第1章　绪论

高分辨率遥感影像的目标识别是在计算机技术辅助下，通过对遥感影像的预处理、分割以及进行特征的表示与描述等步骤，将高分辨率遥感图像中的各类物体依据其独特的光谱特征实现分类的目的，并最终实现对高分变绿遥感影像中的各类物体进行识别与解译。要分辨率遥感图像识别在控制导弹、防暴反恐、导弹防御等军事领域，还是在抢险救灾、灾害监控、交通监管、城市规划等民用领域都发挥着重要作用。在遥感影像精度大大提高以及人工智能技术高速发展的今天，遥感影像的目标识别技术将有着更大的发展空间。

近年来，随着空间传感器技术与计算机硬件与软件的不断发展，空间高分辨率的观测系统的光谱、频谱的分辨率增加，时间和空间分辨率不断提升，遥感图像的分辨率得到了明显地提高。当前，世界上最先进的高分辨率遥感图像的分辨率达到了分米级，合成孔径雷达（Synthetic Aperture Radar, SAR）已达米级，采用干涉测量技术的合成孔径雷达（Interferometric Synthetic Aperture Radar, InSAR）相对位移精度达到了毫米级，高光谱图像（Hyperspectral Image）的分辨率甚至达到了纳米级。另一方面，卫星重访周期的单位已缩短至天。上述特征使得拍摄的高分辨率遥感图像不仅具备数量多、多源化、异构化的特征，还具备高维度、多尺度、非平稳的内部特性，具备大数据的特点[1]。然而，适配于中低分辨率的传统遥感影像解译方法，在识别包含车辆、飞机、轮船等复杂特征的高分辨率遥感影像时效率低下，甚至无法准确识别。面对新时代高分辨率遥感影像目标识别的更高要求，亟需探索全新的、更加高效的高分辨率遥感影像目标识别方法与模型，用来提高高分辨率遥感图像的目标识别精度、识别技术的智能化水平[2-4]。

遥感图像在光谱分辨率上的提高，使得该图像与自然图像的差异逐步缩小，越来越多的研究者将计算机视觉的理论与方法应用到高分辨率遥感图像分类与目标识别等任务中。

1.1　研究背景与意义

遥感技术综合了空间学、电子学、光学以及计算机科学的最新成果，是当代高新技术的一个重要代表。该技术起源于 1983 年摄影技术的发展。19 世纪中期，人们利用热气球对地面进行摄影，观察地面的物体。20 世纪初期，人们利用飞机从空中对地面进行摄影，将拍摄的航空图像用于勘测、军事等方面，使得遥感技

术得以快速发展。

随着空间技术、无线电技术、计算机技术、光学技术以及材料技术的不断提高，遥感技术不断发展。遥感技术获取的信息可以用于各个领域。例如，1960 年美果第一个卫星拍摄了地球的云圈，极大地促进了大气探测领域的发展，为世界人民了解地球大气圈、陆地、海洋以及生物的分布起到了核心的作用。除此之外，遥感技术一般是从高空中获取地面相关事物的图像信息。这种获取信息的方式具有：获取信息速度快、信息的范围大、获取信息的方式限制低以及手段多的特点。基于上述特征，遥感技术受到工业界和学术界的广泛关注，如表 1-1。

表 1-1 世界各国遥感技术的发展

Table 1-1 Countries Invest in Developing Remote Sensing Technology

国家	时间（年）	事件
美国	1960	发射卫星
美国	1999	发射卫星 IKONOS
美国	2001	发射卫星 QuickBird
中国	2005	建立地球观测组织 GEO
美国	2008	发射卫星 GeoEye-1
美国	2009	发射卫星 WorldView-3
法国	2011	发射卫星 Pleiades
中国	2014	发射卫星 GF-2
中国	2015	启动“星、机、地”综合定量遥感系统重大项目
中国	2016	高景-1

目前我们正在进入一个智能化时代（互联网+及工业 4.0 时代），大数据、云计算和人工智能已经成为这个时代进步的三驾马车(（图 1-1 所示），它们分别为智能化时代提供数据、算力和算法层面的支持，从而成为各行各业技术革新和社会发展的重要引擎。

在智能化时代的背景下，云计算为遥感技术提供了大量的存储空间，使得海量高分辨率遥感图像的存储成为可能。研究者可以从大量存储的数据中识别、提取感兴趣的信息，为国家领土安全、城市规划与建设、灾害的监控与预防、国家

经济发展的等领域带来巨大的益处，具体来说：（1）在国家领土安全方面：从高分辨率遥感图像上可以识别各种军事武器，如航母、军舰、基地等关键军事目标，从而获取敌军的战略信息；（2）在城市规划与建设方面：通过识别遥感图像获取

图 1-1 拉动智能时代的三驾马车

Figure 1-1 Pulling the Troika of the Intelligent Age

城市的街道、绿化面积和人员活动等信息，辅助城市规划工程师制定合理的城市建设方案；（3）在灾害监测与预防方面：从高分辨率的遥感图像中获取灾区的受灾面积，受灾类型，周围的环境状况等关键信息，辅助制定急求方案与运输急求物资；（4）在研究国家经济发展领域：通过遥感技术获取与国家发展相关的主要信息，如夜晚的灯光分布信息，从而了解国家不同地区的经济发展情况和比较不同国家的经济发展状况。

基于上述分析，遥感图像识别是运用遥感技术解决各种问题的关键步骤。高分辨遥感图像中的目标具有多样性与复杂性，具体来说：（1）在多样性方面：搭载在航空、航天等设备上的仪器拍摄的遥感图像由于拍摄角度的原因，图像中物体呈现出事物种类多、事物之间联系复杂，既包含要识别的目标，又包含诸多干扰物体，如山川、河流、各种建筑物等；（2）在复杂性方面：目标与目标之间、目标与非目标之间往往存在复杂的联系，且呈现出相互影响的特征。该分辨率遥感图像的这些特点为识别图像中的目标带来了挑战。传统的遥感图像识别技术（如：数字图像处理软件、统计模式识别方法等）存在严重依赖人工解译，解译精度较低等诸多问题。因此，如何高效、准确地识别高分辨率遥感图像中的目标是一个亟待解决的关键问题。

卷积神经网络在自然图像处理以及目标识别领域具有优秀的表现[5]。另一方

面，随机计算机技术以及光学原理的不断发展（表 1-2 展示了近年来我国卫星拍摄遥感图像分辨率的变化），高分辨率遥感图像与自然图像的差别不断缩小，这使得卷积神经网络运用到遥感图像的目标识别中成为可能。与自然图像相比，高分辨率遥感图像呈现出图片尺寸较大、目标的拍摄角度多变、光影变化等特点，给卷积神经网络运用到高分辨率遥感图像的目标识别上增加了阻碍。如何正确地将卷积神经网络运用到高分辨遥感图像的目标识别中，自动、高效地识别目标具有重要的研究价值。

表 1-2 对地观测卫星

Table 1-2

卫星名称	发射时间	传感器	分辨率
资源 3 号	2012 年	全色 TDICCD 相机	< 2.1m
		全色 TDICCD 相机	< 3.5m
		多光谱相机	< 5.8m
北京一号	2005 年	全色波段	4m
		多光谱波段	32m
HJ-1-A	2008 年	CCD 相机	30m
		高光谱成像仪	100m
HJ-1-B	2008 年	CCD 相机	30m
		红外多光谱相机	150m
			300m

1.2 研究内容与成果

在处理自然图像方面，卷积神经网络具有优秀的特征表示能力、强大的自我学习能力和准确的目标识别能力。随着搭载平台与遥感技术的不断革新，高分辨率图像与自然图像的差异也逐步减小，为卷积神经网络应用到高分辨率遥感图像的目标识别提供了便利。本文主要研究卷积神经网络在高分辨率遥感图像的目标识别任务中的有效性与准确性，具体表现在以下 3 个方面：

1) **提出卷积神经网络应用到高分辨率遥感图像分类的框架：**结合高分辨率图像的特征与卷积神经网络的基本原理提出了 C-SMI 框架，首先在原始高分辨率图像的基础上提取特征，然后通过卷积运算和降采样的方法对提取的特征进行抽象和化简。本文通过经验学习的方法运用多个经典卷

积神经网络模型在高分辨率遥感图像的目标识别准确率与传统的词袋模型对比，结果显示本文提出的框架相对于词袋模型在高分辨率遥感图像分类方面的准确性方面有显著的提升。

2) **验证卷积神经网络识别单个高分辨率遥感图像目标的效率与准确性：**本文拟采用经典的 R-CNN 和 Fast R-CNN 卷积神经网络模型与传统的词袋模型在识别单个高分辨率遥感图像目标的效率与准确性。研究结果展示 R-CNN 和 Fast R-CNN 具有较高的目标识别准确率。

