

稠密融合卷积神经网络的多模态地基云状分类^{*}

刘爽 许依琳 张重

(天津师范大学 电子与通信工程学院 天津 300387)

摘要:为了解决现有的地基云状分类方法对多模态信息利用不充分的问题,有效融合地基云样本的视觉特征与多模态特征,提出稠密融合卷积神经网络的多模态地基云状分类方法。稠密融合卷积神经网络采用卷积神经网络作为视觉子网络提取地基云图像的视觉特征,采用多模态子网络提取多模态特征,在网络内部加入了5个稠密融合模块,用于充分融合视觉特征与多模态特征,稠密融合模块在不改变原有网络结构的同时,能够独立地加入到子网络中,具有较大的灵活性。在多模态地基云公开数据集MGCD上的实验达到了89.14%的分类精度,验证了所提出的稠密融合卷积神经网络在地基云状分类任务中的有效性。

关键词:卷积神经网络;多模态地基云状分类;稠密融合

中图分类号:TP391 **文献标识码:**A **国家标准学科分类代码:**510

Multi-modal ground-based cloud classification based on dense fusion convolutional neural network

Liu Shuang Xu Yilin Zhang Zhong

(College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China)

Abstract: In order to solve the issue that the existing ground-based cloud classification methods can not make full use of multi-modal information, we propose the dense fusion convolutional neural network (DFCNN) for multi-modal ground-based cloud classification to effectively integrate the visual features and the multi-modal features of ground-based cloud samples. The DFCNN utilizes convolution neural network as the visual subnet to extract visual features and adopts the multi-modal subnet to extract multi-modal features of cloud samples. There are five dense fusion modules (DFM) in the DFCNN and they are employed to fully fuse visual features and multi-modal features. The DFM could be injected into the subnet independently without changing the original network structure, and therefore it possesses great flexibility. The DFCNN achieves the classification accuracy of 89.14% on the public multi-modal ground-based cloud dataset MGCD, which verifies the effectiveness of the proposed DFCNN for the ground-based cloud classification task.

Keywords: convolutional neural network; multi-modal ground-based cloud classification; dense fusion

0 引言

云对于全球范围内的水文循环、能量均衡等具有重要意义。精准的云观测技术对天气预报、降水估计、光学遥感应用等领域至关重要。云观测技术主要有卫星遥感观测和地基遥感观测^[1]。其中,卫星遥感观测的特点是观测范围广,可以收集较大范围内的云数据,适用于大尺度气候研究。地基遥感观测通过地面特定位置的成像设备进行连续的云观测,得到局部区域云信息,适用于局部区域的气候研究。云观测主要包含3个要素:云量、云底高和云状^[2]。云

状能即时反映当地的大气状况,因此对云状分类的研究是云观测研究中重要的组成部分。本文主要关注局部区域内的云状,因此在收集数据时使用地基遥感观测方法,并对地基云状分类进行研究。

目前,卷积神经网络(convolutional neural network, CNN)已成功应用于诸多领域^[3-5],并在地基云状分类任务上取得优越性能^[6]。相较于其他手工构建特征的方法,CNN的优势在于可以从原始数据中学习较复杂的特征变换。在地基云状分类任务中,大部分方法将地基云图像作为CNN的输入,学习得到地基云图像的特征表示。

收稿日期:2021-07-24

^{*} 基金项目:国家自然科学基金(62171321)、天津市自然科学基金重点项目(20JCZDJC00180, 19JCZDJC31500)、天津师范大学教学改革项目(重点项目)(JGZD01220014)资助

然而,仅从地基云图像中很难提取完备的特征表示,因为云的形成、变化与温度、湿度、大气压强、风速等气象因素息息相关,如温度和大气压强影响云的形成,风速影响云的形状。在地基云状分类任务中,将地基云图像称为视觉信息,将其他气象信息称为多模态信息。将两种信息融合,可以获得完较备的地基云特征表示,有利于地基云状分类。

