

基于EfficientNet的星系形态分类研究*

艾霖嫔^{1†} 徐权峰¹ 杜利婷¹ 许婷婷² 高献军¹ 李广平¹ 周卫红^{1,3‡}

(1 云南民族大学数学与计算机科学学院 昆明 650500)

(2 广州大学天体物理中心 广州 510006)

(3 中国科学院天体结构与演化重点实验室 昆明 650011)

摘要 星系的结构和形态能够反映星系自身的物理性质,其形态的分类是后续分析研究的一个重要环节. EfficientNet模型使用复合系数对深度网络模型的深度、宽度、输入图像分辨率进行更加结构化的统一缩放,是一种新的深度网络优化扩展方法. 将该模型应用于星系数据形态的分类研究中,结果表明基于EfficientNet-B5模型的平均准确率、精确率、召回率以及F1分数(精确率与召回率的调和平均数)都在96.6%以上,与残差网络(Residual network, ResNet)中ResNet-26模型的分类结果相比有较大的提升. 实验结果证明EfficientNet的深度网络优化扩展方法可行且有效,可应用于星系的形态分类.

关键词 方法: 数据分析, 技术: 图像处理, 星系: 普通

中图分类号: P157; **文献标识码**: A

1 引言

星系是由恒星、恒星遗骸、暗物质和星际气体等物质组成的引力束缚系统. 在星系天文学的传统研究中,人们通常根据它的视觉形态来对星系进行分类. 星系的形态反映了星系的综合物理特征,因而与星系的形成与演化过程有着密切的联系. 比如,研究表明,椭圆形态的星系更可能由星系的并合过程形成^[1],因此,对星系按形态特征进行分类,是星系天文学研究中的一个常规且重要的环节. 随着大型巡天项目的巡天深度和探测效率不断提升,巡天探测中得到的星系数目也在不断增加,甚至达到了近亿量级,因此,对常规的数据处理方法提出了新的挑战.

早期,由于观测技术以及观测仪器的限制,星系的观测数目也有限,用人眼便可完成对星系的形

态分类,即目视法. 著名的目视分类系统有哈勃系统^[2]、德沃古勒系统^[3]、叶凯士系统^[4]等. 在这些分类中哈勃音叉图是目前使用最为广泛的分类方法. 但是,目视分类系统只适合星系图像数据较少的情形,且有着较强的主观性. 随着数字技术的发展,一些大天区巡天项目获得大量的星系图像. 此时研究者已不能再依靠目视法来对星系的形态进行大样本的分类了,与此同时一些基于计算机的星系形态的替代分类方法出现了,其中包括了图像轮廓的参数化拟合以及图像特征的非参数测量两种常用方法^[5]. 对于参数化的轮廓拟合来说,该方法对星系的二维表面亮度轮廓进行一维拟合,因此会丢失星系的旋臂在切向方向的图像特征. 此外,对高红移星系而言,由于图像的信噪比一般比较低,往往并不能对星系的图像进行很好的一维参数化

2021-08-30收到原稿, 2021-11-16收到修改稿

*国家自然科学基金项目(61561053)资助

[†]972734136@qq.com

[‡]ynzwh@163.com

轮廓拟合, 特别在进行核球和盘成份分解等复杂任务时. 为了解决参数化拟合对低信噪比图像或者不规则星系图像特征描述不足的缺点, 学者们提出了非参数化的测量方法. 比如星系图像中最为常用的非参数化特征包括了星系的参数如C (聚集度指数)^[6]、A (非对称指数)^[7]、G (基尼系数)^[8]等形态特征参数. 范海英^[9]结合了参数化与非参数化星系形态特征的测量方法来计算星系的形态参数, 发现他们研究的COSMOS (Cosmic Evolution Survey)天区中红移 $z \sim 3$ 的456颗LBG (Lyman Break Galaxy)星系的Sérsic指数(Sérsic指数大于等于2的时候, 代表早型星系, 否则为晚型星系)均较小, 属于晚型星系.