1.3 论文组织结构

本文的组织结构如下：

第一章， 绪论，主要包括课题的研究背景与意义、研究内容与成果以及论文的组织结构。

第二章， 背景介绍，主要包括卷积神经网络和词袋模型的基础概念与理论，以及目标识别、遥感图像的目标识别、卷积神经网络的国内外研究现状。

第三章， 深入讨论本文提出的基于神经网络的高分辨率遥感图像目标识别的框架。

第四章， 在开源的数据集上验证所提框架的有效性与准确性，并与传统的词袋模型进行对比，主要包括研究问题阐述、实验对象说明、实验流程。

第五章， 通过图、表的方式回答所提的问题，并深入分析实验结果。

第六章， 研究工作的总结以及未来的展望。

第2章　背景介绍

本章介绍卷积神经网络和词袋模型的相関概念、基础理论以及国内外研究现状。

2.1　相关技术与概念

本节首先介绍卷积神经网络的基本特征，然后介绍几种经典的神经网络算法，最后介绍一种传统的遥感图像目标识别方法—词袋模型。

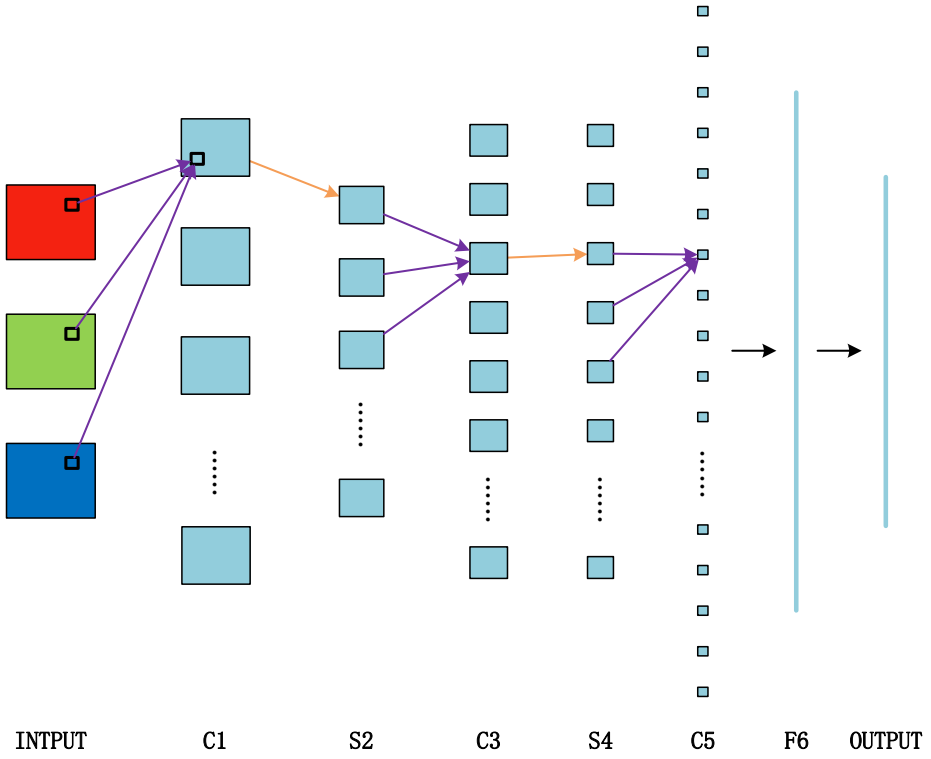

图 2-1 LeNet-5 的网络结构

Figure 2-1

2.1.1　卷积神经网络

现在的神经网络最早是由 Lecun 等人提出的 7 层网络结构（如图 2-1）LeNet-5[7]，主要包括卷积层（C）、池化层（S）、全连接层（F）。通常卷积层紧跟在输入层之后，池化层在卷积层之后。神经网络通过重复输入层、卷积层与池化层这种连接结构增加神经网络的深度。为了将提取的特征映射为实际的输出，全连接层一般出现在神经网络的末端。接下来，详细介绍神经网络中的主要组件：

1)　输入层：在训练和测试卷积神经网络的过程中，输入层的神经元数目通常是

固定的，负责接收大量的多维数组（数据）。与传统的全连接神经网络（如自适应神经网络[11]）相比,卷积神经网络不需要对数据进行归一化处理（即不需要限制数组中每一个维度的值$0 < v_i < 1$），也不需要对数据进行中心化处理（即数据每一个维度的值减去平均值）[12]。

2) 卷积层：在泛函分析中，卷积是通过两个已知函数$\boldsymbol{f_1}$和$\boldsymbol{f_2}$生成新函数$\boldsymbol{V}$的算子，记为公式 2.1。在卷积神经网络中，卷积运算起到了滤波的作用，将输入图像的局部区域与权值进行加权运算，其权值由一个的函数定义。卷积的具体表现

$$\boldsymbol{V(t) = \int_{-\infty}^{+\infty} f_1(p) f_2(t-p) dp} \quad \textbf{(2.1)}$$

形式为公式 2.2，其中$\boldsymbol{f(x,y)}$表示图中坐标为$\boldsymbol{(x,y)}$的灰度值,w(i,j)为卷积核。公式 2.1 的运算结果为卷积核对相关输入的响应值，即为输出图像中的一个像素。

$$\boldsymbol{g(x,y) = f(x,y) \times w(i,j) = \sum_{i,j} f(x+i, y+j) \times w(i,j)} \quad \textbf{(2.2)}$$

卷积神经网络区别于传统的全连接网络的重要特征是存在卷积层。该层首先利用卷积核对上一层的局部输出数据进行卷积运算，然后利用非线性激活函数将卷积的结果限定在某一个范围内，使得模型具备非线性的特征。通常，利用公式 2.3 和 2.4 进行卷积运算，其中$\boldsymbol{Z_j}$表示卷积运算的输出值，$\boldsymbol{X_i}$是卷积层的输入，$\boldsymbol{K_{i,j}}$表

$$\boldsymbol{Z_j = \sum_i X_i * K_{i,j} + B_j} \quad \textbf{(2.3)}$$

$$\boldsymbol{A_j = f(Z_j)} \quad \textbf{(2.4)}$$

示卷积核，$\boldsymbol{B_j}$表示卷积的加性偏项，$\boldsymbol{A_j}$是该卷积层的输出特征图，$\boldsymbol{f}(\cdot)$表示一个激活函数。与传统全连接的权重更新方式类似，在训练模型的过程中卷积神经网络的卷积核通过反向误差传播算法学习得到，并且每一个卷积层包含多个卷积核。不同的卷积核与输入的图像进行卷积运算，得到输入图像的多样化特征。一般情况下，激活函数是非线性函数（常用的激活函数如表 2-1，每一种激活函数对应的函数图像如图 2-2），根据输入的值计算出相应的输出并将输出的值控制在一定的范围内，使得卷积神经网络可以捕获图像的非线性特征。表 2-1 展示的激活函数中，Sigmoid 函数是全连接神经网络以及传统卷积神经网络中常用的激活函数。图 2-2(a)表明 Sigmoid 函数是单挑递增的光滑函数，函数值的范围限制在(0,1)。然而，Sigmoid 函数在数据正向传播时计

表 2-1 常用的激活函数

Table 2-1

激活函数	定义	参数	图像
Sigmoid	$f(x) = 1/(1+e^{-x})$	--	图 2-2(a)
Tanh	$f(x) = 2/(1+e^{-2x}) - 1$	--	图 2-2(b)
ReLU	$f(x) = \begin{cases} xx \geq 0 \\ \alpha xx < 0 \end{cases}$	--	图 2-2(c)
LeakyReLU	$f(x) = \begin{cases} xx \geq 0 \\ \alpha xx < 0 \end{cases}$	$\alpha \in (0,1)$	图 2-2(d)
PReLU	$f(x) = \begin{cases} xx \geq 0 \\ \alpha xx < 0 \end{cases}$	α是学得的参数	图 2-2(d)
RReLU	$f(x) = \begin{cases} xx \geq 0 \\ \alpha xx < 0 \end{cases}$	$\alpha \sim uniform(a,b)$	图 2-2(d)
ELU	$f(x) = \begin{cases} xx \geq 0 \\ \alpha(e^x - 1)\ x < 0 \end{cases}$	A是预先设参数	图 2-2(e)

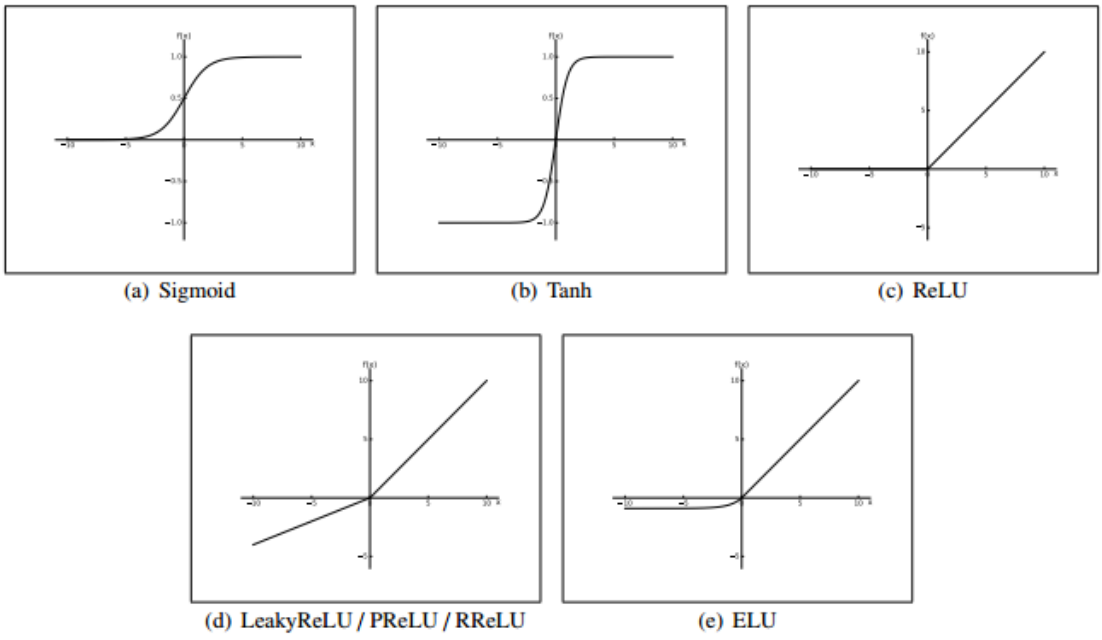

图 2-2 激活函数对应的图像

Figure2-2

算量大，在误差反向传播时容易出现梯度消失的情况，在深层的神经网络中表现欠佳。随着卷积神经网络架构的不断改进以及算法的不断更新，卷积神经网络的层次不断增加（AlexNet 有 8 层，VGGNet 有 19 层，ResNet 最大有 152 层，GooLeNet 有 22 层），ReLU 激活函数逐步代替 Sigmoid 函数，得到广泛的使用。与 Sigmoid 函数相比，ReLU 激活函数有效地缓解了误差反向传播时梯度消失的

问题，直接以监督的模式训练神经网络，并且具有收敛速度快的特点。然而，当一个较大的梯度流经神经元时，传统的 ReLU 激活函数容易使得该神经元不能被激活，造成模型训练不充分的问题。为此，LeakyReLU、PReLU 等改进的 ReLU 激活函数陆续被提出，并取得良好的效果。

3) **池化层：**该层将池化策略通到卷积层输出的图像特征，降低特征图的大小与分辨率，有效地减少了参数的卷积神经网络中参数的个数，也具有防止过拟合的作用。典型的采样方法有两种：（1）对特征图像相邻像素中取最大值；（2）计算相邻像素的平均值。图 2-3 展示了选取最大池化方法作用在4×4图像后的结果，假设池化的窗口为2×2。尽管最大和平均的池化方法在卷积神经网络中表现良好，一些