现有多模态地基云状分类方法通常利用多个子网络,分别提取视觉特征和多模态特征,并利用在网络末端的融合层对各子网络提取的特征进行融合,该融合方法称为末端融合。末端融合方法对信息融合不充分,不利于发挥多模态特征在分类任务中的优势。

针对末端融合的缺点,本文提出稠密融合卷积神经网络(dense fusion convolutional neural network, DFCNN)的多模态地基云状分类方法,进而实现充分融合视觉信息和多模态信息。DFCNN 主要包含用于学习视觉特征和多模态特征的两个子网络,以及稠密融合模块(dense fusion module, DFM)。相比于末端融合方法,DFCNN 的优点体现在两个方面:首先,DFCNN 可以实现在多个子网络中间层的稠密融合,进而使视觉特征和多模态特征充分融合;其次,DFM 可以直接加到子网络的各个层次结构中,具有很大的灵活性。在 MGCD 公开数据集上的实验结果证明了 DFCNN 在地基云状分类任务中的有效性。

1 相关研究

由于 CNN 在多个计算机视觉任务中获得了优越的性能,因此它被研究者们引入到地基云图像分类任务中。基于 CNN 的方法主要有两类:一类是直接利用 CNN 模型提取特征并作为地基云图像的表示。例如,Shi 等^[7]利用 CNN 的浅层提取地基云的局部纹理特征,准确地描述了地基云图像。Zhang 等^[8]提出了 CloudNet 模型,该方法能够有效学习地基云图像的纹理、结构和形状信息。另一类方法是对 CNN 模型提取的特征做进一步处理以获得最终的表示。例如,Ye 等^[9]利用费舍尔向量对 CNN 提取的特征进行编码,进而增加不同类别地基云图像间的差异。Zhang 等^[10]提出深度局部二值模式方法,从 CNN 模型中提取卷积激活图,然后在卷积激活图上利用局部二值模式对地基云图像进行编码。

近年来,为获得完备的特征表示,在地基云状分类研究中引入气象信息,如温度、湿度、大气压强、风速等,即多模态信息。地基云样本的视觉特征与多模态特征的融合方式是地基云状分类任务的关键技术。Liu 等^[11]分别用两个独立的网络提取视觉特征和多模态特征,再将其串联起来作为地基云样本的特征表示,这种方法初步实现了视觉信息与多模态信息的融合。Liu 等^[12]提出联合融合卷积神经网络(joint fusion convolutional neural network, JFCNN),利用 CNN 提取地基云图像的深度视觉特征,同时利用多层感知机提取多模态信息的深度特征,之后将两种特征在深

度神经网络的最后一层进行融合。除此之外,Li 等^[13]提出了深度张量融合网络(deep tensor fusion network, DTFN),在深度神经网络的最后一层使视觉特征与多模态特征在张量级别上融合,从而有效利用地基云图像的空间信息。

2 稠密融合卷积神经网络

2.1 网络结构

现有多模态地基云状分类方法,常利用末端融合方法学习地基云样本的特征表示。现有方法采用末端融合有两个主要原因:1)不同网络的中间层特征往往空间尺寸不统一,因此在中间层融合难度较大;2)网络框架常包含多个预训练的子网络,对中间层的特征进行融合时,往往会改变子网络的结构,使预训练变得复杂,甚至需要重新训练子网络。末端融合方法实现简单,但对多模态信息的融合多模态不够充分。本文提出 DFCNN 能够实现中间层的稠密融合,在不改变子网络结构的同时,实现多模态信息的充分利用。

DFCNN 的整体框架如图 1 所示。它主要由视觉子网络、多模态子网络、DFM、融合层、全连接层和损失函数等 6 个部分组成。其中,视觉子网络采用 ResNet-50^[14]的部分结构学习地基云图像的视觉特征。在视觉子网络中,将输出相同尺寸卷积激活图的卷积层归为一类,由浅到深分别记为 Cov1、Cov2-x、Cov3-x、Cov4-x 和 Cov5-x。其中 Cov2-x 到 Cov5-x 是由多个卷积层组成的卷积模块。多模态子网络由 5 个全连接层组成,用于学习地基云的多模态特征。

