2。

4.2 实验细节

使用加入残差结构的 Swin Transformer 网络进行训练,采用迁移学习的方法,将 Swin Transformer 网络在 ImageNet 数据集上预训练的权重迁移到改进模型。为能正确识别出数据集中的所有类别,将模型最后的 FC 层改为 11 个输出,预训练模型的所有层均设置为可训练。训练过程使用 Ranger 优化器,初始学习率为 3×10^{-5},将单次训练所取样本数设为 64 张图片,共训练的迭代次数为 300 个。最终,选择验证集上正确率最高的模型进行测试实验,并重复实验 10 次,将 10 次的平均值作为模型最终结果。

4.3 模型评价指标

对于一个多分类任务,可以将每一种类别样本预测结果分为 4 种:TN(True Negative),表示把负样本正确地预测为负样本;FN(False Negative), 表示把正样本错误地预测为负样本;TP(True Positive),表示把正样本正确地预测为正样本;FP(False Positive),表示把负样本错误地预测为负样本。针对模型效果的评价,本文采用以下评价指标。

准确率 P(Precision)指在被判定为正的样本中,实际上为正样本所占的比例。公式如下:

$$P=\frac{TP}{TP+FP}$$

召回率 R(Recall)指原本为正的样本中被判定为正的样本所占的比例。公式如下:

$$R=\frac{TP}{TP+FN}$$

F_1 分数(F_1-measure)指准确率 P 和召回率 R 的调和平均值。公式如下:

$$F_1=\frac{2\times TP}{2\times TP+FP+FN}$$

正确率 Acc(Accuracy)指所有样本中分类结果正确的样本所占的比例。公式如下:

$$Acc=\frac{TP+TN}{TP+FP+TN+FN}$$

宏观准确率 ma_P(macro-average of Precision)指各类别准确率的平均值,公式如下:

$$ma\_P=\frac{1}{N}\sum_{i=1}^{N}P_i$$

宏观召回率 ma_R(macro-average of Recall),指各类别召回率的平均值,公式如下:

$$ma\_R=\frac{1}{N}\sum_{i=1}^{N}R_i$$

宏观 F_1 分数 ma_F_1(macro-average of F_1 score)指各类别 F_1 分数的平均值,公式如下:

$$ma\_F_1=2\times\frac{ma\_P\times ma\_R}{ma\_P+ma\_R}$$

上述评价指标的值都在 0 ~ 1。准确率、正确率、召回率和它们的宏观平均值越高,代表模型的分类性能越好。

4.4 实验结果分析

4.4.1 Swin Transformer 选择实验

由于 Swin Transformer 提出 4 个网络模型,即 Swin-T、Swin-S、Swin-B、Swin-L,为更好地建立适合本文数据集的分类模型并从模型部署和性能方面考虑,本文分别使用 Swin-T、Swin-S、Swin-B 来构建天气识别模型,并从中选出一个正确率最高的模型来做实验的基础模型。实验结果如表 1 所示,从表 1 可以看出 Swin-B 具有更好的精度优势。

表 1 Swin Transformer 模型选择实验

模型名称	Acc
Swin-T	0.883
Swin-S	0.899
Swin-B	0.904

4.4.2 残差模块嵌入实验

为验证在堆叠 STB 模块之间加入残差结构的有效性,对原 Swin-B 模型和加入残差结构的 Swin-B 模型进行对比实验,实验结果如表 2 所示。从表 2 可以看出,加入残差结构的 Swin-B 模型相对原模型的正确率显著提高。

表 2 残差模块的消融实验

模型名称	Acc
Swin-B	0.904
残差+Swin-B	0.926

4.4.3 优化器对比实验

由于优化器在深度学习过程中扮演着重要的角色,本文设计了基于残差 Swin-B 模型的 5 个优化器对比实验,使用的优化器分别为 SGD、RMSprop、Radam、