随着斯隆数字巡天(Sloan Digital Sky Survey, SDSS)、暗能量巡天(The Dark Energy Survey, DES)等项目巡天深度和探测效率的不断提升, 星系图像数据呈爆炸式增长, 传统的目视分类系统和参数化的轮廓拟合都无法满足如此大规模的星系图像数据处理需求. 对于非参数化的图像特征参数来说, 虽然其测量较为简单, 但是并不能有效地替代星系的形态特征这一经典描述. 幸运的是, 研究人员开展了一项名为星系动物园(Galaxy Zoo)的公众科学计划(详见3.1节). 这项科学计划利用了公众的力量对包含几十万颗星系图像的样本来进行形态分类, 从而提供了一个很好的可以利用机器学习来对星系图像的形态特征进行自动分类的训练样本. Hocking等^[10]提出了一种无监督的机器学习技术, 与传统的无监督机器学习方法不同, 该方法不经过任何形式的预训练便可分离早期和晚期类型的星系, 将基于机器的分类与星系动物园的人工分类结果进行比较, 证明了人工和机器分类之间的一致性; Gauci等^[11]基于机器学习开发了不同的分类模型应用于星系形态分类, 进行了比较分析包括Classification And Regression Tree (CART)、C4.5 (可产生决策树用作分类)、随机森林和模糊逻辑算法. 实验表明, 使用这几种算法的训练时间都不长, 其中当随机森林的树为50棵的时候, 其分类精确率在几种分类算法中是最高的; Barchi等^[12]提出了两种不同的星系形态学方法: 一种基于非参数形态学 and 传统机器学习算法,

另一种基于深度学习. 实验结果是深度和传统的机器学习方法准确率都超过了94.5%; Cheng等^[13]探索了无监督机器学习在星系形态分类中的应用, 将特征提取与向量量化变分自动编码器(Vector Quantized Variational AutoEncoder, VQ-VAE)和层次聚类(Hierarchical Clustering, HC)相结合, 在早期类型星系和晚期类型星系上均达到了87%的准确率.

深度学习方面戴加明等^[14]通过改进残差网络(Residual network, ResNet)的残差单元, 得到一个名为ResNet-26的深度残差网络模型, 对星系的形态特征进行提取, 以95.12%的精度验证了该模型对星系形态分类的有效性; Chen^[15]基于Galaxy Zoo的Kaggle公共数据集(详见3.1节), 设计了一种用于星系形态图像分类的深度卷积神经网络. 该网络包含8个卷积层、5个最大池化层、一个Flatten层、一个包含150个神经元的全连接层, 最终输出了每个图像类别的概率分布. 结果表明该模型在验证集上余弦相似度(Cosine Similarity, CS)达到了-0.8652 (越接近-1意味着预测类别和实际类别越接近), 证实了该模型对于星系形态分类是非常有效的; Cavanagh等^[16]开发了一种新的卷积神经网络架构, 以3类(椭圆、透镜状、螺旋形)和4类(加不规则)模式对星系形态进行分类, 在3类和4类分类方面都优于现有模型, 总体分类准确率分别为83%和81%. 并比较了所有4个类别之间双向/二进制分类的准确性, 表明椭圆和螺旋最容易区分, 而螺旋和不规则最难区分.

近年来的研究成果表明, 深度学习网络逐渐成为海量天文数据处理分析的有效工具和方法, 如何设计构建、扩展以及优化高效的深度网络模型, 是深度学习方法应用中的一个主要难点问题.

2 EfficientNet

以往对于深度网络模型的优化通常从增加卷积神经网络的深度、宽度以及增加输入图像的分辨率其中的一个维度着手, 但这些模型扩展方法具有任意性, 还需要手动调参, 这不仅增加了网络设计的难度, 还不能保证网络性能最优.

2019年谷歌大脑的研究人员Tan等^[17]提出了

一种新的针对网络模型的深度、宽度、输入图像分辨率进行统一缩放的方法. 与以往对模型扩展时任意对网络深度、宽度以及输入图像分辨率扩展不同, 该方法使用复合系数对网络模型的深度、宽度、输入图像分辨率进行更加结构化的统一缩放, 谷歌大脑基于该缩放方法提出一系列的卷积神经网络模型, 命名为EfficientNet.

