

doi: 10.3969/j.issn.1000-8349.2018.04.03

基于深度残差网络的星系形态分类

戴加明^{1,2}, 佟继周¹

(1. 中国科学院 国家空间科学中心, 北京 100190; 2. 中国科学院大学, 北京 100049)

摘要: 星系形态与星系的形成和演化有着密切的联系, 因此星系形态分类 (galaxy morphology classification) 成为研究不同星系物理特征的重要过程之一。斯隆数字巡天 (Sloan Digital Sky Survey, SDSS) 等大型巡天计划产生的海量星系图像数据对星系形态的准确、实时分类提出了新的挑战, 而深度学习 (deep learning) 算法能有效应对这类海量星系图片的自动分类考验。面向星系形态分类问题提出了一种改进的深度残差网络 (residual network, ResNet), 即 ResNet-26 模型。该模型对残差单元进行改进, 减少了网络深度, 并增加了网络宽度, 实现了对星系形态特征的自动提取、识别和分类。实验结果表明, 与 Dieleman 和 ResNet-50 等其他流行的卷积神经网络 (convolution neural network, CNN) 模型相比, ResNet-26 模型具有更优的分类性能, 可应用于未来大型巡天计划的大规模星系形态分类系统。

关键词: 星系; 形态分类; 卷积神经网络; 残差网络

中图分类号: P152 **文献标识码:** A

1 引 言

星系形态是描述星系结构最直观的观测特征, 它是不同运动状态的恒星轨道在天空中的投影。星系形态与星系的形成和演化有着密切的联系, 是探究星系物理的重要参数。如早型星系在测光上颜色偏红, 星族 (stellar population) 年龄偏老, 中心大多存在以速度弥散为主导的核球 (bulge), 并且有其经历了多次星系并合留下的壳层 (shell) 结构。大质量早型星系大多处于致密环境, 形态上表现为椭圆星系。晚型星系在测光上表现为颜色偏蓝, 星族成分较为年轻, 含有大量冷气体, 有剧烈的恒星形成活动。晚型星系中恒星的运动状态以绕中心旋转为主导, 形态上表现为有恒星盘和旋臂结构的盘状星系。为了研究不同类型星系的物理特征, 首先要做的是有效区分星系的不同形态^[1, 2]。

星系形态分类可以基于不同的标准。哈勃于 1926 年提出的哈勃序列 (Hubble sequence)

收稿日期: 2018-03-01; 修回日期: 2018-04-16

资助项目: 中国科学院“十三五”信息化建设专项 (XXH13505-04)

通讯作者: 佟继周, tongjz@nssc.ac.cn

是早期用于目视分类的星系形态分类标准中最著名的一种^[3]。随后, 研究者们又逐渐发展了模型化分类和非模型化分类等标准^[4], 其中, 模型化分类的方法是, 假设星系的面亮度轮廓满足一定的函数分布, 并据此对星系形态进行分类, 如核球 (bulge) + 盘 (disk) 分解方法^[5, 6]和 Sérsic 轮廓拟合方法^[7]等。非模型化分类的方法是, 基于聚集度指数、非对称指数、基尼系数等星系形态的结构参数来对星系形态进行分类^[8-14]。在众多星系形态分类标准中, 哈勃序列由于其与中性氢的质量、星系的积分颜色、星系光度和环境等物理参数显著相关, 至今还具有很强的生命力。2007 年推出的星系动物园 (Galaxy Zoo, GZ) 项目的星系形态分类标准便是基于哈勃序列的思想而制定的^[15]。

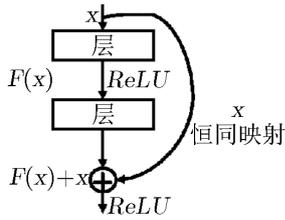
随着观测技术的进步, 以及 SDSS, COSMOS 和 LSST 等大天区星系图像巡天计划的逐步实施, 星系观测数据呈现爆炸式增长趋势, 因此, 对目标证认、特征提取、天体识别、随动观测优先级确定等工作的实时性要求也越来越高, 传统的数据处理和分析方式面临着前所未有的瓶颈, 必须通过创新的手段实现对海量数据的稳定、灵活、实时的处理和挖掘^[16]。在过去的 20 年里, 星系形态自动化分类方法发展迅速, 研究者们尝试使用人工神经网络 (artificial neural network)、朴素贝叶斯 (naive Bayes)、决策树 (decision tree) 和局部加权回归 (locally weighted regression) 等方法进行星系分类^[17-22]。但上述基于传统机器学习算法实现的星系自动分类方法的分类性能取决于特征提取的好坏^[23], 而特征设计和提取需要具有专业领域知识和背景的资深研究人员完成, 设计过程耗时。近 3 年来, 基于深度学习算法的星系形态分类成为非常活跃的研究方向。深度神经网络允许直接输入原始数据, 实现特征自动提取和抽象, 自动找出复杂且有效的高阶特征。越深层的数据特征越抽象, 且越有益于识别和分类^[23, 24]。这在极大程度上解决了基于传统机器学习的星系形态分类方法需要人工提取特征的问题, 减少了人为设计特征的主观性。例如, 2015 年 Dieleman 等人^[25]首次使用深度卷积神经网络进行星系形态分类, 利用 GZ 的 50 000 多张星系图片进行模型训练, 取得 Galaxy Zoo-the Galaxy Challenge 比赛的冠军。同年, Gravit 等人^[26]使用 Dieleman 模型, 利用 CANDELS 巡天数据的高红移星系图像, 将星系分成 5 类, 取得了非常不错的分类效果。