AdamW、Ranger,结果如表 3 所示。由表 3 可知,在残差 Swin-B 模型的基础上使用 Ranger 优化器可以使识别结果达到0.936的正确率,同时,进一步证明了该优化器的优点。

表 3　优化器对比实验

模型名称	Acc
残差 Swin-B + SGD	0.868
残差 Swin-B + RMSprop	0.921
残差 Swin-B + RAdam	0.920
残差 Swin-B + AdamW	0.926
残差 Swin-B +Ranger	0.936

4.4.4　本文方法与其他模型对比实验

表 4 显示了本文方法与一些主流模型的性能对比。在 WEAPD 数据集上,总体分类最佳结构是本文方法,正确率为 93.64%。同时,该模型的宏观准确率、宏观召回率和宏观 F_1 分数为 94% 左右,均超过其他模型。

表 4　本文方法与其他模型结果对比

模型名称	ma_P	ma_R	ma_F1	Acc
Vgg16	0.8776	0.8750	0.8750	0.8721
Vgg19	0.8531	0.8555	0.8531	0.8521
ResNet18	0.8868	0.8953	0.8901	0.8873
Resnet34	0.8932	0.8955	0.8928	0.8858
Effcientnet-B7	0.8819	0.8804	0.8805	0.8741
Wang et al[16]	0.8604	0.8522	0.8546	0.8521
MeteCNN[8]	0.9355	0.9331	0.9340	0.9268
本文	0.9443	0.9443	0.9423	0.9364

4.4.5　方法结果

本文方法的评价指标结果如表 5 所示,实现了对 11 种天气现象的有效分类,特别是对冰雹、雷暴、沙尘暴、雪这些极端天气能够准确识别。

表 5　本文方法的分类性能展示

类别	准确率 P	召回率 R	F_1 分数
dew	0.98	0.94	0.96
fog/smog	0.94	0.95	0.95
frost	0.98	0.83	0.89
glaze	0.83	0.92	0.87
hail	0.98	0.97	0.99
lightning	1.0	1.0	1.0
rain	0.93	0.94	0.93
rainbow	0.92	1.0	0.96
rime	0.88	0.92	0.90
sandstorm	0.97	0.99	0.98
snow	0.96	0.87	0.92
Average	0.94	0.94	0.94

为进一步证明残差 Swin Transformer 模型对天气现象分类的性能,使用混淆矩阵来说明分类精度,如图 5 所示。

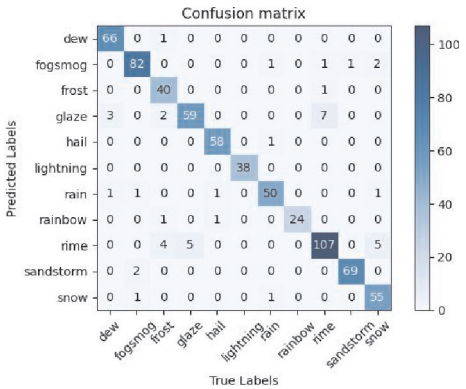

图 5　本文模型的混淆矩阵

5　结束语

本文尝试将视觉 Transformer 应用到天气图像识别领域,并提出一个基于残差结构的 Swin Transformer 模型,该模型可以很好地学习天气现象的特征。实验结果表明,本文所提出的模型对天气现象识别是有效的,且能避免因主观误差造成的错误,优于多数深度学习方法。虽然天气图像存在相似性、复杂性等问题,但总体来说,残差 Swin Transformer 模型的识别正确率达 93.6%,能够满足日常生活的需要。因此,本文所提出的模型可以广泛应用于天气现象的日常观测,也可为环境监测、农业和交通运输提供天气指导,特别在天气变化和预报方面。

由于本文是在 Swin-B 模型的基础上加入残差结构,Swin-B 网络模型参数较多,对计算机的性能要求较高,这对天气识别模型的时效性提出了挑战。后期研究中将考虑使用知识蒸馏、网络剪枝等思想对模型进行优化,在保证网络性能不变的情况下,提高时效性,实现一个轻量级天气识别模型。

参考文献:

[1] Lu C, Lin D, Jia J, et al. Two-class weather classifica-tion[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015:3718-3725.