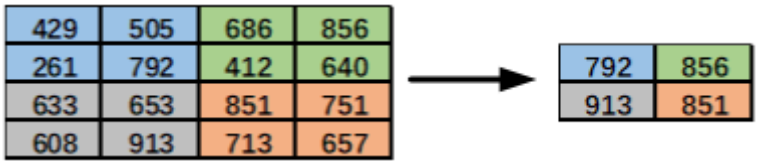

图 2-3 最大池化法作用实例

Figure2-3

研究者提出了更好的池化方法：He 等人提出了空间金字塔池化方法[13]，对于任意大小的特征图首先分成 16、4、1 个块，然后在每个块上运用最大池化方法，最后将池化后的特征拼接得到一个固定维度的输出上。该方法的主要有点是对于任意尺寸的输入产生固定大小的输出。

4) 全连接层和输出层：卷积神经网络的末端是全连接层和输出层。输入的图像经过若干次卷积运算与池化后，得到高水平的抽象特征，然后每一个神经元将提取的图像特征均输入到全连接层（即该层与上一层的连接结构具有全连接的特征），最后通过输出层的神经元输出模型最后结果。这种连接方式不仅有效地降低了卷积神经网络中参数的个数还充分地提取图像复杂、多样化的特征，还能挖掘图片更深层次的信息。输出层是卷积神经网络的最后一层，该层的神经元数目在训练和测试的过程中具有不变性，主要功能是对图像做出最终的判断并输出。

2.1.2 卷积神经网络的特点

从卷积神经网络的提出到第一次成功的运用在图像识别上，该技术一方面收益于计算机硬件与软件的发展，另一方面卷积神经网络自身也具有成功的特点。

在图像分类与图像的目标识别领域，卷积神经网络在工作的过程中不需要人员的参与，自动地从图像中提取特征，从数据中获得识别图像目标或者对图像进行分类的逻辑，极大地解放了人员的参与；另一方面，卷积神经网络内部层与层之间不同层内部的机制，也使得卷积神经网络的成功成为必然，例如卷积神经网络通过局部感受野、权值共享和降采样的方式使得图像做位移、缩放等变换时具有不变性，具体来说：

1） **局部感知野：**该特征受到生物局部神经元获取视野方式的启发，研究者认为只需要对输入图像的局部进行感知，即某一输出层一个点的值对应输入层某一区域，随着卷积神经网络层数不断加深，高层神经元可以获得高水平的原始图像的全局抽象特征。与全连接的神经网络相比，在保证提取特征效果不变的条件下卷积神经网络中参数数目更少。假设输入图像的像素为$100*100$，全连接神经网络的隐藏层和卷积神经网络的卷积层均具有 100 个神经元且每一个神经元与$10*10$的局部图像相连，则对于全连接网络，参数的数目为：$100*100*100$；然卷积神经网络的参数数目为：$100*10*10$（参数的数目相对于全连接神经网络减小了 100 倍）。通过上述分析可知，卷积神经网络极大地化简了神经网络，提高了神经网络的训练与测试时间，节省了成本。

2） **权值共享：**随着问题要解决问题的复杂程度不断增大，卷积神经网络的深度呈现出加深的趋势，这种架构使得卷积神经网络的参数不断增多。另一方面对了提高模型的准确度，训练数据的个数是海量的，势必需要更多的时间进行训练与测试。基于上述问题，研究者探究进一步减小卷积神经网络参数的方法，提出了权值共享的方法。这种想法基于生物的神经网络和图像不同部位具有一定的相似性这一观察。该想法具体的实现是同一个卷积核在与图像的不同部位进行卷积运算时的参数保持不变，从而在卷积层的感受野的基础上进一步减小参数的数目。在上述例子中，卷积层每一个神经元与$10*10$的局部图像相连，即参数的数目是 100，由于这 1000 个神经元共用这个 100 个参数，因此一个卷积核的参数数目为 100；否则参数数目为$1000*100$。一般情况下，为了充分提取图像的特征信息，卷积层往往具有多个卷积核，然而多个卷积核需要的参数数目与卷积核的个数呈现线性关系，避免出现参数数目爆炸的情况。

3） **降采样：**局部感受野和权值共享有效地减少了卷积神经网络中的参数数目，

极大地降低了模型的训练与测试时间。但是当输入图像比较大时，仍会降低模型的训练与测试速度。基于上述例子，输入的图像经过第一个卷积层之后得到$90*90=8100$个特征图，每一个特征图与输入的图像进行卷积得到$91*91=8281$维的卷积特征，一共存在 8100 个特征图，总计66,248,000维的特征向量。如此庞大的维数不仅需要大量的时间，还容易出现过拟合的现象。为了缓解这一问题，研究者提出了降采样的方法，即将特征图映射成一个小尺寸的特征图，但是扔保留了特征图中的信息，相当于对特征图进行了抽象。由此，在提取输入图像特征信息的基础上不仅降低了卷积神经网络的参数数目，还降低了输入的维数。

4） **误差反向传播：**传统的全连接神经网络由于存在阈值函数，使得基于误差调整权重是一个难以解决的问题。直到梯度下降法的提出，缓解了全连接神经网络的权重更新问题。该算法需要数据标签，并根据模型的输出值与标签进行对比，得到隐藏层与输出层每一个神经元的误差。由于输出层每一个神经元的输出误差是由隐藏层中每一个神经元的误差与权重加权的和导致的，因此输出层所有神经元的误差与隐藏层某一个神经元的权重加权并求和，得到隐藏层的误差。重复以上步骤直到更新第一个隐藏层与输入层之间的权重。研究者基于全连接神经网络更新误差的方式，提出了批量梯度下降法，并成功运用到卷积神经网络中：假设存在一个包含m个样本的训练集 $X=\{(x_1,y_1),(x_2,y_2),\dots,(x_m,y_m)\}$。从X中选取样本并输入模型后，运用梯度下降法对其求解，神经网络的损失函数为：

$$J(W,b;x,y)=\frac{1}{2}\left\|h_{W,b}(x)-y\right\|^2$$

则包含m个样本的数据集的代价函数可以表示为

$$J(W,b)=\left[\frac{1}{m}\sum_{i=1}^{m}\left\|h_{W,b}(x^{(i)})-y^{(i)}\right\|^2\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(W_{ji}^{l})^2$$

从上式可以看出，代价函数是凹函数，即具有最小值，但存在多个极小值。为了避免结果出现在错误的极小值位置，可以多次输入来进行调整。在实际的问题解决过程中，梯度下降法的表现良好。通过判定目标函数是否收敛来判断模型是否训练结束。通过求导，可以得出每一次输入数据之后参数的变化情况，并通过以

下公式进行更新：

$$W_{ij}^{l+1} = W_{ij}^{l} - \alpha \frac{\partial}{\partial W_{ij}^{l}} J(W,b)$$

$$b_{i}^{l+1} = b_{i}^{l} - \alpha \frac{\partial}{\partial b_{i}^{l}} J(W,b)$$

其中，\alph 是学习速率。对于整个数据集，代价函数的偏导如下：

$$\frac{\partial}{\partial W_{ij}^{l}} J(W,b) = [\frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial W_{ij}^{l}} J(W,b;x^{(i)},y^{i})] + \lambda W_{ij}^{l}$$

$$\frac{\partial}{\partial b_{i}^{l}} J(W,b) = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial}{\partial b_{i}^{l}} J(W,b;x^{(i)},y^{(i)})$$

综上所述，卷积神经网络具有强大能力的同时，其内部层次之间以及层次之中的特点使得卷积神经网络的参数极大地降低，输入数据的维度得到进一步降低，极大地节省了测试与训练的时间。使得卷积神经网络在实际的应用中表现良好，甚至达到了人类的水平。针对某一种具体类型的问题，或者在 LeNet-5 模型的基础上，研究者们先后提出了 AlexNet[5]、GooLeNet[8]、VGG-16[9]、ResNet[10]等被广泛学习的神经网络模型，表 2-2 总结了常见神经网络最重要的特征，本文涉及的神经网络模型详细地描述在 2.1.3 章节。

2.1.3 卷积神经网络框架简介

深度学习技术在各个领域取得了令人惊叹的成绩，例如特斯拉、谷歌以及百度的无人驾驶汽车；人脸识别；AlphGo 战胜了世界排名第一的柯洁等，使得深度学习收到了广泛的关注。为了方便更多的人从事深度学习相关的研究，研究者们基于不同的高级语言（例如 Java、python、C++等）开发出了一系列的框架，例如 TensorFlow、Caffe、Keras 等。Clark[14]等人为了研究开源的深度学习框架受欢迎的情况，基于框架的关注度以及提交代码的数量，对 16 个常用框架流行程度进行了排序，结果展示在图 2-3。从图中可以明显看出 TensorFlow[15]的关

注度最高，说明该框架是目前使用最多的深度学习框架。Dlib 框架的关注度最低，说明该框架使用的人数最少。TensorFlow 框架不仅可以用来研究深度学习相关的问题，还可以用来做其它方面的数值研究，该框架具有可移植性，即用户可以在

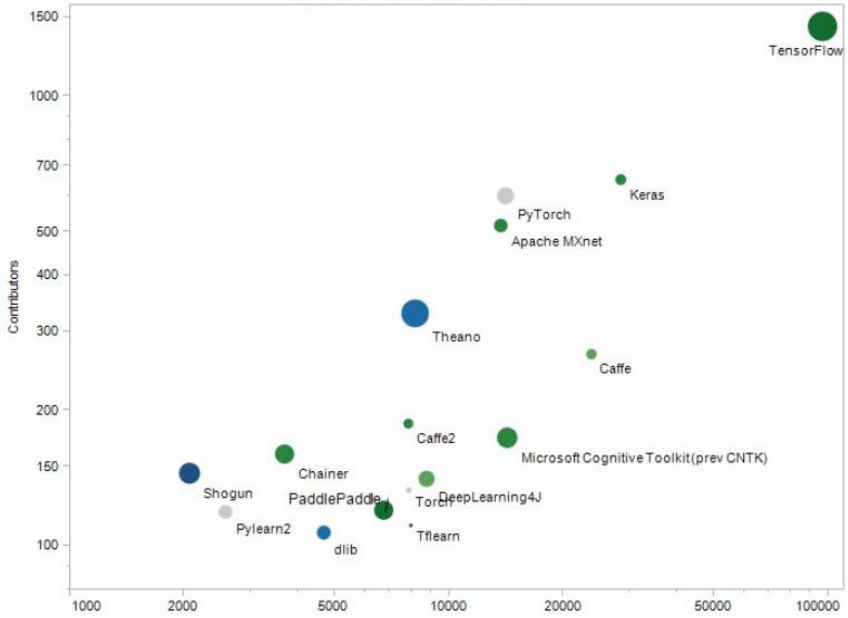

图 2-3 16 种深度学习开源框架关注度排名

Figure2-3

多个平台使用该框架。例如台式计算机中的一个或多个 CPU（或 GPU），服务器，移动设备等。TensorFlow 最初是由 Google 公司开发出来的，由于其简介的代码、友好的接口以及可移植性，使得该框架收到广泛的关注。后来，发布到开源社区，越来越多的研究者改进架构、修复故障等，使得该框架具有非常高的可用性。该框架的主要特征如下：