本文对星系形态分类提出一种改进的深度残差网络, 该网络结合 Dieleman 模型和残差网络的优点, 具有较高的分类准确率。

2 深度残差网络

2.1 深度残差网络

卷积神经网络 (convolutional neural network, CNN) 是专门用来处理具有类似网络结构数据 (如图像数据) 的神经网络^[27]。CNN 可以直接使用图像的原始像素作为输入, 自动提取有效特征, 避免复杂的特征工程, 同时具有缩放、平移、旋转等畸变不变性, 因而有着很强的泛化性。CNN 一般包含卷积层、池化层和全连接层。目前它已发展形成 AlexNet^[28], VGG^[29], Inception^[30-33] 和 ResNet^[34, 35] 等一系列模型。

图 1 ResNet 残差单元^[34]

深度残差网络^[34]通过使用残差单元 (residual unit), 可以构建极深的神经网络, 并极快地加速网络训练, 从而使模型的准确率得到很大的提升。ResNet 的核心在于残差单元, 如图 1 所示。其基本思想为: 允许原始输入信息直接传输到后面的层中, 从而避免因卷积层数过多引起的准确率先上升, 然后达到饱和, 再因持续增加网络深度导致准确率下降的问题。假设某段神经网络的输入是 x , 期望输出是 $H(x)$ 。如果直接把

输入 x 传到输出, 作为初始结果, 那么需要学习的目标就是 $F(x) = H(x) - x$ 。ResNet 相当于将学习目标改变, 不再是学习一个完整的输出 $H(x)$, 而只是学习输出和输入的差别 $H(x) - x$, 即残差。 x 通过旁路的支线将输入直接连到后面的层, 使得后面的层可以直接学习残差, 这种结构被称为短连接。这些短连接跳过 2 层或者更多层, 直接执行恒同映射 (identity mapping)。残差单元可以定义为:

$$x_{l+1} = f[x_l + F(x_l, W_l)] \quad (1)$$

其中, x_l 和 x_{l+1} 分别为第 l 个残差单元的输入和输出; f 是激活函数, 一般为修正线性单元 (rectified linear unit, ReLU)^[36], $ReLU = \max\{0, x\}$; F 为残差函数; W 为卷积核。

在 ResNet 的基础上, He 等人^[35]提出了 ResNet V2。在该模型中, 激活函数 f 也采用恒同映射, 每一层都采用批归一化 (batch normalization, BN)^[31], 并采用预激活 (pre-activation) 方式, 即用 BN-ReLU-Conv 取代传统的 Conv-BN-ReLU。

在 ResNet 模型中, 通过大量残差单元的堆叠所构建的极深神经网络看起来非常瘦且高, 通常在其应用过程中会面临特征复用减少等问题^[37]。

2.2 改进的深度残差网络

本文在 ResNet V2 的基础上, 尝试改进残差单元, 减少网络层数, 增加网络宽度。同时, 结合星系图片本身的特点, 我们提出了一种可用于星系形态分类的改进的深度残差网络模型。

改进的模型结构如图 2 和表 1 所示。在第一个卷积层中, 我们吸取 Dieleman 模型的优点, 使用一个相对较大的尺寸为 6×6 , 通道数为 64, 步长为 1 的卷积核, 然后使用一个尺寸为 2×2 , 步长为 2 的最大池化层。最大池化层之后为 4 个卷积组: 卷积组 2、卷积组 3、卷积组 4 和卷积组 5, 其中每个卷积组包含 N 个残差单元, 如卷积组 2 中的第 1 个残差单元的结构为: 尺寸为 1×1 , 通道数为 $64 \times k$ 的卷积 (k 为加宽因子); 尺寸为 3×3 , 通道数为 $64 \times k$ 的卷积; 尺寸为 1×1 , 通道数为 $256 \times k$ 的卷积。这些卷积堆叠成卷积组 2 的 N 个残差单元。下采样在卷积组 2、卷积组 3 和卷积组 4 的最后一层执行, 取步长为 2 的卷积。最后一层是平均池化层, 大小为 4×4 , 输出为 $1 \times 1 \times 4096$, 最后作用于一个包括 5 个神经元的全连接层 *softmax* 函数。

改进的残差单元结构如图 3 所示, 我们使用的 3 个卷积层分别为 1×1 , 3×3 , 1×1 卷积, 其中, m, n 为通道数量, k 为加宽因子。 1×1 卷积的作用是先降维, 再升维, 以减少模型的参数, 并加速训练。 3×3 卷积的设计来自 VGG 模型。该设计的目的是减少模型的参