[2] Song H, Chen Y, Gao Y. Weather condition recognition based on feature extraction and K-NN. Berlin: Springer, 2014:199-210.

[3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet

classification with deep convolutional neural networks[J]. Communications of the ACM,2017,60(6):84-90.

[4] Lin D,Lu C,Huang H,et al. RSCM:Region selection and concurrency model for multi-class weather recognition[J]. IEEE Transactions on Image Processing,2017,26(9):4154-4167.

[5] Zhao B,Li X,Lu X,et al. A CNN - RNN architecture for multi-label weather recognition[J]. Neurocomputing,2018,322: 47-57.

[6] Wang C,Liu P,Jia K,et al. Identification of weather phenomena based on lightweight convolutional neural networks[J]. CMC-COMPUTERS MATERIALS & CONTINUA,2020,64(3):2043-2055.

[7] Tan L,Xuan D,Xia J,et al. Weather Recognition Based on 3C-CNN[J]. KSII Transactions on Internet and Information Systems(TIIS),2020,14(8): 3567-3582.

[8] Xiao H,Zhang F,Shen Z,et al. Classification of Weather Phenomenon From Images by Using Deep Convolutional Neural Network[J]. Earth and Space Science,2021,8(5):e2020EA001604.

[9] Dosovitskiy A,Beyer L,Kolesnikov A,et al. An image is worth 16x16 words:Transformers for image recognition at scale[C]. International Conference on Learning Representations,2021.

[10] 刘文婷,卢新明. 基于计算机视觉的 Transformer 研究进展[J]. 计算机工程与应用,2022,58(6):1-16.

[11] Liu Z,Lin Y,Cao Y,et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10012-10022.

[12] He K,Zhang X,Ren S,et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016:770-778.

[13] Huang G,Liu Z,Laurens V,et al. Densely Connected Convolutional Networks[C]. IEEE Computer Society. IEEE Computer Society,2016.

[14] Wright L, Demeure N. Ranger21: a synergistic deep learning optimizer[J]. arXiv preprint arXiv: 2106.13731,2021.

[15] Liu L,Jiang H,He P,et al. On the variance of the adaptive learning rate and beyond[C]. ICLR,2020.

[16] Zhang M,Lucas J,Ba J,et al. Lookahead optimizer:k steps forward,1 step back[J]. Advances in neural information processing systems,2019,32.

[17] Wang C,Liu P,Jia K,et al. Identification of weather phenomena based on lightweight convolutional neural networks[J]. CMC-COMPUTERS MATERIALS & CONTINUA,2020,64(3):2043-2055.

Research on Weather Image Recognition based on Residual Swin Transformer

ZHANG Zhuoran[1], ZHANG Qian[2], SONG Zhi[3], HE Jia[1]

(1. College of Computer Science,Chengdu University of Information Technology,Chengdu 610225,China;2. Active Network(Chengdu), Ltd.,Chengdu 610000 China;3. Sichuan Meteorological Detection Data Center,Chengdu 610072,China)

Abstract:Human activities are often affected by weather conditions, and automatic weather recognition-based image is of great importance in practical applications. However, existing methods all use convolutional neural networks, which fail to effectively utilize the global information of images and the long-distance dependency between pixel points, and recognize fewer weather types with low recognition accuracy. To solve these problems, we try to apply the visual Transformer to the field of weather recognition, and also propose a model based on the residual Swin Transformer and use the advanced optimizer Ranger to improve the weather recognition rate. The model is validated on WEAPD, a publicly available dataset containing 11 weather phenomena, and the results show that its overall performance is better than other advanced recognition networks, with a 93.6% correct recognition rate. It can benefit the research of weather image recognition and weather forecasting.

Keywords:weather phenomena; image recognition; deep learning; Swin Transformer