1) **高度的灵活性：**TensorFlow 不仅可以用来进行神经元网络的研究还可以将研究者的计算转化为为一个数据流图。该框架提供了有用的工具来帮助研究者组装“子图”（常用于神经网络）。此外，研究者可以自己在该框架的基础上定制自己的“上层库”。

2) 真正的可移植性：很多开源框架（如 caffe）只能在含有 CPU 和 GPU 的平台上运行，例如台式机、服务器、手机移动设备等。Tensorflow 可以在没有特殊硬件的前提下，调用接口，实现既定的功能。

3) 自动求微分：基于梯度的机器学习算法会受益于 Tensorflow 具有自动求微分的能力，极大地便利了基于梯度的机器学习算法。研究者只需要定义预测模型的结构，将该结构和目标函数结合在一起，然后添加数据，该框架将自动为研究者计算相关的微分导数。

4) **多语言支持：**为了适应开发者技能具有多元化的特征，该框架不仅提供了多

种高级语言（C++、python 等）的接口（实现了某种功能的函数）而且提供了用户交互的界面。研究者在用户交互界面可以使用自己掌握的语言来构建与执行数据。

5) **性能最优化：**该框架给予了线程、队列、异步操作等最佳的支持，可以将研究者可用的硬件计算潜能全部发挥出来。该框架还支持自动地将图像中的计算元素分配到不同设备上，通过这种方式来实现充分利用计算资源的目的。

2.1.4 AlexNet 卷积神经网络

AlexNet 卷积神经网络[5]是由 Alex Krizhevsky 等人在 2010 年图像分类竞赛中提出的，该神经网络成功地将 120W 张高分辨率的图像分为 1000 个类别，并取得了冠军。该框架需要两个 GPU 分别跑前五个卷积层，从第一个到第三个全连接层共享参数，每个 GPU 分摊二分之一的参数，最后将生成的特征图分配给 GPU 进行计算，具体结构展示在图 2-4。

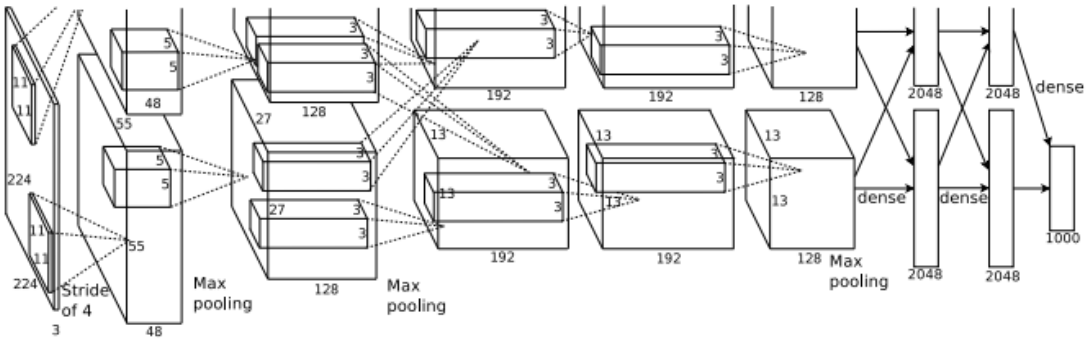

图 2-4 AlexNet 卷积神经网络架构

Figure 2-4 The Architecture of AlexNet

AlexNet 神经网络具有以下特征：

1) 收敛/训练速度加快，用了非饱和神经元（ReLU，即线性矫正单元）。ReLU 是一种非线性神经元（当输入 x 大于 0，则输出为 x；输入 x 小于 0，则输出 0）较之前的激活函数计算速度更快，收敛也更快；

2) 模型并行。文中并没写是模型并行，但我看介绍认为是一种（有数据并行的）模型并行。训练过程用了两个 GTX580 GPU 3GB（一个 gpu 无法 cover 住），除了第三层卷积、最后的两个全连接层（两个 GPU 将全连接拆成各自两部分）和 Softmax（汇聚到单个 GPU 处理），其余部分处理可以看作两个 GPU 的数据并行（AlexNet 架构图的上部分和下部分对应两个 GPU 各自的处理流程，单 GPU 的话，就是 group=2 这个可以看后文中通过 netscope 可视化出的网络结构）。相

比单 GPU 没加快多少，同时某些层两个 GPU 共享参数，可以减少显存占用这样以便在一次参数更新放更多的图片加入训练。发现双卡比单卡 top1 和 top5 的 error 有下降（我认为是作者忘了改 SGD 学习率导致，先跑 one-gpu net 再跑 two-gpu net）；

3) 正则化——数据扩增和 dropout。在数据扩增方面，一般的做法是在已有数据的基础上做某种变化，进而在变化的过程中找到某种规律。在保证数据多样性的同时，相似的图像输入到卷积神经网络中，这种方式使得卷积神经网络的权重不断更新，在输入到输出的过程中映射图像的本质；在 dropout 方面，模型在学习的过程中，缺少 dropout 模型的训练很快会出现过拟合的现象。

2.1.5 GooLeNet 卷积神经网络

GoogLeNet 提出最直接提升深度神经网络的方法就是增加网络的尺寸，包括宽度和深度。深度也就是网络中的层数，宽度指每层中所用到的神经元的个数。但是这种简单直接的解决方式存在的两个重大的缺点。

(1) 网络尺寸的增加也意味着参数的增加，也就使得网络更加容易过拟合。

(2) 计算资源的增加。

因此想到将全连接的方式改为稀疏连接来解决这两个问题。由 Provable bounds for learning some deep representations.提到数据集的概率分布由大又稀疏的深度神经网络表达时，网络拓扑结构可由逐层分析与输出高度相关的上一层的激活值和聚类神经元的相关统计信息来优化。但是这有非常多的限制条件。因此提出运用 Hebbian 原理，它可以使得上述想法在少量限制条件下就变得实际可行。通常全连接是为了更好的优化并行计算，而稀疏连接是为了打破对称来改善学习，传统常常利用卷积来利用空间域上的稀疏性，但卷积在网络的早期层中的与 patches 的连接也是稠密连接，因此考虑到能不能在滤波器层面上利用稀疏性，而不是神经元上。但是在非均匀稀疏数据结构上进行数值计算效率很低，并且查找和缓存未定义的开销很大，而且对计算的基础设施要求过高，因此考虑到将稀疏矩阵聚类成相对稠密子空间来倾向于对稀疏矩阵的计算优化。因此提出了 VGG-16 卷积神经网络。inception 结构的主要思想在于卷积视觉网络中一个优化的局部稀疏结构怎么样能由一系列易获得的稠密子结构来近似和覆盖。上面提到网络拓扑结构是由逐层分析上一层的相关统计信息并聚集到一个高度相关的单

元组中，这些簇（单元组）表达下一层的单元（神经元）并与之前的单元相连接，而靠近输入图像的底层相关的单元在一块局部区域聚集，这就意味着我们可以在一块单一区域上聚集簇来结尾，并且他们能在下一层由一层 1x1 的卷积层覆盖，也即利用更少的数量在更大空间扩散的簇可由更大 patches 上的卷积来覆盖，也将减少越来越大的区域上 patches 的数量。研究者在设计 inception 的结构时，限制了 inception 结构中滤波器的大小（$1*1,3*3,5*5$），来达到避免 patch 对齐的问题。为了在高层能提取更抽象的特征，就要减少其空间聚集性，因此通过增加高层 inception 结构中的$3*3,5*5$积数量，捕获更大面积的特征。

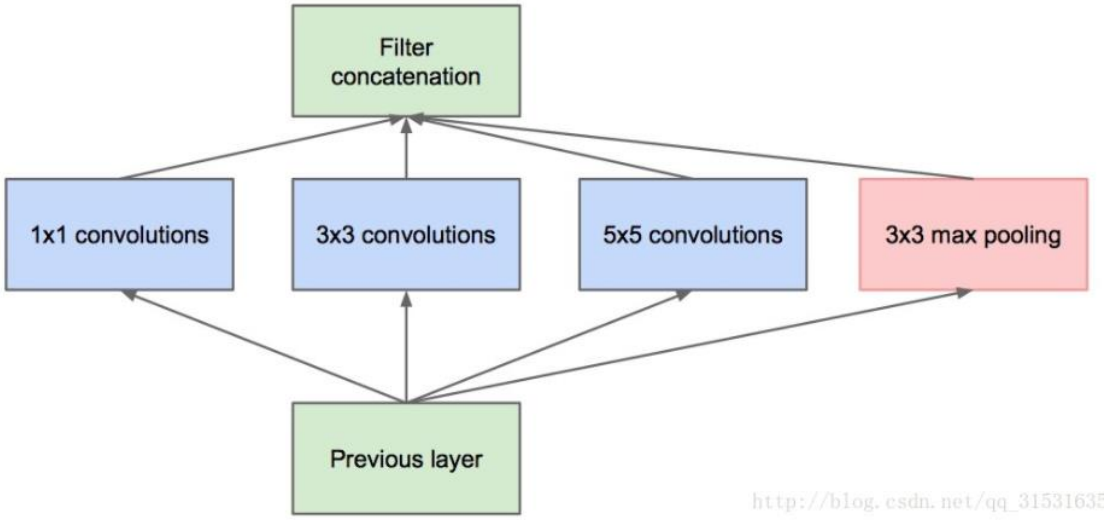

图 2-5 inception 的架构

Figure 2-5 The Architecture of inception

2.1.6 ResNet 卷积神经网络

它差不多是当前应用最为广泛的CNN特征提取网络。它的提出始于2015年，作者中间有大名鼎鼎的三位人物He-Kaiming, Ren-Shaoqing, Sun-Jian。绝对是华人学者的骄傲啊。