扩展网络的性能取决于基线网络, 因此一个构架简单、干净、容易扩展、泛化的基线网络至关重要. Tan等^[17]通过自动机器学习(Automated Machine Learning, AutoML) Mobile Neural Architecture Search (MNAS)框架进行神经网络架构搜索, 得到一个基线模型命名为EfficientNet-B0. EfficientNet-B0使用移动翻转瓶颈卷积(Mobile inverted Bottleneck Convolution, MBConv), 继续对EfficientNet-B0进行缩放得到一系列扩展模型—EfficientNet, 图1为EfficientNet-B0结构图. 其中包含移动翻转瓶颈卷积模块16个、卷积层2个、全局平均池化层以及分类层各1个. 进行复合缩放首先需要进行网格搜索, 探索在固定的计算机内存资源条件下EfficientNet-B0的网络深度、宽度、输入图像分辨率缩放之间的联系. 发现常规卷积运算的每秒浮点计算次数与网络深度、宽度的平方、输入图像分辨率的平方成正比, 即网络深度翻倍会使每秒浮点计算次数翻倍, 而网络宽度或输入图像分辨率翻倍则会使每秒浮点计算次数增加4倍. 如下式所示:

$$\left\{ \begin{array}{l} \text{depth} : d = \alpha^\phi, \\ \text{width} : w = \beta^\phi, \\ \text{resolution} : r = \gamma^\phi, \\ \text{s.t. } \alpha\beta^2\gamma^2 \approx 2, \\ \alpha \geq 1, \beta \geq 1, \gamma \geq 1. \end{array} \right. \quad (1)$$

其中 d 、 w 、 r 分别表示网络深度、宽度、输入图像分辨率, ϕ 是一个用户指定的系数, 它规定了有多少资源可以用于模型缩放, α 、 β 、 γ 是常数, 由网格搜索来确定, 分别指定了如何将这些额外的资源分配给网络宽度、深度和输入图像分辨率. 最后, 应用这些系数, 将EfficientNet-B0扩展到目标模型大

小或目标计算成本.