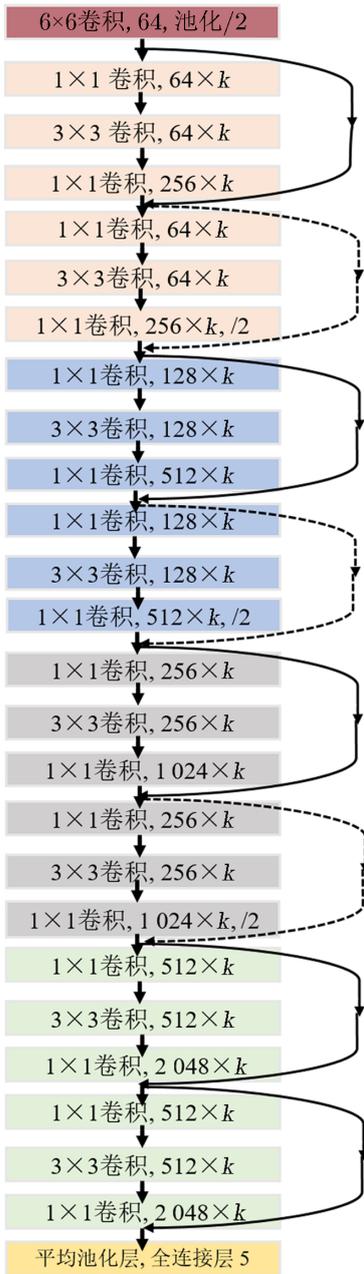


图 2 改进的深度残差网络结构

表 1 改进的深度残差网络结构

组名称	输出尺寸	深度
卷积层 1	64×64	6×6 , 通道数 64
最大池化层	32×32	2×2 , 步长 2
卷积组 2	16×16	$\begin{bmatrix} 1 \times 1, 64 \times k \\ 3 \times 3, 64 \times k \\ 1 \times 1, 256 \times k \end{bmatrix} \times N$
卷积组 3	8×8	$\begin{bmatrix} 1 \times 1, 128 \times k \\ 3 \times 3, 128 \times k \\ 1 \times 1, 512 \times k \end{bmatrix} \times N$
卷积组 4	4×4	$\begin{bmatrix} 1 \times 1, 256 \times k \\ 3 \times 3, 256 \times k \\ 1 \times 1, 1024 \times k \end{bmatrix} \times N$
卷积组 5	4×4	$\begin{bmatrix} 1 \times 1, 512 \times k \\ 3 \times 3, 512 \times k \\ 1 \times 1, 2048 \times k \end{bmatrix} \times N$
平均池化层	1×1	4×4 , 5 维, <i>softmax</i>

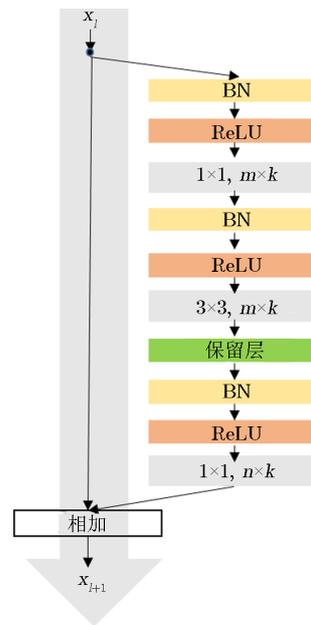


图 3 改进的三层残差单元

数, 从而获得更好的性能^[29]。前 2 个卷积层的通道数一样, 第 3 个卷积层的通道数一般为前 2 个卷积层通道数的 4 倍。残差单元采用预激活方式, 即 BN-ReLU-Conv, 并在其中的 3×3 卷积之后加入保留层 (dropout)^[38], 以减轻过拟合现象。改进的残差单元可以定义为:

$$x_{l+1} = x_l + W_3 \sigma \{ W_2 \sigma [W_1 \sigma (x_l)] \} , \quad (2)$$

其中, x_l, x_{l+1} 分别为第 l 层的输入和输出; σ 包括 BN 和 ReLU; W_1, W_2, W_3 分别表示 3 个卷积核。保留层被放在 W_2 操作之后, 偏置 (bias) 项为了简化符号被省略。

除了最后一层全连接层使用 *softmax* 函数以外, 所有卷积层都使用 ReLU 作为激活函数。训练使用的代价函数为交叉熵函数, 采用动量梯度下降算法, 动量值为 0.9, 批大小为 128。初始学习率设置为 0.1, 然后每 3×10^4 步下降 10 倍。权重延迟为 0.0001, 保留层保留率为 0.8, 网络初始化权重值采用 He 等人^[38] 的结果。

在设计的残差网络模型中, 我们改进了残差单元, 减少了网络的层数, 加宽了网络的宽度 (通道数目), 同时融合了 Dieleman 模型的优点。网络的总层数为 $2N + 2$ 层。

3 数据集选取与预处理

3.1 数据集选取

星系动物园^[39] 是一个在线众包项目, 邀请了大量志愿者根据所给星系彩色图像的形态对星系进行分类。本文采用的是星系动物园 2 (Galaxy Zoo 2, GZ2) 提供的数据和分类标准, 其中, 实验数据集来自^[40] 星系动物园在 Kaggle 平台上举办的比赛 Galaxy Zoo-the Galaxy Challenge。该比赛的训练集包含来自斯隆数字化巡天数据 SDSS DR7^[41] 的 61 578 张带标签的星系观测彩色 RGB 图片, 图片的尺寸为 $424 \times 424 \times 3$ 像素。SDSS 对星系的观测包括 5 个光学波段 (u, g, r, i 和 z), 人们通常取前 3 个波段 (u, g 和 r) 的数据合成相应的 RGB 星系图像。每一张图片的标签为 1×37 的向量, 来自于 GZ2 志愿者投票分数的修正累计频率值。GZ2 将星系形态分为 11 个问题和 37 个答案, 本文选取其中 5 类星系进行模型训练和实验。这 5 类星系分别为圆形星系 (即完全圆的平滑星系)、中间星系 (即介于完全圆的平滑星系和雪茄状平滑星系之间的星系)、雪茄状星系 (即雪茄状平滑星系)、侧向星系和旋涡星系。