将地基云图像与多模态信息分别输入相应的子网络,并将各层子网络输出的卷积激活图及特征向量作为各层 DFM 的输入。由 DFM 实现视觉特征与多模态特征在中间层的稠密融合。DFM 的输出是一个卷积激活图,它的通道数 C 与输入当前模块的卷积激活图 f_{cov} 的通道数 C 相同。

如图 1 所示,以第 1 次稠密融合过程为例,将 DFM 输出的特征向量 f_{DFM} 与 f_{cov} 按对应通道相乘,得到卷积激活图 f'_{cov} :

$$f'_{cov} = f_{cov} \odot f_{DFM} \quad (1)$$

其中,符号 $\odot$ 表示按通道相乘。在视觉子网络中,将式(1)得到的卷积激活图作为下一个卷积模块的输入,在多模态子网络中,将当前层输出的特征向量作为下一个全连接层的输入。之后将视觉子网络输出的卷积激活图经过全局平均池化(global average pooling, GAP)和全连接层后与多模态子网络的输出共同作为融合层的输入。其中,GAP 表示对每个通道的卷积激活图中的像素取平均值,因此将 C 个通道的卷积激活图转化为 C 维向量。

在融合层中将两个子网络的输出特征串联,之后输入到全连接层 FC6。FC6 神经元的数量与云的类别数量 K 一致。之后利用 Softmax 函数,将 K 个预测值变为 $0 \sim 1$

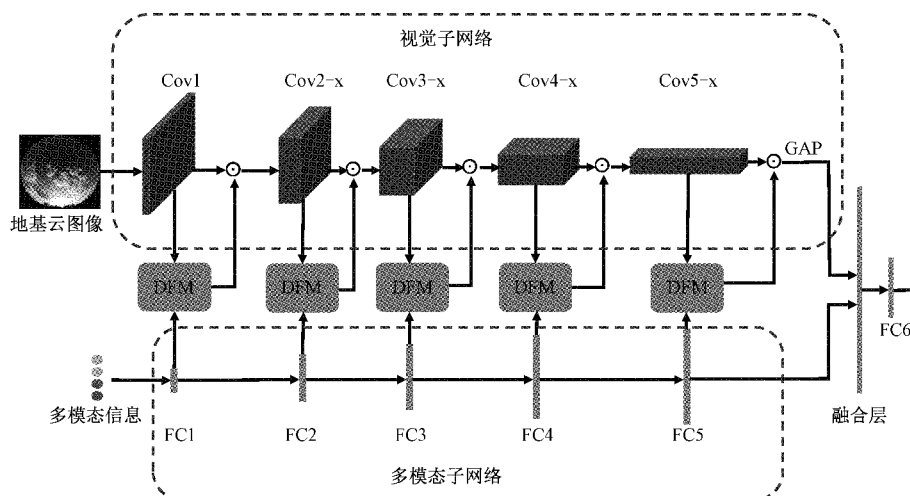


图 1 DFCNN 整体框架

之间的概率分布。Softmax 运算如下:

$$y_k = \frac{e^{x_k}}{\sum_{i=1}^K e^{x_i}} \quad (2)$$

其中, x_k 为 FC6 中第 k 个神经元的输出, K 为云状类型的数量, y_k 为属于第 k 类云的预测值。之后, 用交叉熵损失函数对 DFCNN 训练, 公式如下:

$$L = - \sum_{k=1}^K q_k \log y_k \quad (3)$$

其中, q_k 是地基云样本的真实概率, 当 k 与真实标签相同时, $q_k = 1$, 反之, $q_k = 0$ 。

2.2 稠密融合模块

现有多模态特征融合方法常采用末端融合, 导致多模态特征融合不够充分。DFM 能够实现视觉特征与多模态特征在中间层的融合, 不受限于中间层卷积激活图的大小。DFM 能够作为独立的模块添加到视觉子网络和多模态子网络间, 实现两种特征的稠密融合, 同时不需要改变原有网络的结构, 因此具有较大的灵活性。DFM 结构如图 2 所示。