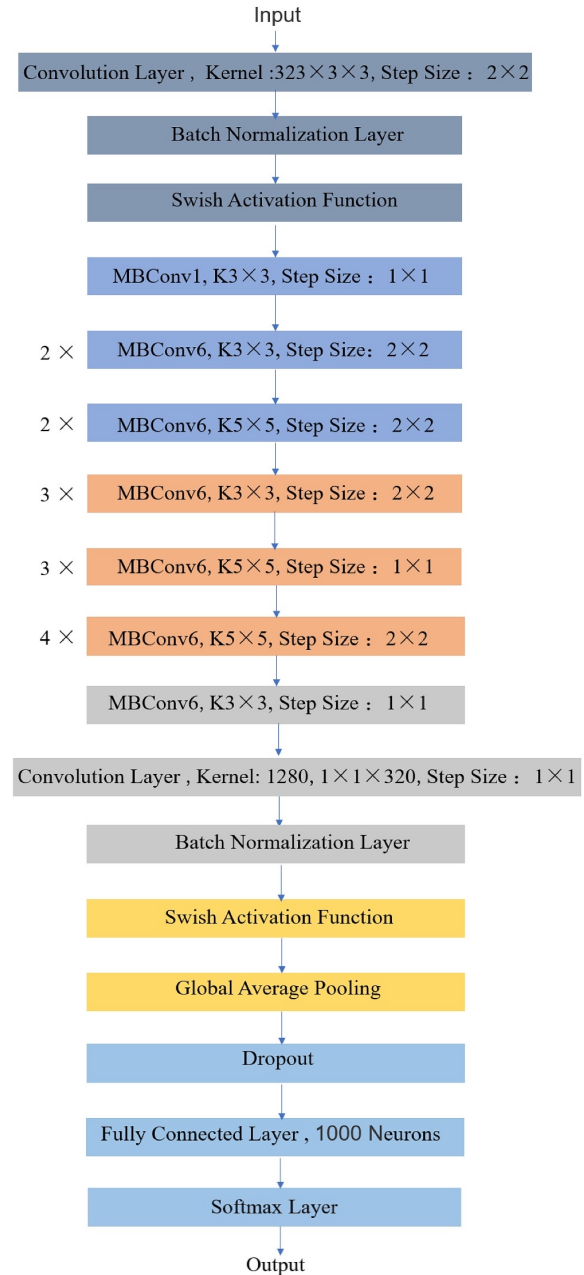


图1 EfficientNet-B0模型的结构图

Fig. 1 Structure of EfficientNet-B0 model

在EfficientNet-B0的基础上, 根据宽度系数、深度系数分别对网络的宽度、深度进行缩放, 得到不同的EfficientNet模型, 分别命名为EfficientNet-B1至EfficientNet-B7. 具体缩放比例如表1所示.

图2为文献[17]的模型扩展图, 其中a是基线网络, b是对基线网络的宽度进行扩展, c是对基线网络的深度进行扩展, d是对基线网络的输入图像分辨率进行扩展, e是对基线网络用复合系数进行统一的扩展.

3 样本数据准备

3.1 Galaxy Zoo简介

Galaxy Zoo是一个志愿者科学计划, 邀请志愿者在网站上对星系形态进行分类. 多个志愿者经过多次识别同一张星系图像, 同一张星系图像的分类

结果经过多名志愿者的判断若存在差异, 天文学家将对这张星系图像做出专业而统一的判断.

表 1 EfficientNet-B0至EfficientNet-B7的缩放尺度
Table 1 Scaling of EfficientNet-B0 to

EfficientNet-B7				
EfficientNet	Depth	Width	Resolution	Dropout
EfficientNet-B0	1.0	1.0	224	0.2
EfficientNet-B1	1.0	1.1	240	0.2
EfficientNet-B2	1.1	1.2	260	0.3
EfficientNet-B3	1.2	1.4	300	0.3
EfficientNet-B4	1.4	1.8	380	0.4
EfficientNet-B5	1.6	2.2	456	0.4
EfficientNet-B6	1.8	2.6	528	0.5
EfficientNet-B7	2.0	3.1	600	0.5

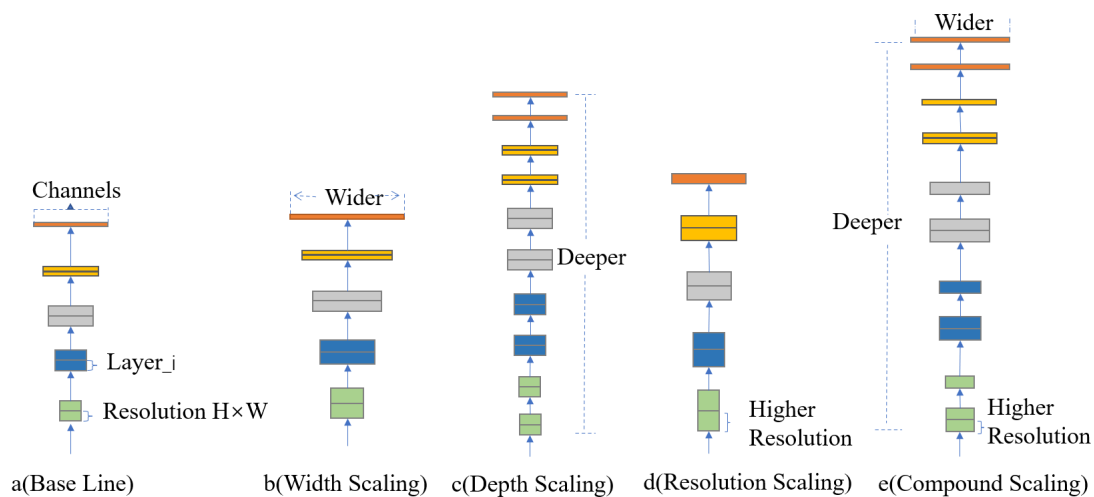


图 2 EfficientNet的扩展图^[17]