GZ2 数据发布白皮书中有明确的干净样本 (well-sampled galaxies) 阈值选取规则^[42], 即对于同一张星系图片, 对其进行分类的志愿者人数必须大于 20 个, 且计算得到的累计投票分数修正值满足一定的阈值, 该图片才可被分类到某个星系类别。例如, 一张图片须满足 3 个阈值条件 ($f_{\text{features}/\text{disk}} \geq 0.430$, $f_{\text{edge-on,no}} \geq 0.715$, $f_{\text{spiral,yes}} \geq 0.619$) 才可被分类到旋涡星系。因 GZ2 的干净样本阈值选取规则过于保守, 针对上述 5 类中的平滑星系 (圆形星系、中间星系和雪茄状星系), 能够选取到的数据样本数略显不足。为了得到足够数量的样本进行模型训练与测试, 本文将平滑星系的阈值选取标准适当放宽, 从 0.8 降到了 0.5, 而侧向星系和旋涡星系的阈值选取规则仍参照 GZ2 数据发布白皮书中的默认取值^[42]。表 2 为本文实际采用的 5 类干净样本数据的选取规则。

依据上述规则最终获得的干净样本共包含 28 790 张星系图片, 其中圆形星系、中间星系、雪茄状星系、侧向星系和旋涡星系的图片分别有 8 434, 8 069, 578, 3 903 和 7 806 张。图 4 为从干净样本原始数据集中随机抽取的 5 类星系图像。

表 2 干净样本选取规则^[42]

星系类型	任务	阈值	N_{sample}
圆形星系	T01	$f_{\text{smooth}} \geq 0.469$	8 434
	T07	$f_{\text{completely round}} \geq 0.50$	
中间星系	T01	$f_{\text{smooth}} \geq 0.469$	8 069
	T07	$f_{\text{in between}} \geq 0.50$	
雪茄状星系	T01	$f_{\text{smooth}} \geq 0.469$	578
	T07	$f_{\text{cigar-shaped}} \geq 0.50$	
侧向星系	T01	$f_{\text{features/disk}} \geq 0.430$	3 903
	T02	$f_{\text{edge-on,yes}} \geq 0.602$	
旋涡星系	T01	$f_{\text{features/disk}} \geq 0.430$	7 806
	T02	$f_{\text{edge-on,no}} \geq 0.715$	
	T04	$f_{\text{spiral,yes}} \geq 0.619$	

注: T01—T11 为 GZ2 中的 11 个分类问题; f_{smooth} 为一张图片被分类为平滑星系的频率; $f_{\text{completely round}}$, $f_{\text{in between}}$ 和 $f_{\text{cigar-shaped}}$ 分别为一张图片被分类为完全圆的、处于圆和雪茄状之间的、雪茄状的平滑星系的频率; $f_{\text{features/disk}}$ 为一张图片被分类为有特征/盘状结构的频率; $f_{\text{edge-on,yes}}$ 和 $f_{\text{edge-on,no}}$ 分别为一张图片被分类为侧向星系和非侧向星系的频率; $f_{\text{spiral,yes}}$ 为一张图片被分类为旋涡星系的频率; N_{sample} 为选取到的干净样本数目。



图 4 GZ2 中随机抽取的星系图片

将 28 790 张干净样本图片按 9:1 的比例划分为训练集和测试集, 即训练集包含 25 911 张图片, 测试集包含 2 879 张图片。训练集用于训练模型, 而测试集用于测试模型的泛化能力。表 3 展示了训练集和测试集中每一类星系的图片数目。可以看出, 在训练集和测试集中, 5

类星系的图片数目满足同分布要求。

表 3 训练集和测试集中星系图片样本统计

	圆形星系		中间星系		雪茄状星系		侧向星系		旋涡星系		总计
	N_1	占比 / (%)	N_2	占比 / (%)	N_3	占比 / (%)	N_4	占比 / (%)	N_5	占比 / (%)	
训练集	7 591	29	7 262	28	520	2	3 513	14	7 025	27	25 911
测试集	843	29	807	28	58	2	390	14	781	27	2 879
数据集	8 434	29	8 069	28	578	2	3 903	14	7 806	27	28 790