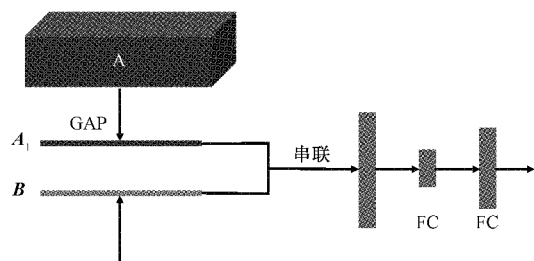


图 2 DFM 结构

在 DFM 中, 由视觉子网络输出的卷积激活图 A 经过 GAP 得到特征向量 A_1 。由多模态子网络输出特征向量 B , 将 A_1 与 B 串联, 之后将其通过两个全连接层, 在两个全连接层后分别加入非线性激活函数 ELU 和 Sigmoid。ELU 函数公式如下:

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases} \quad (4)$$

其中, α 是常数参数, 在实验中取 $\alpha = 1.0$ 。ELU 函数在神经元输入为负数时仍有输出, 具有一定的抗干扰能力。Sigmoid 函数公式如下:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (5)$$

非线性激活函数用于调节各神经元的输出分布, 在训练时加速收敛, 并避免发生梯度消失等问题。

3 实验与结果分析

3.1 MGCD 数据集

MGCD^[15] 由 8 000 个多模态地基云样本组成, 共分为 7 类, 每个地基云样本包括 1 幅地基云图像和对应的 4 种多模态信息, 地基云样本如图 3 所示。各类地基云类别及数量如表 1 所示。MGCD 中的地基云图像由鱼镜头相机捕获, 分辨率为 $1\,024 \times 1\,024$ 。每个地基云样本的多模态信息包含温度、湿度、大气压强、风速 4 种, 由小型气象站收集, 以矢量形式存储。MGCD 划分为训练集和测试集两部分, 各包含 4 000 个样本。

3.2 实验设置

在实验中, 首先对地基云图像和多模态信息进行预处理。具体方法是: 将地基云图像的大小调整为 224×224 , 对像素值进行归一化处理, 并先后在垂直方向和水平方向上对地基云图像做镜像翻转, 翻转概率为 0.5, 之后, 在平面中 $-25^\circ \sim 25^\circ$ 的范围内对地基云图像做平面旋转。将多模态信息归一化到 $[0, 255]$ 的范围内。使用随机梯度下降法对提出的 DFCNN 进行训练, 设置小批量训练样本数量为 32, 迭代次数(Epoch)为 50, 初始学习率为 4.7×10^{-4} , 并分别在第 15、35 个 Epoch 时进行学习率衰减, 衰减率为 0.2。通过精度(Accuracy)指标对模型进行评估。精度定义如下:

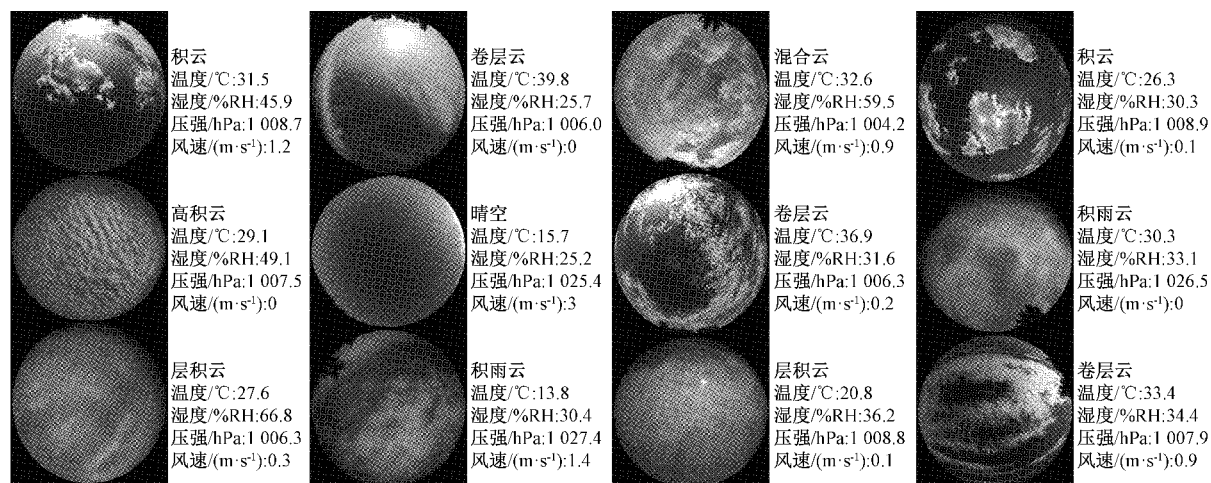


图 3 MGCD 中的云样本

表 1 MGCD 数据集中地基云样本的类别及数量

地基云类别	数量
积云	1 438
高积云和卷积云	731
卷云和卷层云	1 323
晴空	1 338
层积云、层云和高层云	963
积雨云和雨层云	1 187
混合云	1 020

$$Accuracy = \frac{\sum_{k=1}^K m_k}{N} \quad (6)$$

其中, m_k 是第 k 类地基云样本中被正确分类的数量, N 是测试集中的样本数量。精度越高, 表示分类效果越好。

3.3 实验结果

将 DFCNN 与其他方法在 MGCD 数据集上进行比较, 这些方法包括手工构建特征的方法和基于 CNN 的方法。手工构建的方法包括词包模型 (bag-of-words, BoW)、金字塔词包模型 (pyramid BoW, PBoW)、局部二值模式 (local binary pattern, LBP) 和完备局部二值模式 (completed LBP, CLBP); 基于 CNN 的方法包括 ResNet-50、CloudNet、JFCNN、DTFN。对比结果如表 2 所示。

表 2 实验结果对比

算法	Accuracy	算法	Accuracy
BoW	66.15	BoW+MI	67.20
PBoW	66.13	PBoW+MI	67.15
LBP	50.20	LBP+MI	50.53
CLBP	69.18	CLBP+MI	69.68
ResNet-50	83.15	ResNet-50+MI	84.48
CloudNet	79.93	JFCNN	84.13
—	—	DTFN	87.90
—	—	DFCNN	89.14

从表 2 中, 可以观察得到如下 3 个结论: 1) 将手工构建特征的方法与基于 CNN 的方法进行对比, 可以得出, 基于 CNN 的地基云状分类方法可以得到更高的分类精度。这是因为 CNN 的多层级结构可以学习得到高度复杂的非线性变换, 进而学习得到判别性较强的特征。2) 将表格左右两侧进行对比, 其中左侧是仅利用地基云图像的云状分类方法, 右侧是加入多模态信息 (MI) 的地基云状分类方法。通过对比可以得出, 加入多模态信息的方法得到了更高的分类精度, 这说明加入多模态信息后, 可以得到更加完备的特征表示, 因此将多模态特征与视觉特征融合有利于提高云状分类精度。3) JFCNN、DTFN 和 DFCNN 均是在卷积神经网络框架下融合视觉特征和多模态特征的方法, 其中 JFCNN 和 DTFN 采用末端融合, 本文提出的 DFCNN 获得了更高的分类精度。这说明融合视觉特征与多模态特征的具体方式以及融合程度会影响分类效果, 相较于其他融合程度较低的末端融合方法, 提出的稠密融合能获得更加完备的地基云样本的特征表示, 充分发挥多模态信息的优势。