Fig. 2 Expansion diagram of EfficientNet^[17]

为进一步提高星系图像分类的准确率, 星系动物园管理者将星系动物园1升级为星系动物园2 (Galaxy Zoo 2, GZ2), 该阶段致力于利用计算机的力量提升星系形态分类的准确率. 管理者举办了一场比赛, 该比赛依托Kaggle平台. 志愿者在星系动物园1阶段完成的对星系形态进行分类的成果, 在这一阶段被用于训练网络. 该比赛被命名为Galaxy

Zoo-the Galaxy Challenge, 其比赛数据采用SDSS发布的Data Release 7 (DR7)数据集的星系图像, 共61578张. 本文针对Galaxy Zoo-the Galaxy Challenge中的5类星系(圆形星系、中间星系、雪茄状星系、侧向星系和旋涡星系)应用EfficientNet对其进行分类研究, 进一步提高这5类星系的形态分类准确率.

3.2 样本数据选取

基于GZ2数据选取规则, 在选取实验数据时, 为了应对雪茄状、中间、圆形星系得到的样本数量较少的问题, 我们采取了与文献[14]相同的方法, 即将任务7 (GZ2中有11个分类问题, 其中任务7的问题为: 星系有多圆? 对应回答有3个分别为: 非常圆、介于两者之间、雪茄形的)对应的3个响应的投票分数从0.8降到了0.5. 基于此, 选取出来的星系图像数量如表2所示, 0到4分别表示圆形星系、中间星系、雪茄状星系、侧向星系和旋涡星系(下同).

表 2 样本集中5类星系数量
Table 2 Number of 5 types of galaxies in the sample

Class	0	1	2	3	4
The number of sample	8434	8069	578	3903	7806

选取出来的彩色星系图像如图3所示, 从左至

右依次为: 圆形星系、中间星系、雪茄状星系、侧向星系和旋涡星系.

3.3 星系图像预处理

图4为本文图像预处理的流程图. 基于5种星系的选取规则得到的星系图像是彩色图像, 我们数据预处理的第1步是将彩色星系图像转化为灰度图像, 再对灰度图像进行开运算, 即先进行腐蚀(Erosion)运算, 再进行膨胀(Dilation)运算. 开运算能够除去孤立的小点, 而总的位置和形状不变. 其中腐蚀运算将图像中的高亮部分腐蚀掉, 领域缩减, 效果图拥有比原图更小的高亮区域, 操作的时候表现为相邻区域用极小值代替, 高亮区域减少. 而膨胀运算是图像中的高亮部分进行膨胀、领域扩张, 效果图拥有比原图更大的高亮区域, 操作的时候表现为相邻区域用极大值代替, 高亮区域增加. 整体表现为除去较小的明亮细节, 保持整体的灰度级和较大的明亮区域不变, 去除噪声.



图 3 各类星系的彩色图像(从左至右, 它们的标签依次为: 圆形星系、中间星系、雪茄状星系、侧向星系和旋涡星系.)

Fig. 3 Color images of various galaxies (from left to right, their labels are: completely round smooth, in-between smooth, cigar-shaped smooth, edge-on and spiral.)