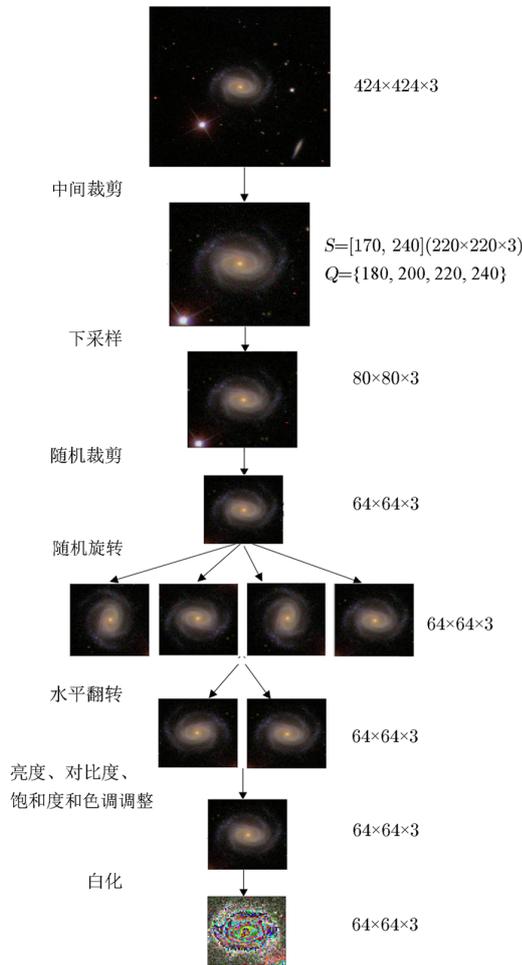


图 5 训练集预处理流程图

3.2 数据预处理

选取后的训练集和测试集星系图片需进行数据预处理后，方可作为模型的训练和测试输入。训练集和测试集的数据预处理过程是不同的。训练集预处理流程如图 5 所示。

网络结构有上千万参数，而训练集图片数量有限，不到 30 000 张。为了减轻过拟合现象，需使用数据增强处理方法，以尽量增大训练集数量。训练集预处理的第一步是中间裁剪处理。中间裁剪是将图片裁剪到一个范围值 S ，称为训练尺度抖动 (training scale jittering)^[29]。图片被随机裁剪到 S 范围中的某个值时，不同的图片会被裁剪成不同的尺寸。即使是同一张图片，在不同的迭代次数时也会被裁剪成不同的尺寸。裁剪不仅可以起到降维的作用，还可以消除一些噪声，如次级物体。由于数据集中几乎所有星系都位于图片的正中间，四周包含了大量深色背景，所以人们选择从图片中间裁剪到一个范围值 $S = [170, 240]$ 。这样处理可在保证星系完整信息的情况下，将图片的尺寸降到原来的 $1/4$ 。

第二步是下采样处理。将训练集中的图片下采样至 $80 \times 80 \times 3$ 像素，以达到降维和节省计算资源的目的。第三步是数据增强处理，包括随机裁剪、随机旋转、水平翻转和光学畸变等，其中，随机裁剪处理是将图片随机裁剪至 $64 \times 64 \times 3$ ，以便在不改变原星系图片标签的情况下，训练集

图片数目增加至原来的 256 倍。随机旋转处理是指对图片随机旋转 $0^\circ, 90^\circ, 180^\circ, 270^\circ$ 。由于

星系图片具有旋转不变性, 旋转之后星系图片的标签不会改变, 因此, 数据量可增大至原来的 4 倍。水平翻转亦不会改变星系图片的标签, 因此, 数据量可增大至原来的 2 倍。光学畸变处理包括改变亮度、对比度、饱和度和色调等。最后一步是图像白化处理。图像中相邻像素之间具有很强的相关性, 图像白化处理可以降低输入数据的冗余性。最后, 网络输入的图片大小为 $64 \times 64 \times 3$ 。

测试集预处理步骤相对简单。第一步是中间裁剪处理, 即通过测试尺度抖动 (testing scale jittering), 将图片裁剪到 4 个具体的值, 即 $Q = \{180, 200, 220, 240\}$; 第二步是下采样处理, 即把测试集中的图片下采样至 $80 \times 80 \times 3$ 像素; 第三步再次进行中间裁剪处理, 把图片裁剪到 $64 \times 64 \times 3$; 第四是图像白化处理。最后, 得到网络的测试输入。

4 实验及分析

4.1 超参数选择

在改进的深度残差网络中, 各超参数的设置是决定模型最终分类性能的关键问题之一。本文针对第一个卷积层卷积核大小、加宽因子 k 、每组残差单元数 N 和保留率大小等超参数开展一系列实验, 以确定其最佳取值。具体实验过程为: 首先, 针对每类超参数, 选取数个不同值进行模型训练; 其次, 将训练得到的模型在测试集的 4 个测试尺度上分别测试 10 次; 最后, 取每个测试尺度的最好结果进行分类性能的比较分析。

(1) 第一个卷积层的卷积核大小与网络性能的关系

为了探索第一个卷积层的卷积核大小与网络性能的关系, 我们分别设计了 3 种卷积核大小: 3×3 , 6×6 和 7×7 。 3×3 的设计源于 VGG 中采用的小型卷积核的思路; 6×6 是 Dieleman 模型中的第一个卷积层的卷积核大小; 7×7 是原 ResNet 中的策略。其他参数保持不变。实验结果见表 4。

从表 4 可以看出, 当第一层卷积核的尺寸选择为 6×6 时, 模型的准确率最高。所以, 第一层卷积核的尺寸定为 6×6 。

(2) 加宽因子 k 和每组残差单元数 N 与网络性能的关系

为了探索加宽因子 k 和每组残差单元数 N 与网络性能的关系, 我们设计了 4 组实验: $k = 1, N = 1$; $k = 1, N = 2$; $k = 2, N = 1$; $k = 2, N = 2$ 。其他参数控制不变。实验结果如表 5 所示。从表 5 可以看出, 当 $k = 2, N = 2$ 时, 网络的分类准确率最高。

表 4 第一层卷积核尺寸与网络性能的关系

第一层卷积核尺寸	准确率/(%)
3×3	92.1181
6×6	95.2083
7×7	93.7153

表 5 加宽因子 k 和每组残差单元数 N 与网络性能的关系

k	N	层数	准确率/(%)
$k = 1$	$N = 1$	14	93.4028
$k = 1$	$N = 2$	26	94.7569
$k = 2$	$N = 1$	14	93.0556
$k = 2$	$N = 2$	26	95.2083

表 6 保留率大小与网络性能的关系

保留率	准确率/(%)
0.5	92.8819
0.7	94.2917
0.8	95.2083

(3) 保留率大小与网络性能的关系

为了探索保留率大小与网络性能的关系,保留层被放置在每个残差单元中的 3×3 卷积之后,用来防止过拟合。其他参数保持不变。实验结果如表 6 所示。可以看出,当保留率为 0.8 时,网络的分类准确率最高。