4 结 论

本文提出了一种将地基云图像与多模态信息充分融合后用于地基云状分类的方法, 在 DFCNN 网络中设计了视觉子网络和多模态子网络分别用于提取视觉特征和多模态特征, 设计了 DFM 实现特征在网络中间层的稠密融合。最后在融合层中将上述两种特征串联, 得到地基云样本的特征表示。实验结果证明, DFCNN 克服了常用的末端融合方法中无法对多模态信息充分利用的困难, 能够有效地融合地基云图像与多模态信息, 获得完备的地基云样本的特征表示, 在地基云状分类任务中获得较好的性能。

参考文献

- [1] TAPAKIS R, CHARALAMBIDES A. Equipment and methodologies for cloud detection and classification;

- A review[J]. Sol. Energy, 2013, 95: 392-430.
- [2] ZHONG B, CHEN W H, WU S L, et al. A cloud detection method based on relationship between objects of cloud and cloud-shadow for Chinese moderate to high resolution satellite imagery[J]. IEEE J-Stars, 2017, 10(11): 4898-4908.
- [3] 陈星. 基于卷积神经网络算法的飞机发电机故障诊断研究[J]. 国外电子测量技术, 2020, 39(11): 57-60.
- [4] 唐波, 陈慎慎. 基于深度卷积神经网络的轴承故障诊断方法[J]. 电子测量与仪器学报, 2020, 34(3): 88-93.
- [5] 蓝金辉, 王迪, 申小盼. 卷积神经网络在视觉图像检测的研究进展[J]. 仪器仪表学报, 2020, 41(4): 167-182.
- [6] YE L, CAO Z G, XIAO Y, et al. Supervised fine-grained cloud detection and recognition in whole-sky images[J]. IEEE T Geosci Remote, 2019, 57(10): 7972-7985.
- [7] SHI C Z, WANG C H, WANG Y, et al. Deep convolutional activations-based features for ground-based cloud classification[J]. IEEE Geosci Remote S, 2017, 14(6): 816-820.
- [8] ZHANG J L, LIU P, ZHANG F, et al. CloudNet: Ground-based cloud classification with deep convolutional neural network[J]. Geophys Res Lett, 2018, 45(16): 8665-8672.
- [9] YE L, CAO Z G, XIAO Y. DeepCloud: ground-based cloud image categorization using deep convolutional features[J]. IEEE T Geosci Remote, 2017, 55(10): 5729-5740.
- [10] ZHANG Z, DUAN L L, LIU S. Multi-view ground-based cloud recognition by transferring deep visual information[J]. Appl Sci-Basel, 2018, 8(5): 748.
- [11] LIU S, LI M. Deep multimodal fusion for ground-based cloud classification in weather station networks [J]. Eurasip J Wirel Comm, 2018, 2018(1): 48.
- [12] LIU S, LI M, ZHANG Z, et al. Multimodal ground-based cloud classification using joint fusion convolutional neural network[J]. Remote Sens-Basel, 2018, 10(6): 822.
- [13] LI M, LIU S, ZHANG Z. Deep tensor fusion network for multimodal ground-based cloud classification in weather station networks[J]. Ad Hoc Netw, 2020, 96: 101991.
- [14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016: 583-592.
- [15] LIU S, LI M, ZHANG Z, et al. Multi-evidence and multi-modal fusion network for ground-based cloud recognition[J]. Remote Sens-Basel, 2020, 12(3): 464.

作者简介

刘爽, 博士, 副教授, 主要研究方向为计算机视觉、模式识别、地基云图分析等。

E-mail: s.liu@tjnu.edu.cn

许依琳, 硕士研究生, 主要研究方向为计算机视觉、模式识别、地基云图分析等。

E-mail: xuyilin416@163.com

张重(通信作者), 博士, 教授, 主要研究方向为计算机视觉、模式识别、机器学习等。

E-mail: zhangz@tjnu.edu.cn