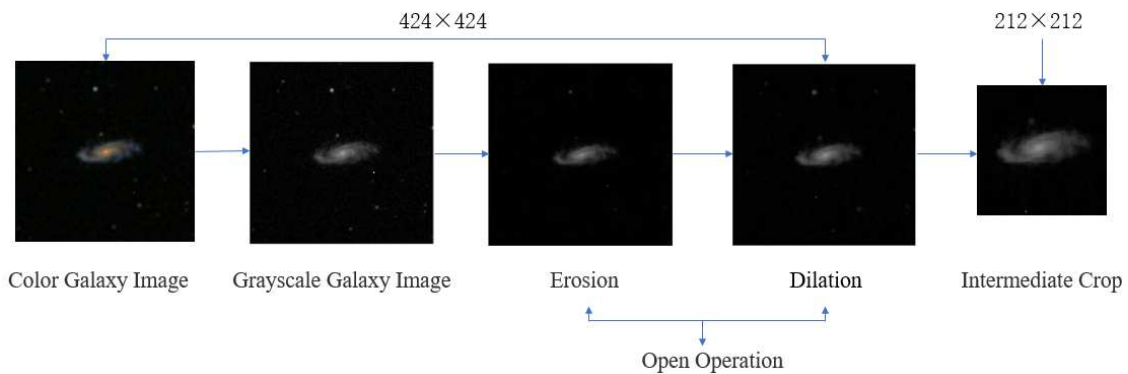


图 4 星系图像预处理流程图

Fig. 4 The data-flow diagram of Galaxy images pre-processing

从星系的图像可以看出, 星系一般位于图像中间, 所以我们对开运算后的图像进行了中间裁剪, 裁剪到 212×212 . 这样可以进一步消除噪声, 裁剪后的5类星系图像按照9:1^[14]的比例随机被划分为训练集和测试集.

4 实验结果与分析

在我们的实验中用于EfficientNet的数据来源相同, 训练集、测试集比例都为9:1, EfficientNet-B0至EfficientNet-B7准确率的对比如表3所示. 通过表3可以看出EfficientNet-B0至EfficientNet-B7的准确率都达到了92%以上, 在EfficientNet-B5上我们获得了最高的准确率为96.76%, 说明EfficientNet-B5对基线网络的缩放比例对于星系的形态分类而言更加有效.

表 3 不同EfficientNet模型在星系图像上总体分类准确率
Table 3 Overall classification accuracy of different EfficientNet models on galaxy images

EfficientNet	Accuracy
EfficientNet-B0	92.23%
EfficientNet-B1	93.75%
EfficientNet-B2	93.91%
EfficientNet-B3	94.33%
EfficientNet-B4	94.93%
EfficientNet-B5	96.76%
EfficientNet-B6	94.62%
EfficientNet-B7	95.54%

表4为EfficientNet-B5在各类星系中的分类效果, 我们采用准确率、精确率、召回率以及F1分数来衡量EfficientNet-B5的分类性能. 可以看出采用EfficientNet-B5对5类星系形态进行分类, 各类的分类准确率都在98%以上, 而精确率、召回率以及F1分数除雪茄状星系外也都在95%以上, 雪茄状星系可能由于数据量过少导致精确率、召回率以及F1分数没有其他几类星系高. 5类星系的平均分类准确率为96.76%, 平均精确率为96.67%, 平均召回率为96.76%, 平均F1分数为96.71%, 全部在96%

以上, 说明了EfficientNet-B5对星系的形态分类而言有着很好的泛化能力.

表 4 EfficientNet-B5在各类星系中的分类效果
Table 4 Classification effect of EfficientNet-B5 in various galaxies

Class	Accuracy	Precision	Recall	F1 Score
0	98.68%	97.63%	97.86%	97.75%
1	98.19%	97.01%	96.53%	96.77%
2	98.78%	73.47%	62.07%	67.29%
3	98.85%	96.12%	95.38%	95.75%
4	99.03%	97.36%	99.10%	98.22%
Average	96.76%	96.67%	96.76%	96.71%

表5是测试数据集在EfficientNet-B5上的混淆矩阵, 其中行为真实分类, 列为预测类别. 可以看出分类正确的圆形星系、中间星系以及漩涡星系数量分别为825、779和774, 而这3类星系中分类错误时也均是错分给了彼此, 并没有被错分为雪茄状星系和侧向星系. 对于雪茄状星系和侧向星系而言, 正确分类的数量分别为36和372. 其中, 雪茄状星系分别有4、15和3个被错分为中间星系、侧向星系和漩涡星系, 而侧向星系分别有3、13和2个被错分为中间星系、雪茄状星系以及漩涡星系. 我们分析认为圆形星系、中间星系和漩涡星系在形态上存在一定的相似性, 所以被错误地分给了彼此. 对于雪茄状星系和侧向星系则可能由于这两类星系数据量太少, 导致网络并没有很好地学习到它们的形态特征而导致了分类结果不佳.