通过上述一系列实验,当第一层卷积核的尺寸为 6×6 ,加宽因子 k 为 2,每组残差单元数为 2,以及保留率为 0.8 时,得到的模型具有相对最优的分类性能。由于整个网络共计 26 层(包括 25 个卷积层和 1 个全连接层),因此该模型被称为 ResNet-26。ResNet-26 共计有 2.6×10^7 个参数,在测试尺度为 220,训练迭代次数为 4.2×10^4 时取得最好结果。

(4) 星系图像类型与网络性能的关系

为了验证星系图像类型对模型结果的影响,本文将星系 RGB 图像转化成灰度图像,然后采用 ResNet-26 对 RGB 图像和灰度图像进行比较,结果如表 7 所示。可以看出,对于灰度图像,在测试尺度为 240 时结果最好,但最高准确率仅为 93.4722%;而对于 RGB 图像,最高准确率达到 95.2083%。在分类性能评价指标 F1 值、受试者工作特征(receiver operating characteristic, ROC) 曲线下的面积(area under ROC curve, AUC) 等其他评价指标上,RGB 图像的模型结果也均优于灰度图像的模型结果。上述结果表明,星系 RGB 图像的 3 个颜色通道值对模型的分类性能提升是有利的,因此本文中的模型使用星系 RGB 图像作为输入。

表 7 星系 RGB 图像和灰度图像单测试尺度的最高准确率、F1 值和 AUC 值

图像类型	测试尺度(Q)	准确率/(%)	F1 值	AUC 值
RGB 图像	220	95.2083	0.9515	0.9823
灰度图像	240	93.4722	0.9342	0.9786

4.2 ResNet-26 与其他模型比较

为了验证 ResNet-26 模型的分类性能,在相同的测试环境和测试集情况下,将 ResNet-26 与 Dieleman 和 ResNet-50 等流行的深度卷积神经网络模型进行实验对比。

表 8 测试集的精确率、召回率和 F1 值

星系类别	精确率	召回率	F1 值
圆形星系	0.9611	0.9634	0.9622
中间星系	0.9561	0.9431	0.9495
雪茄状星系	0.7234	0.5862	0.6476
侧向星系	0.9412	0.9485	0.9448
旋涡星系	0.9573	0.9782	0.9677
平均	0.9512	0.9521	0.9515

4.2.1 ResNet-26 分类结果

ResNet-26 的最好结果在测试尺度为 220 时取得,获得的精确率、召回率、F1 值、混淆矩阵、ROC 曲线和 AUC 值如表 8 和表 9 所示。

表 8 给出了测试集中每一类星系分类的精确率、召回率和 F1 值,它们分别为 0.9512, 0.9521 和 0.9515,其中,圆形星系取得了最高的精确率值 0.9611,旋涡星系取得了最好的召回率 0.9782 和 F1 值 0.9515。

表 9 为测试集的混淆矩阵,其中,列为真实标签,行为预测标签。从混淆矩阵中可以看

出, 有 815 张圆形星系图片、762 张中间星系图片、34 张雪茄状星系图片、368 张侧向星系图片和 763 张旋涡星系图片被正确分类, 由此可以计算出, 5 个类别星系的分类准确率分别为: 圆形星系, 96.6785%; 中间星系, 94.4238%; 雪茄状星系, 58.6207%; 侧向星系, 94.3590%; 旋涡星系, 97.6953%。另外, 有 29 张圆形星系图片被错分给了中间星系, 12 张雪茄状星系图片被错分给了侧向星系, 18 张侧向星系图片被错分给了雪茄状星系。我们通过分析认为, 圆形星系图片被错分给中间星系属于正常现象, 因为这两个星系都属于平滑星系, 且两者没有严格的区分界限; 雪茄状星系和侧向星系错分则较为意外。经多次实验对比, 我们初步判定, 这是因为两个星系在形状上的相似性所致。

表 9 测试集的混淆矩阵

	圆形星系	中间星系	雪茄状星系	侧向星系	旋涡星系
圆形星系	815	21	0	0	10
中间星系	29	762	0	0	17
雪茄状星系	0	4	34	18	2
侧向星系	0	3	12	368	5
旋涡星系	4	7	1	5	763

图 6 为测试集的 ROC 曲线, 其中每一种颜色代表一个星系类别。真正率越接近于 1, 假正率越接近于 0, 表示曲线预测得越好。也就是说, 曲线越接近左上角, 代表结果越好。从图中可看出, 每一类都预测得很好, 其中侧向星系预测得最好, 其 AUC 值为 0.993 9; 雪茄状星系预测得相对差些, 其 AUC 值为 0.948 1。模型的平均 AUC 值为 0.982 3, 说明模型的整体预测性能很好。

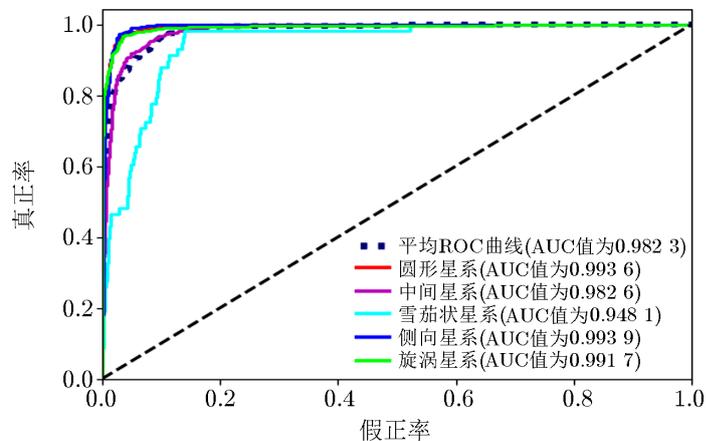


图 6 测试集上 5 个星系类别的 ROC 曲线

从 ResNet-26 模型所获得的精确率、召回率、F1 值、混淆矩阵、ROC 曲线和 AUC 值这 6 个指标可以看出, 该模型在圆形星系、中间星系、侧向星系和旋涡星系的分类识别过程中均表现优秀, F1 值都在 0.94 以上, 其中圆形星系、侧向星系和旋涡星系的 AUC 值都超过