VGG 网络试着探寻了一下深度学习网络的深度究竟可以深几许以能持续地提高分类准确率。我们的一般印象当中，深度学习愈是深（复杂，参数多）愈是有着更强的表达能力。凭着这一基本准则 CNN 分类网络自 Alexnet 的 7 层发展到了 VGG 的 16 乃至 19 层，后来更有了 Googlenet 的 22 层。可后来我们发现深度 CNN 网络达到一定深度后再一味地增加层数并不能带来进一步地分类性能提高，反而会招致网络收敛变得更慢，test dataset 的分类准确率也变得更差。排除数据集过小带来的模型过拟合等问题后，我们发现过深的网络仍然还会使分类准

确度下降（相对于较浅些的网络而言

正是受制于此不清不楚的问题，VGG 网络达到 19 层后再增加层数就开始导致分类性能的下降。而 Resnet 网络作者则想到了常规计算机视觉领域常用的 residual representation 的概念,并进一步将它应用在了 CNN 模型的构建当中，于是就有了基本的 residual learning 的 block。它通过使用多个有参层来学习输入输出之间的残差表示，而非像一般 CNN 网络（如 Alexnet/VGG 等）那样使用有参层来直接尝试学习输入、输出之间的映射。实验表明使用一般意义上的有参层来直接学习残差比直接学习输入、输出间映射要容易得多（收敛速度更快），也有效得多（可通过使用更多的层来达到更高的分类精度）。

当下 Resnet 已经代替 VGG 成为一般计算机视觉领域问题中的基础特征提取网络。当下 Facebook 乍提出的可有效生成多尺度特征表达的 FPN 网络也可通过将 Resnet 作为其发挥能力的基础网络从而得到一张图片最优的 CNN 特征组合集合。

若将输入设为 X，将某一有参网络层设为 H，那么以 X 为输入的此层的输出将为 H(X)。一般的 CNN 网络如 Alexnet/VGG 等会直接通过训练学习出参数函数 H 的表达，从而直接学习 X->H(X)。

而残差学习则是致力于使用多个有参网络层来学习输入、输出之间的参差即 $H(X)-X$即学习$X \to (H(X)-X)+X$。其中X这一部分为直接的识别映射，而 $H(X)-X$则为有参网络层中输入输出之间的残差。

2.1.7 Fast R-CNN 卷积神经网络

Fast R-CNN 借助多任务损失函数，实现了同时识别目标和位置修正两大功能，使得研究者不再需要对网络分布训练。Fast R-CNN 独特的架构（如图 2-6）使得研究者在训练该模型的过程中不需要大量的内存来存储中间数据，极大地降低了卷积神经网络对计算机硬件的硬性要求。图 2-6 中，卷积-池化层的功能与上文讲述的 LeNet-5 的功能类似，具有自动提取输入图像特征、降低卷积神经网络的参数数目以及较少输入数维数的作用。该结构识别的图像特征包括了要识别的目标特征。当前，大多数研究者利用选择搜索算法对自动识别输入图像的目标区域（TF）。一般情况下，TF 包含两部分内容：（1）要识别的目标；（2）目标背景或与目标紧密的相关的其它物体。该算法提取的目标区域的大小具有多样性，然

而该结构的全连接层对输入的特征向量的维数具有不变性的要求。因此，该模型将 LeNet-5 模型最后一个池化层替换为兴趣池化层，用来将不同位数的特征图统一化，即生成维数相同的特征向量，最后输入每一个候选区域特征的类型得分。

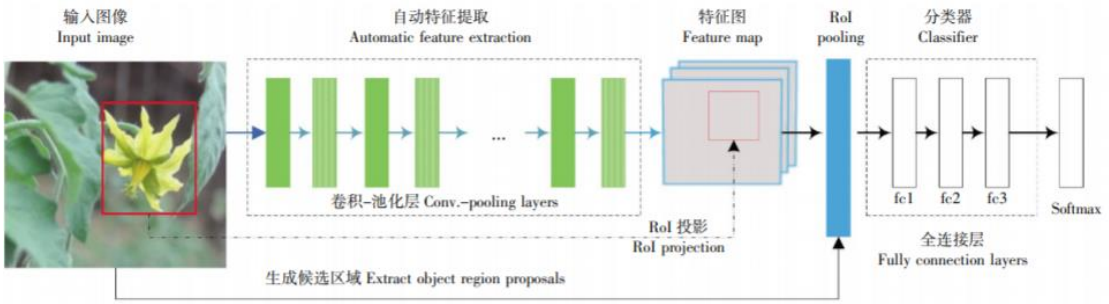

图 2-6 Fast R-CNN 卷积神经网络架构

Figure 2-6 The Architecture of Fast R-CNN

2.1.8 词袋模型

词袋模型最早是用来解决文档领域中的匹配问题，认为文档中每一个词都是独立出现的，也就是文档中词语与词语之间不具有关联性，将文档当成一个纯粹的填装词语的袋子，忽略了词语之间的顺序、语句的语法。该模型首先提取感兴趣的关键词，然后将文档表示成与顺序无关的关键词组合，最后利用统计的方法对文档中关键词出现的频率进行计算。该模型具有简单、易于理解的特点，研究者们将该模型成功地应用到自然图像的目标识别中并取得了较好的结果[16,17]，首先将图像中多个局部特征转化为视觉单词，然后将输入的每一幅图像表示为视觉单词的直方图。将词袋模型应用到图像的目标识别中主要包括提取和描述特征，创建字典，量化特征和训练分类器 4 个过程，具体来说：

1) **特征提取和描述：**一般情况下，输入图像的内容具有多元化与多样化的特征，识别图像中的目标首先需要将目标中图像中区分并且提取出来。因此，图像特征提取和描述是图像分类以及图像目标识别任务中的第一步，该步骤主要任务是提取图像最核心的特征，删除无关元素，用提取的特征来表示输入的图像。最主要的图像特征提取与描述的传统技术是 SIFT 描述子。为了提高图像特征提取与描述的精度，本文首先利用稠密采样的方法，提取输入图像中的局部特征，然后再利用 SIFT 描述子提取图像的特征。

2) **视觉词典生成：**在上一步的基础上，得到了输入图像大量的特征，生成视觉词典的过程就是对特征域分区的过程，每一个分区的中心就是所谓的视觉单词。

分区的方式有很多，常用的技术为 K-means 聚类。当对特征域分区之后，新的特征利用最近邻方法技术到每一个视觉单词的最近距离，并认为该视觉单词可以表达该特征。根据上述分析，可以得到视觉单词的数量与 K-means 的聚类中心数目相同。

3) **特征量化：**在量化特征的过程中，最近邻算法用来计算输入的特征图与所有视觉单词之间的距离的最小值，并得到可以表达该输入特征的视觉单词，然后利用统计学的知识，得到该视觉单词出现的数目，并绘制相应的直方图。传统的机器学习的分类算法中，支持向量机（SVM）是广受欢迎的分类器，在卷积神经网络出现前，该分类器是被研究最多的分类器。SVM 希望找到一个超平面，并通过该超平面对不同的数据进行分类，其目标函数可以表示为：

$$\min_{w,\xi}\{\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i\}$$

约束条件为：

$$y_i(w\cdot x_i-b)\geqslant 1-\xi_i,\ \xi_i\geqslant 0,\ \text{任意的}\ i=1,2,\cdots,n。$$

其中w是超平面的法向量，c为惩罚因子，ξ_i为松弛向量。本文采用 SVM 进行分类，选用径向基核函数。

2.2 国内外研究现状

本节介绍相关技术的研究现状，包括：目标识别和卷积神经网络。

2.2.1 目标识别的研究现状

传统的目标识别方法通常需要设计人员设计复杂的特征，需要借鉴大量的领域知识并经过巧妙设计后，在特定情况下取得了较好的效果，但这些人工设计特征无法充分体现图像中的具体含义与复杂的高层语义信息，使得特征的抽象和泛化能力较弱。

近年来，大数据为科学发展提供了大量、多元的数据，云计算技术为这些大量的数据提供了足够的存储空间，研究者利用不同的技术与原理来充分利用这些技术带来的益处，推动社会不断发展，经济不断提高。其中深度学习最近收到了人们的广泛关注，卷积神经网络在世界级各大竞赛以及实际的应用过程中表现良好，受到研究者与工业界的普遍认可。

在图像识别领域内，传统的技术需要大量的人工参与，例如人工设计或者提

取研究对象的特征，模型参数调优等，极大地限制了图像分类以及图像目标识别技术的准确性。卷积神经网络的出现，为该领域带来了曙光，该模型不需要人工提取图像特征，并且参数的调整过程是完全自动化的。该模型首先将输入的图像经过卷积层、池化层中的相关、重复操作，将图像中的关键信息提取成图像特征，并用这些提取出来的关键的特征对图像进行分类或者识别图像中的目标。从上文的论述中可以看出，卷积神经网络在图像的分类与图像目标识别的任务中训练模型阶段与与测试阶段不需要人工参与，并且具有计算速度快（参数共享、降采样等级制）鲁棒性强的优点。目前，高精度的图像目标识别以及图片分类任务，卷积神经网络已经成为最重要的关键技术[16]。传统的图像识别技术需要利用图像特征描述子 SIFT 提取图像的特征，然而该技术具有提取特征不准确并且时间开销比较大的缺点。基于上述问题，研究者相继提出了一些列的改进算法：$PCA-SIFT, Color-SIFT, GLOH$[17-19]等。传统的图像分类与图像目标识别的技术利用描述子提取了大量的特征之后，需要用其它技术手段对提取的特征进行分类，即简历图像的视觉词典，然后在视觉词典的基础上，计算图像的特征与词典中视觉单词的距离，得到与目标距离最近的视觉单词，最后统计图像中视觉单词的数目，并以直方图的形式表现出来。但是该方式不能提取图像的抽象含义并且在模型搭建的过程中需要经验丰富的工程师利用大量的领域知识进行建模，不仅速度慢，准确率也不高。研究者为了得到更加精准的图像特征，提出了一系列的技术，包括再统计、编码等[20-22]。1997 年 Joachims 等人认为文档中词语与词语之间不具有关联性，将文档当成一个纯粹的填装词语的袋子，提出了词袋模型(BoW)。2003 年 Sivic 等人将图像当做特殊的“文档”，用来进行图像分类与图像目标的识别。该模型首先提取感兴趣的关键词，然后将文档表示成与顺序无关的关键词组合，最后利用统计的方法对文档中关键词出现的频率进行计算。该模型具有简单、易于理解的特点，研究者们将该模型成功地应用到自然图像的目标识别中并取得了较好的结果[16,17]，首先将图像中多个局部特征转化为视觉单词，然后将输入的每一幅图像表示为视觉单词的直方图。将词袋模型应用到图像的目标识别中主要包括提取和描述特征，创建字典，量化特征等过程。随后，大量的研究者在词袋模型的基础上做了改进：Lazebni 等人在词袋模型的基础上引入了金字塔模型[23]；Yang 等人在 Lazebni 的工作基础上增加了稀疏编码理论，提出了