由于本文与ResNet-26^[14]采用的实验数据来源相同, 且都是对圆形星系、中间星系、雪茄状星系、侧向星系和漩涡星系进行形态分类, 所以将本文实验结果与ResNet-26的实验结果作了对比, 如表6所示. 得益于EfficientNet模型从网络宽度、网络深度以及输入图像分辨率3个维度的综合优化, 能看出相对于ResNet-26通过扩宽网络的宽度、减少网络深度来优化网络, EfficientNet-B5的准确率、精确率、召回率以及F1值都有了较大的提升, 从表6的结果来看同时从3个维度进行优化能进一步提升网络的性能.

表 5 EfficientNet-B5在不同形态的星系分类中的混淆矩阵
Table 5 Confusion matrix of EfficientNet-B5 in the classification of galaxies with different morphologies

Class	0	1	2	3	4
0	825	13	0	0	5
1	17	779	0	0	11
2	0	4	36	15	3
3	0	3	13	372	2
4	3	4	0	0	774

表 6 EfficientNet-B5与ResNet-26模型的分类结果对比

Table 6 Comparison of the classification results between EfficientNet-B5 and ResNet-26

Model	Accuracy	Precision	Recall	F1 Score
ResNet-26 ^[14]	95.20%	95.12%	95.21%	95.15%
EfficientNet-B5	96.76%	96.67%	96.76%	96.71%

表 7 不同程度高斯噪声下的分类结果对比

Table 7 Comparison of classification results under different levels of Gaussian noise

Model	Add Gaussian noise level	Accuracy	Precision	Recall	F1 Score
EfficientNet-B5	5%	78.33%	78.65%	78.27%	78.25%
	10%	79.47%	79.933%	79.63%	79.75%
	20%	84.67%	84.13%	84.52%	84.25%
	50%	81.51%	81.01%	81.72%	81.58%
	0%	96.76%	96.67%	96.76%	96.71%

本文鉴于深度学习在天文数据中的应用越来越广泛,但深度网络存在网络设计难度大、调参难的问题,将同时从网络的宽度、深度以及输入图像分辨率3个维度综合优化后的EfficientNet应用于星系形态的分类研究,通过实验可以看出:

(1) EfficientNet-B0至EfficientNet-B7的准确率都达到了92%以上,其中EfficientNet-B5上获得了最高的准确率为96.76%,可见EfficientNet-B5对基线网络的缩放比例对于星系的形态分类而言更加有效;

(2) ResNet-26通过扩宽网络的宽度、减少网络深度来优化网络,而EfficientNet同时对网络宽

为验证本文模型在低信噪比星系图像上的泛化能力,我们往原始星系图像上添加了不同程度的高斯噪声,分别为5%、10%、20%、50%.再对添加了噪声的星系图像使用本文预处理方法,并用之前训练好的网络来进行分类测试,结果如表7所示.由于噪声的添加,模型提取到的星系形态特征受到影响,与未加噪声的星系图像分类效果相比整体有所下降.但添加了高斯噪声的整体分类精度稳定在80%附近,说明EfficientNet-B5对于低信噪比星系图像分类效果稳定,有着不错的泛化能力.

度、网络深度以及输入图像分辨率进行缩放来优化网络,从两者的对比结果来看,同时对3个维度进行优化能进一步提升网络的性能;

(3) EfficientNet-B5对于低信噪比星系图像的分类精度波动不大,稳定在80%附近,说明该模型对于低信噪比星系图像也有着较好的泛化能力.