了 0.99。ResNet-26 针对旋涡星系的分类性能最好，其分类准确率达到 97.695 3%；针对雪茄状星系的分类性能表现一般。训练集中的雪茄状星系图片数量相对较少是造成分类性能一般的原因之一。

4.2.2 不同模型的分类型性能对比

本文在相同的测试环境下，使用相同的测试集对 3 个模型进行对比实验。Dieleman 模型为 7 层卷积神经网络，包括 4 个卷积层和 3 个全连接层，其权重初始化和训练算法全部来自文献[25]；Resnet-50 为 50 层的残差网络，其权重初始化和训练算法来自文献[34, 35]。3 个模型的对比实验结果如表 10 和表 11 所示，其中，表 10 为不同模型在多个测试尺度上的平均准确率，即在每个测试尺度上测试 10 次所取得的最好结果平均值。这样可以使模型的测试结果更加稳定。表 11 为不同模型在单个测试尺度上的最高准确率、F1 值和 AUC 值，即在每个测试尺度上测试 10 次所取得的每个测试尺度上最好结果。比较表 10 和表 11 可以看出，表 11 的结果比表 10 中的结果好。

表 10 不同模型的多测试尺度平均准确率

模型	图片尺寸		准确率/(%)
	训练尺度 (S)	测试尺度 (Q)	
Dieleman ^[25]	[170, 240]	180, 200, 220, 240	93.880 0
ResNet-50 ^[35]	[170, 240]	180, 200, 220, 240	94.097 2
ResNet-26	[170, 240]	180, 200, 220, 240	94.687 5

表 11 不同模型的单测试尺度最高准确率、F1 值和 AUC 值

模型	测试尺度 (Q)	准确率/(%)	F1 值	AUC 值
Dieleman ^[25]	180	94.652 8	0.945 6	0.979 3
ResNet-50 ^[35]	220	94.687 5	0.946 1	0.982 3
ResNet-26	220	95.208 3	0.951 5	0.982 3

由表 10 可以看出，ResNet-26 取得了最高的准确率 (94.687 5%)。Dieleman 模型是专门为星系图片设计的卷积神经网络模型，其准确率为 93.880 0%。ResNet-50 尽管是专为 ImageNet 数据集而设计的，但在星系数据集上也表现出了很不错的泛化能力，其准确率达到 94.097 2%。

由表 11 可知，ResNet-26 在测试尺度为 220 时取得了最高的准确率 (95.208 3%)，以及最高的 F1 值 (0.951 5) 和 AUC 值 (0.982 3)。Dieleman 模型在测试尺度为 180 时取得其最好结果。ResNet-50 在测试尺度 220 时取得其最好结果，并获得了最高的 AUC 值 (0.982 3)。

总之，通过以上对比结果可以看出，ResNet-26 的分类性能最好。

4.3 实验环境

为实施模型训练和测试，我们配置了 1 台 GPU 服务器，其配置分别为 2 块 NVIDIA Tesla K80 GPU、CentOS Linux 7 操作系统、8.044 版本 CUDA 和 5.1 版本 CUDNN。此

外, 我们还采用了 Python, Pandas, Scikit-learn^[44], Scikit-image^[45], TensorFlow^[46]等语言和库。本文模型代码可以在 <https://github.com/Adaydl/GalaxyClassification> 上下载。

5 总 结

将深度卷积神经网络应用于星系形态分类是天文大数据时代解决星系图像需要快速、高效地自动分类问题的一个有效方法。本文在深度残差网络的基础上, 通过改进残差单元, 减少网络深度, 加宽网络宽度等方法, 并结合星系图片本身的特点, 设计了基于改进的深度残差网络的星系形态分类框架模型 ResNet-26。实验结果为, ResNet-26 模型与 Dieleman 模型和 ResNet-50 模型相比, 其平均准确率、F1 值和 AUC 值等评价指标均为最优。未来, 我们将获取更多高质量的星系图片数据, 以进一步验证 ResNet-26 模型的适用性, 并在更细粒度星系形态分类方面进行深入研究。同时, 我们将尝试把人工设计的特征加入到神经网络中, 以期提升神经网络的分类性能。

参考文献:

- [1] 汪敏, 孔旭. 天文学进展, 2007, 25: 215
- [2] 方官文, 马仲阳, 孔旭. 天文学报, 2015, 56: 425
- [3] Hubble E. ApJ, 1926, 64: 321
- [4] Sandage A. Annu Rev Astron Astrophys, 2005, 43: 581
- [5] Schade D, Lilly S J, Crampton D, et al. ApJ, 1995, 451: L1
- [6] Ratnatunga K U, Griffiths R E, Ostrander E J. ApJ, 1999, 118: 86
- [7] Sersic J L. Bulletin of the Astronomical Institutes of Czechoslovakia, 1968, 19: 105
- [8] Kent S M. ApJ, 1985, 59: S115
- [9] Abraham R G, Valdes F, Yee H K C, et al. ApJ, 1994, 432: 75
- [10] Conselice C J, Bershady M A, Jangren A. ApJ, 2000, 529: 886
- [11] Abraham R G, Tanvir N R, Santiago B X, et al. MNRAS, 1996, 279: L47
- [12] Conselice C J. ApJ, 2003, 147: S1
- [13] Abraham R G, Van Den Bergh S, Nair P. ApJ, 2003, 588: 218
- [14] Lotz J M, Primack J, Madau P. ApJ, 2004, 128: 163
- [15] Zooniverse. Edwin Hubble, the Man Behind HST. (2010-08-16). <https://blog.galaxyzoo.org/2010/08/16/edwin-hubble-the-man-behind-hst/>
- [16] 崔辰州, 于策, 肖健, 等. 科学通报, 2015, 5: 445
- [17] Naim A, Lahav O, Sodre Jr L, et al. MNRAS, 1995, 275: 567
- [18] Owens E A, Griffiths R E, Ratnatunga K U. MNRAS, 1996, 281: 153
- [19] Bazell D, Aha D W. ApJ, 2001, 548: 219
- [20] De La Calleja J, Fuentes O. MNRAS, 2004, 349: 87
- [21] Banerji M, Lahav O, Lintott C J, et al. MNRAS, 2010, 406: 342
- [22] Ferrari F, de Carvalho R R, Trevisan M. ApJ, 2015, 814: 55
- [23] LeCun Y, Bengio Y, Hinton G. Nature, 2015, 521: 436
- [24] Bengio Y, Courville A, Vincent P. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35: 1798

- [25] Dieleman S, Willett K W, Dambre J. *MNRAS*, 2015, 450: 1441
- [26] Gravet R, Cabrera-Vives G, Pérez-González P G, et al. *ApJ*, 2015, 221: S8
- [27] Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: MIT Press, 2016
- [28] Krizhevsky A, Sutskever I, Hinton G E. *ImageNet Classification with Deep Convolutional Neural Networks*. (2012). <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>
- [29] Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. (2015). <https://arxiv.org/abs/1409.1556>
- [30] Szegedy C, Liu W, Jia Y, et al. *Going Deeper with Convolutions*. (2015). https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html
- [31] Ioffe S, Szegedy C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. (2015). <https://arxiv.org/abs/1502.03167>
- [32] Szegedy C, Vanhoucke V, Ioffe S, et al. *Rethinking the Inception Architecture for Computer Vision*. (2016). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html
- [33] Szegedy C, Ioffe S, Vanhoucke V, et al. *AAAI*, 2017, 4: 12
- [34] He K, Zhang X, Ren S, et al. *Deep Residual Learning for Image Recognition*. (2016). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html
- [35] He K, Zhang X, Ren S, et al. *Identity Mappings in Deep Residual Networks*. (2016). https://link.springer.com/chapter/10.1007/978-3-319-46493-0_38
- [36] Nair V, Hinton G E. *Rectified Linear Units Improve Restricted Boltzmann Machines*. (2010). <https://www.cs.toronto.edu/hinton/absps/reluICML.pdf>
- [37] Zagoruyko S, Komodakis N. *Wide Residual Networks*. (2016). <https://arxiv.org/abs/1605.07146>
- [38] He K, Zhang X, Ren S, et al. *Delving Deep into Rectifiers: Surpassing Human-level Performance on Imagenet Classification*. (2015). https://www.cv-foundation.org/openaccess/content_iccv_2015/html/He_Delving_Deep_into_ICCV_2015_paper.html
- [39] [https://www.galaxyzoo.org/\(2018-03-13\)](https://www.galaxyzoo.org/(2018-03-13))
- [40] <https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge> (2018-03-13)
- [41] <http://www.sdss.org/> (2018-03-13)
- [42] Willett K W, Galloway M A, Bamford S P, et al. *MNRAS*, 2013, 435: 2835
- [43] Ceri S, Bozzon A, Brambilla M. *An Introduction to Information Retrieval*. Berlin: Springer, 2013: 96
- [44] Pedregosa F, Varoquaux G, Gramfort A, et al. *Journal of Machine Learning Research*, 2011, 12: 2825
- [45] Van der Walt S, Schonberger J L, Nunez-Iglesias J, et al. *PeerJ*, 2014, 2: e453
- [46] Abadi M, Barham P, Chen J, et al. *OSDI*, 2016, 16: 265

Galaxy Morphology Classification Using Deep Residual Networks

DAI Jia-ming^{1,2}, TONG Ji-zhou¹

(1. National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Galaxy morphology is closely related to the galaxy formation and evolution, so galaxy morphology classification is one of the most important processes in the study of the physical characteristics of different galaxies. The massive galaxy images data produced by the large scale surveys, such as Sloan Digital Sky Survey (SDSS), poses a new challenge to classifying galaxy images accurately and real time, and deep learning algorithm can effectively and automatically deal with the kind of very large collections of galaxy images. In this paper, a modified residual network (ResNet), namely ResNet-26, is proposed for galaxy morphology classification. This model improves the residual unit, while reduces the depth of the network and widens the width of the network, and realizes the automatic extraction of galaxy morphological features to identification and classification. The experimental results show that ResNet-26 has better classification performance compared with other popular convolution neural network models such as Dieleman and ResNet-50, and can be applied to large-scale galaxy classification in forthcoming surveys.

Key words: galaxy; morphology classification; ResNet; CNN