基于稀疏编码的金字塔模型ScSPM。

2.2.2 卷积神经网络在图像识别中的发展

2014 年 Girshick 等人在卷积神经网络的基础上，增加了线性回归、支持向量机等技术，提出了一种全新的卷积神经网络架构 R-CNN，该结构是第一个将卷积神经网络运用到图像的目标识别中。传统的图像识别技术需要利用图像特征描述子 SIFT 提取图像的特征，然而该技术具有提取特征不准确并且时间开销比较大的缺点。在模型训练阶段人工的模型参数调优的方式，极大地限制了传统图像分类以及图像目标识别技术的准确性。R-CNN 模型不需要人工提取图像特征，并且参数的调整过程是完全自动化的。该模型首先将输入的图像经过卷积层、池化层中的相关、重复操作，将图像中的关键信息提取成图像特征，并用这些提取出来的关键的特征对图像进行分类或者识别图像中的目标，具有高度的自动化、识别准确度高的特点，很快成为目前图像目标识别领域的主流技术。

基于卷积神经网络的图像目标识别技术首先将输入的图像划分分区，即将输入的图像划分为不同的兴趣区，然后将兴趣区进行裁剪并输入到卷积神经网络中，接着在卷积层与池化层提取高纬度、抽象的图像特征，最后利用支持向量机对兴趣区进行分类。在该类模型中兴趣区的分类结果视为目标的类别，对应的位置 Wie 目标位置，相关的算法包括：R-CNN、Fast R-CNN 和 Taster R-CNN 等。R-CNN 在对输入的图像进行兴趣区划分时，通常产生 2000 多个区域，极大地增加了模型的训练与测试开销。在 R-CNN 的基础上，Fast R-CNN 在 R-CNN 的基础上增加了区域池化层，减少了大量的重复计算，大大降低了模型训练与测试的时间。

2.3 小结

本章首先介绍了卷积神经网络的基本概念、开源框架，然后介绍了卷积神经网络、目标识别以及卷积神经网络在图像识别领域中的研究工作。

第3章　基于卷积神经网络的高分辨率遥感图像识别方法

本章首先分析卷积神经网络运用到高分辨率图像的必要性，然后介绍提出的基于卷积神经网络的高分辨率遥感图像目标识别技术，包括方法框架、算法。

3.1　卷积神经网络识别遥感图像目标的必要性

遥感技术综合了空间学、电子学、光学以及计算机科学的最新成果，是当代高新技术的一个重要代表。从 1983 年摄影技术发明到如今，遥感技术用于勘测、军事、防灾等各个领域，并取得了显著的成效，使得遥感技术得以快速发展。

随着空间技术、无线电技术、计算机技术、光学技术以及材料技术的不断提高，遥感技术不断发展。遥感技术获取的信息可以用于各个领域。通过拍摄地球的云圈，促进了大气探测领域的发展，为世界人民了解地球大气圈、陆地、海洋以及生物的分布起到了核心的作用。除此之外，遥感技术一般是从高空中获取地面相关事物的图像信息。这种获取信息的方式具有：获取信息速度快、信息的范围大、获取信息的方式限制低以及手段多的特点。基于上述特征，遥感技术受到工业界和学术界的广泛关注，目前我们正在进入一个智能化时代（互联网+及工业 4.0 时代），大数据、云计算和人工智能已经成为这个时代进步的三驾马车，它们分别为智能化时代提供数据、算力和算法层面的支持，从而成为各行各业技术革新和社会发展的重要引擎。

在智能化时代的背景下，云计算为遥感技术提供了大量的存储空间，使得海量高分辨率遥感图像的存储成为可能。研究者可以从大量存储的数据中识别、提取感兴趣的信息，为国家领土安全、城市规划与建设、灾害的监控与预防、国家经济发展的等领域带来巨大的益处，具体来说：（1）在国家领土安全方面：从高分辨率遥感图像上可以识别各种军事武器，如航母、军舰、基地等关键军事目标，从而获取敌军的战略信息；（2）在城市规划与建设方面：通过识别遥感图像获取城市的街道、绿化面积和人员活动等信息，辅助城市规划工程师制定合理的城市建设方案；（3）在灾害监测与预防方面：从高分辨率的遥感图像中获取灾区的受灾面积，受灾类型，周围的环境状况等关键信息，辅助制定急求方案与运输急求物资；（4）在研究国家经济发展领域：通过遥感技术获取与国家发展相关的主要信息，如夜晚的灯光分布信息，从而了解国家不同地区的经济发展情况和比较不

同国家的经济发展状况。

基于上述分析，遥感图像识别是运用遥感技术解决各种问题的关键步骤。高分辨遥感图像中的目标具有多样性与复杂性，具体来说：（1）在多样性方面：搭载在航空、航天等设备上的仪器拍摄的遥感图像由于拍摄角度的原因，图像中物体呈现出事物种类多、事物之间联系复杂，既包含要识别的目标，又包含诸多干扰物体，如山川、河流、各种建筑物等；（2）在复杂性方面：目标与目标之间、目标与非目标之间往往存在复杂的联系，且呈现出相互影响的特征。该分辨率遥感图像的这些特点为识别图像中的目标带来了挑战。传统的遥感图像识别技术（如：数字图像处理软件、统计模式识别方法等）存在严重依赖人工解译，解译精度较低等诸多问题。因此，如何高效、准确地识别高分辨率遥感图像中的目标是一个亟待解决的关键问题。

卷积神经网络在自然图像处理以及目标识别领域具有优秀的表现[5]。另一方面，随机计算机技术以及光学原理的不断发展（表 1-2 展示了近年来我国卫星拍摄遥感图像分辨率的变化），高分辨率遥感图像与自然图像的差别不断缩小，这使得卷积神经网络运用到遥感图像的目标识别中成为可能。与自然图像相比，高分辨率遥感图像呈现出图片尺寸较大、目标的拍摄角度多变、光影变化等特点，给卷积神经网络运用到高分辨率遥感图像的目标识别上增加了阻碍。如何正确地将卷积神经网络运用到高分辨遥感图像的目标识别中，自动、高效地识别目标具有重要的研究价值。

3.2 卷积神经网络识别遥感图像的框架

高分辨率遥感图像分类是遥感技术的一个热门研究领域，首先需要通过一定的手段获取图像的特征（体现图像的主要信息），并通过一定的技术手段基于提取的特征对图像进行分类。传统的高分辨遥感图像需要在研究者建模的基础上，提取图像的特征，该方式具有明显的主观因素。词袋模型是一种典型方法。然而提取的特征直接影响图像分类的精度，因此研究者利用具有强大学习能力的卷积神经网络自动地提取特征，并进行分类。本节主要建立词袋模型以及卷积神经网络在高分辨率遥感图像分类任务中的模型

3.2.1 基于词袋模型的高分辨率图像目标识别框架（B-SMI）

词袋模型最早是用来解决文档领域中的匹配问题，认为文档中每一个词都是

独立出现的，也就是文档中词语与词语之间不具有关联性，将文档当成一个纯粹的填装词语的袋子，忽略了词语之间的顺序、语句的语法。该模型首先提取感兴趣的关键词，然后将文档表示成与顺序无关的关键词组合，最后利用统计的方法对文档中关键词出现的频率进行计算。该模型具有简单、易于理解的特点，研究者们将该模型成功地应用到自然图像的目标识别中并取得了较好的结果[16,17]，首先将图像中多个局部特征转化为视觉单词，然后将输入的每一幅图像表示为视觉单词的直方图。将词袋模型应用到图像的目标识别中主要包括提取和描述特征，创建字典，量化特征三个过程，具体来说：

1) **特征提取和描述：**一般情况下，输入图像的内容具有多元化与多样化的特征，识别图像中的目标首先需要将目标中图像中区分并且提取出来。因此，图像特征提取和描述是图像分类以及图像目标识别任务中的第一步，该步骤主要任务是提取图像最核心的特征，删除无关元素，用提取的特征来表示输入的图像。最主要的图像特征提取与描述的传统技术是 SIFT 描述子。为了提高图像特征提取与描述的精度，本文首先利用稠密采样的方法，提取输入图像中的局部特征，然后再利用 SIFT 描述子提取图像的特征。

2) **视觉词典生成：**在上一步的基础上，得到了输入图像大量的特征，生成视觉词典的过程就是对特征域分区的过程，每一个分区的中心就是所谓的视觉单词。分区的方式有很多，常用的技术为 K-means 聚类。当对特征域分区之后，新的特征利用最近邻方法技术到每一个视觉单词的最近距离，并认为该视觉单词可以表达该特征。根据上述分析，可以得到视觉单词的数量与 K-means 的聚类中心数目相同。

3) **特征量化：**在量化特征的过程中，最近邻算法用来计算输入的特征图与所有视觉单词之间的距离的最小值，并得到可以表达该输入特征的视觉单词，然后利用统计学的知识，得到该视觉单词出现的数目，并绘制相应的直方图。传统的机器学习的分类算法中，支持向量机（SVM）是广受欢迎的分类器，在卷积神经网络出现前，该分类器是被研究最多的分类器。SVM 希望找到一个超平面，并通过该超平面对不同的数据进行分类，