5 总结与展望

随着大型巡天项目巡天深度和探测效率的不断提升,积累的海量天文数据对其处理方法提出了新的挑战.本文鉴于深度网络存在网络设计难度大、调参难的问题,将EfficientNet应用于星系的

形态分类, 希望进一步提高星系的形态分类精度. 其中EfficientNet-B0至EfficientNet-B7的准确率都达到了92%以上, 在EfficientNet-B5中平均准确率达到96.76%, 平均精确率达到96.67%, 平均召回率达到96.76%, 平均F1分数达到96.71%, 相对于ResNet-26有了较大的提升, 证明了EfficientNet可以应用于星系的形态分类中, 同时该模型对于低信噪比星系图像也有着不错的泛化能力. 本文采用的EfficientNet模型对未来更多大型巡天项目望远镜的后续数据分析提供了更多可能, 例如我国计划于2023年发射的中国空间站望远镜(China Space Station Telescope, CSST)以及大型综合巡天望远镜(Large Synoptic Survey Telescope, LSST). 在未来的工作中我们将用EfficientNet进行不同于本文形态的星系图像分类研究, 研究EfficientNet中网络结构对其形态分类效果的影响, 进一步验证EfficientNet对于星系形态分类的有效性.

参考文献

- [1] 沈世银, 袁方婷, 侯金良. 物理, 2015, 44: 237
- [2] Hubble E P. ApJ, 1926, 64: 321
- [3] de Vaucouleurs G. ApJS, 1961, 5: 233
- [4] Morgan W W. PASP, 1958, 70: 364
- [5] 汪敏, 孔旭. 天文学进展, 2007, 25: 215
- [6] Abraham R G, Valdes F, Yee H K C, et al. ApJ, 1994, 432: 75
- [7] Schade D, Lilly S J, Crampton D, et al. ApJ, 1995, 451: L1
- [8] Abraham R G, van den Bergh S, Nair P. ApJ, 2003, 588: 218
- [9] 范海英. COSMOS天区红移z 3的LBGs星系的形态研究. 上海: 上海师范大学, 2020: 13-23
- [10] Hocking A, Geach J E, Sun Y, et al. MNRAS, 2018, 473: 1108
- [11] Gauci A, Adami K Z, Abela J. Physi, 2010, 39: 34
- [12] Barchi P H, de Carvalho R R, Rosa R R, et al. A&C, 2020, 30: 100334
- [13] Cheng T Y, Huertas-Company M, Conselice C J, et al. MNRAS, 2021, 503: 4446
- [14] 戴加明, 佟继周. 天文学进展, 2018, 36: 384
- [15] Chen J W. Journal of Physics: Conference Series, 2020, 1549: 042033
- [16] Cavanagh M K, Bekki K, Groves B A. MNRAS, 2021, 506: 659
- [17] Tan M X, Le Q V. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019, 97: 6105

Research on Galaxy Morphology Classification Based on EfficientNet

AI Lin-pin¹ XU Quan-feng¹ DU Li-ting¹ XU Ting-ting² GAO Xian-jun¹ LI Guang-ping¹
ZHOu Wei-hong^{1,3}

(1 School of Mathematics and Computer Science, Yunnan Minzu University, Kunming 650500)

(2 Center for Astrophysics, Guangzhou University, Guangzhou 510006)

(3 Key Laboratory of the Structure and Evolution of Celestial Objects, Chinese Academy of Sciences, Kunming 650011)

ABSTRACT The structure and morphology of galaxies can reflect the physical properties of the galaxy itself, and the classification of its morphology is an important part of subsequent analysis and research. The EfficientNet model uses composite coefficients to unify the depth, width, and input image resolution of the deep network model in a more structured manner. This is a new deep network optimization and extension method. This study applies the model to the classification of galaxy data morphology, and the results show that the average accuracy, precision, recall and F1 score (Harmonic mean of precision and recall) based on the EfficientNet-B5 model are all large than 96.6%, which is a significant improvement compared with the classification results of the ResNet-26 model. The experimental results prove that the deep network optimization extension method of EfficientNet is feasible and effective, and can be applied to the morphological classification of galaxies.

Key words methods: data analysis, techniques: image processing, galaxies: general