4） **图像特征表达**:遥感图像特征表达在高分辨率遥感图像分类中处于重要地位，基于输入的高分辨率遥感图像提取到的高分辨率遥感图像的特征往往具有冗余、

噪声等其它因素，不利于对高分辨率遥感图像进行分类，影响高分辨率遥感图像分类的精度。该步骤的提出主要是将提取到的高分辨率遥感图像特征映射成固定长度的向量，表达高分辨率遥感图像的特征，剔除特征中的糟粕，用于后续的高分辨率遥感图像分类，提高高分辨率遥感图像的分类精度。因此，该步骤直接关乎高分辨率遥感图像分类的准确度。基于视觉单词的数量，高分辨率遥感图像的特征表达可以有两种方式：（1）硬量化，即向量量化。该方式广泛应用于传统的高分辨率遥感图像分类任务中，首先研究者需要统计视觉词典中每一个视觉单词在高分辨率遥感图像中出现的频率，并构建直方图，然后向量量化编码只在最近的视觉单词上相应为 1，并具有累加作用。该方式具有直观、简单、易于实现的特点；（2）软量化。研究者在实际的任务场景中发现了高分辨率遥感图像具有模糊的特性，即多个视觉单词的距离可能非常近。在这种情形下，如果采用硬量化的方式将导致多个能够表达高分辨率遥感图像特征的视觉单词被忽略，拉低了高分辨率遥感图像分类的精度。研究者在硬量化的基础上提出了软量化，用多个距离最近的视觉单词表达一个图像特征，该方式具有一定的实际意义。后来研究者在生物体的研究中发现，生物的细胞在大部分时间是不被激活的，即如果将一个细胞的生命周期视为一个时间轴，将激活状态视为时间轴上的突变，那么该时间轴具有凹凸性和稀疏性。基于这样的观察，研究者提出；稀疏编码的方法，该方法在最小二乘的基础上增加了稀疏约束，实现了在一个完备基上的稀疏性，并且在解决实际的问题过程中，利用该方法得到的特征向量，在简单的现行分类器上就得到了很好的实验结果。因此，该方法一度成为研究者在高分辨率遥感图像分类任务中的首选量化方式。然而，研究者在实际的任务中发现，该方式容易将相似的图像特征映射到不同的视觉单词，对高分辨率遥感图像分类的精度产生影响。为了解决这一难题，研究者提出了局部线性约束的方式。接下来对这种编码方式做详细地介绍。该方式的目标函数如下：

$$\min_{s} \sum_{i=1}^{zn} \| x_i - Bc_i \|^2 + \lambda \| d_i \odot c_i \| \quad ..$$
$$\text{s.t.} \quad 1^T c_i = 1, \forall i$$

其中B为输入的高分辨率遥感图像对应的视觉词典，c_i是特征向量系数，⊙表示

内积运算，d_i为词典中每一个视觉单词的自由度，用以下公式计算：

$$d_i = \exp(\frac{dist(x_i, B)}{\sigma})$$

其中，$\boldsymbol{d_i(x_i, \mathbf{B}) = [dist(x_i, b_1), dist(x_i, b_2), \ldots, dist(x_i, b_K))]}$是$\boldsymbol{x_i}$和$\boldsymbol{b_j}$的欧氏距离。

基于拉格朗日算子，目标函数的解为：

$$c_i = [C_i + \lambda diag(d_i^2)] / \mathbf{1}$$

$$c_i = \frac{c_i}{\mathbf{1}^T c_i} \quad \cdots$$

其中，diag(·)表示对角矩阵；$\boldsymbol{C_i = (1x_i^T - \mathbf{B})(\mathbf{B} - 1x_i^T)}$为协方差矩阵。

3.2.2 基于卷积神经网络的高分辨率目标识别框架与算法（B-SMI）

本文基于卷积神经网络的基本流程与高分辨率遥感图像的特点，提出了基于卷积神经网络的高分辨率遥感图像目标识别框架（图 3-2）。高分辨率遥感图像目标识别分为两个步骤：图像特征提取和分类。基于卷积神经网络的基本原理，在识别高分辨率遥感图像目标的过程中，首先卷积层的卷积核与输入的高分辨率遥感图像作用，得到特征图，然后将特征图作为输入前馈到下一层，并且利用误差反向传播的算法更新权重，得到特征向量集，最后利用支持向量机（SVM）对高分辨率遥感图像进行分类并统计结果。

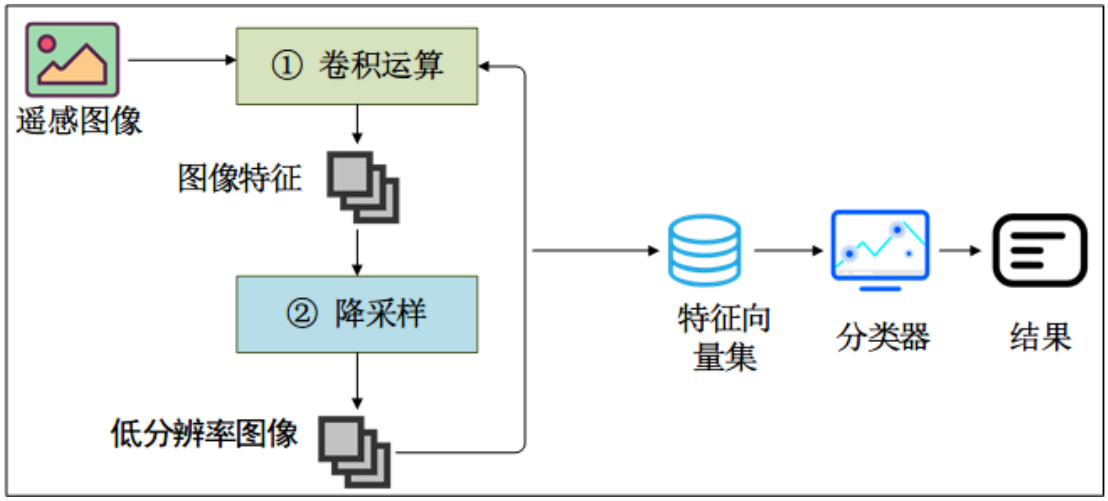

图 3-2 基于卷积神经网络的高分辨率遥感图像目标识别框架

Figure 3-2 High-Resolution Remote Sensing Image Target Recognition Framework Based on Convolutional Neural Network

高分辨遥感图像中呈现出信息的多样性与复杂性，具体来说：(1)在多样性方面：搭载在航空、航天等设备上的仪器拍摄的遥感图像由于拍摄角度的原因，图像中物体呈现出事物种类多、事物之间联系复杂，既包含要识别的目标，又包含诸多干扰物体，如山川、河流、各种建筑物等；(2)在复杂性方面：目标与目标之间、目标与非目标之间往往存在复杂的联系，且呈现出相互影响的特征。该分辨率遥感图像的这些特点为识别图像中的目标带来了挑战。传统的遥感图像识别技术（如：数字图像处理软件、统计模式识别方法等）存在严重依赖人工解译，解译精度较低等诸多问题。因此，如何高效、准确地识别高分辨率遥感图像中的目标是一个亟待解决的关键问题。卷积神经网络在自然图像处理以及目标识别领域具有优秀的表现[5]。另一方面，随机计算机技术以及光学原理的不断发展（表 1-2 展示了近年来我国卫星拍摄遥感图像分辨率的变化），高分辨率遥感图像与自然图像的差别不断缩小，这使得卷积神经网络运用到遥感图像的目标识别中成为可能。与自然图像相比，高分辨率遥感图像呈现出图片尺寸较大、目标的拍摄角度多变、光影变化等特点，给卷积神经网络运用到高分辨率遥感图像的目标识别上增加了阻碍。接下来对架构中的每一部分做详细解释：

卷积运算：卷积层的卷积核、滤波器通常为较小尺寸的矩阵，比如 3×3、5×5 等，数字图像是相对较大尺寸的 2 维（多维）矩阵。滤波器在图像上滑动，对应位置相乘求和；卷积则先将滤波器旋转 180 度（行列均对称翻转），然后使用旋转后的滤波器进行相关运算。在实际的问题中，输入的高分辨遥感图像具有多维度的特征（如图 3-3），卷积时，仍是以滑动窗口的形式，从左至右，从上至下，3 个通道的对应位置相乘求和，输出结果为 2 张 4×4 的特征图。一般地，在卷积神经网络中要求卷积核的通道数要与输入的图像的通道数相同，即当输入为$m*n*c$时，卷积核为$s*s*c$。卷积过程是在图像每个位置进行线性变换映射成新值的过程，将卷积核看成权重，若拉成向量记为 w，图像对应位置的像素拉成向量记为 x，则该位置卷积结果为 $y=w' x+b$，即向量内积+偏置，将 x 变换为 y。从这个角度看，多层卷积是在进行逐层映射，整体构成一个复杂函数，训练过程

是在学习每个局部映射所需的权重，训练过程可以看成是函数拟合的过程。

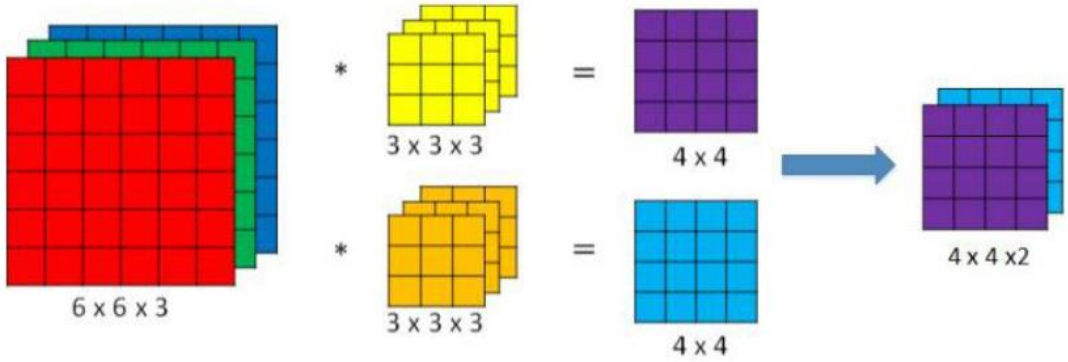

图 3-3 多维图像的卷积过程

Figure 3-3 Convolution Process of Multidimensional Images

1） **降采样**：局部感受野和权值共享有效地减少了卷积神经网络中的参数数目，极大地降低了模型的训练与测试时间。但是当输入图像比较大时，仍会降低模型的训练与测试速度。假设输入的图像经过第一个卷积层之后得到$90 * 90 = 8100$个特征图，每一个特征图与输入的图像进行卷积得到$91 * 91 = 8281$维的卷积特征，一共存在 8100 个特征图，总计$66{,}248{,}000$维的特征向量。如此庞大的维数不仅需要大量的时间，还容易出现过拟合的现象。为了缓解这一问题，研究者提出了降采样的方法，即将特征图映射成一个小尺寸的特征图，但是扔保留了特征图中的信息，相当于对特征图进行了抽象。由此，在提取输入图像特征信息的基础上不仅降低了卷积神经网络的参数数目，还降低了输入的维数。因此该层的主要目的是（1）使得高分辨率遥感图像特征符合下一层的输入尺寸；（2）生成对应该遥感图像的缩略图。

2） **分类器**：卷积以及降采样综合考虑了训练、测试的时间与高分辨率遥感图像的分类精度，并产生了大量图像的特征向量。C-SMI 的分类器主要是将得到的大量特征向量进行分类。一般而言分类的方式有多种，总体上可以分为两大类：（1）基于联合概率的分类模型主要是通过数据学习数据与数据之间的相似性；（2）基于条件概率的分类模型主要是判断数据之间的不相似性。其中广为人知的支持向量机就是基于条件概率的分类模型。支持向量机被研究者广泛研究，在实际的任务中表现良好，并且具有简单、易于实现的特点。本文使用支持向量机对得到的大量特征向量进行分类。假设样本集为$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$，支持向量机希望找到一个超平面将样本分割成两部分（如图 3-4）。在样本中找到一

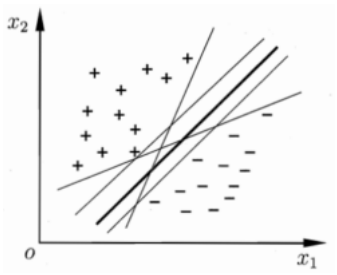

图 3-4 简单支持向量机实例

Figure 3-4 The Example of SVM

个超平面的过程就是一个求解一下方程的过程：

$$w^T x + b = 0,$$

其中，$\mathrm{w} = \{w_1; w_2; \ldots, w_d;\}$为超平面的法向量，该向量决定了平面的方向；b是位移项，决定了超平面与圆点孩子间的距离。显然地，超平面可以通过w和b的值决定。为了找到一个最合适的超平面，首先需要计算每一个样本到超平面的距离 $\mathrm{r} = \frac{|w^T x+b|}{\|w\|}$，并使得所有样本距离超平面的距离最远，表示为

$$\max_{w,b} \frac{2}{\|w\|}$$

$$\text{s.t.} \quad y_i(w^T x + b) \geq 1$$

3.3 小结

本章讨论了如何将卷积神经网络与词袋模型应用到高分辨率遥感图像目标识别领域，提出了两种识别框架与相应的算法，其中，C-SMI 利用卷积神经网络来识别高分辨率遥感图像中的目标；B-SMI 利用词袋模型来识别高分辨率遥感图像中的目标。

第4章　经验研究

为了评估基于卷积神经网络的遥感图像目标识别的效率、有效性以及准确率，本章选择 UcMerced 数据集作为研究对象进行经验研究。

4.1　研究问题

本章实验围绕以下三个问题展开讨论：

1) 基于卷积神经网络的高分辨率遥感图像目标识别框架 C-SMI 和基于词袋模型的高分辨率遥感图像目标识别框架 B-SMI 场景分类的准确率。本文探究 C-SMI 和 B-SMI 在 UcMerced 和 SIRI-WHU 数据集上场景分类的准确率。

2) 基于卷积神经网络的高分辨率遥感图像目标识别框架 C-SMI 和基于词袋模型的高分辨率遥感图像目标识别框架 B-SMI 的目标检测准确率。本文探究 C-SMI 和 B-SMI 在 UcMerced 和 SIRI-WHU 数据集上的目标识别的准确率。

3) 验证 C-SMI 在高分辨率遥感图像的目标识别准确率更高：本文通过经验研究的方式验证 C-SMI 比 B-SMI 在高分辨率高干图像的场景分类和目标识别任务中的表现。

4.2　实验对象

本节对详细地介绍本文的用到的数据集。

4.2.1　UC Merced 数据集

为了研究遥感图像托类与目标识别技术，很多国家基于不同的场景和设备获取了大量的遥感图像，并建立数据库，方便研究者进行研究。其中，美国建立了一个遥感图像集，包括 21 类图像，每一类的图像有 100 张，每一张图像的大小是固定的为$256 * 256$。该数据集具有数量多、分辨率高的特点，被广大研究者广泛使用。图 4-1 显示了数据集中的 21 个样本示例，依次为：农田、机场、棒球场、海滩、建筑物、丛林、密集住宅区、森林、高速公路、高尔夫球场、海港、十字路、中型住宅区、安置房、立交桥、停车场、河流、机场跑道、稀疏住宅区、储油罐和网球场。

4.2.2 SIRI-WHU 数据集

为了获得区分度更高的遥感图像，武汉大学建立了SIRI－WHU数据库，包括12 类样本，其中每一类样本包含 200 个图像，每一个图像的大小固定为200＊200。该数据集中的每一类样本示例展示在图 4-2，具体为农田、 商业区、 港湾、闲置土地、 工业区、草地、 立交桥、 公园、 池塘、 住宅区、 河流和水。

图 4-1 UC Merced 数据集样本

Figure 4-1 The Samples of UC Merced dataset

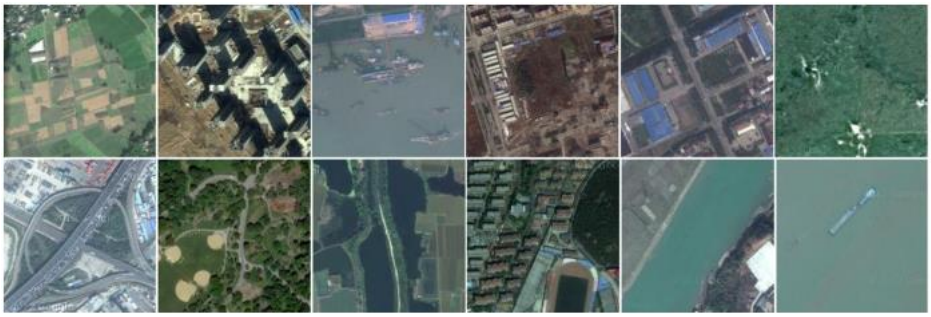

图 4-2 SIRI-WHU 数据集样本

Figure 4-1 The Samples of SIRI-WHU dataset

4.2.3 Google Earth 飞机数据集

本文不仅探究卷积神经网络在高分辨率遥感图像的分类中的表现，还探究卷积神经网络在高分辨率遥感图像目标识别中的表现情况。Google Earth包含了大量的飞机数据（如图 4-3），一共400幅图像，1294架飞机。该数据集中每一副图像的分辨率并不固定在300＊300－700＊700之间。该数据集中的飞机具有多场景、多姿态、数量大的特点。

图 4-3 Google Earth 飞机数据集样本

Figure 4-3 The Samples of Google Earth dataset

4.3 实验设置

本节主要介绍本文实验相关的设计细节。

4.3.1 待评估的技术

本文提出的 C-SMI 和 B-SMI 作为研究的对象。为了突出基于卷积神经网络作用于高分辨率遥感图像目标识别的有效性与准确率，本文评估与比较 C-SMI 和基于词袋模型（S-SMI）的有效性与效率。

4.3.2 度量标准

在机器学习领域中，分类和预测是主要的研究问题。然而任务场景的多样性导致以某一个技术适用于所有的情形。为了判断研究者所提技术的有效性，以下几个度量标准受到广泛的应用：（1）准确率：技术识别的目标中正确的目标占的比例；（2）召回率：识别的目标种类占所有目标种类的比率；（3）平均准确率：反应了某一个目标的全局性能；（4）平均准确率均值：反应了多个目标的全局性能。

对于研究问题（1），本文利用各个类别的平均准确率和所有类别的平均准确率均来衡量。

对于研究问题（2），本文利用平均准确率来衡量所提框架的有效性。此外，目标识别所需时间是衡量目标识别技术的重要指标，本文提出一种度量指标：T-measure，表示从没一幅图像中识别出目标的所需的时间。

对于研究问题（3），在问题（1）和（2）的基础上进行对比，得到不同技术的表现情况。

4.3.3 实验环境

本文使用 python 语言编写测试脚本， 运行在 Windos10 64 位操作系统上。该系统有 4 个 CPU 和 16GB 内存。

4.3.4 实验流程

本文的实验主要分为两个：基于 C-SMI 和 B-SMI 的高分辨率遥感图像的场景分类和基于 C-SMI 和 B-SMI 的高分辨率遥感图像的目标识别。接下来详细描述具体的实验流程。

1） **基于 C-SMI 和 B-SMI 的高分辨率遥感图像的场景分类：**为了训练 C-SMI 框架中的模型，调整参数，评估技术的有效性，本文将UcMerced数据集以及SIRI－WHU数据集分成三类：训练集、验证集合测试集。对于UcMerced数据集，本文将该数据集中的图像均分到三个子集中；对于SIRI－WHU数据集，本文设置训练集包含380幅图像，验证集包含190幅图像，测试集包含435幅图像。然后利用在ImagNet数据集中训练好的模型提取输入的高分辨率遥感图像特征，对于B－SMI模型中的词袋模型，本文利用LLC编码分别获取100、1000、5000、10000个视觉单词。最后利用支持向量机对两个模型提取的特征进行分类。

2） **基于 C-SMI 和 B-SMI 的高分辨率遥感图像的目标识别：**在使用AdaBoost算法对样本进行分类时，需要准备大量的样本，本文选择用的Google Earth数据集中的高分辨率遥感图像数量不足。为了得到充足的样本，本文在已有数据的基础上，对高分辨率遥感图像中的目标组旋转（90、180、270 度），得到1952幅高分辨率遥感图像作为正样本。由于负样本中不能包含目标，因此本文提取高分辨率遥感图像中不包含飞机的背景部分（共计15000个）作为负样本。然后本文利用 AlexNet 模型提取高分辨率遥感图像中的特征。

4.3.5 潜在的风险分析

影响本文实验结果的一个潜在风险是测试脚本实现的正确性。 明显地，不正确的测试脚本直接影响实验结果的正确性。测试脚本经过不同开发人员检查、修正， 因此相关测试技术的实验脚本都是正确的。另一个影响实验结果的风险是度量指标（平均准确率和 T-measure） 的可靠性。 单次实验结果不能有效地反映度量指标的真实值，本文重复 5 次实验保证度量指标在统计上的可靠性。

第5章　实验结果与分析

5.1　卷积神经网络的分类效果

每一个数据集中对象的准确率均值分别记录在表 5-1 和表 5-2 中。每一个数据集中所有对象所有重复实验的的结果分布展示在图 5-1 和图 5-2 中。 在盒图中，盒子的上边界以及下边界代表一个度量标准的上四分位数和下四分位数，中间的横线表示一个度量标准的中位数，实心圆点表示度量标准的均值，下“胡须”的最小值为下四分位数$-1.5*S$，上“胡须” 的最大值为上四分位数$+1.5*S$，其中S表示盒子的长度。不在两个“胡须”范围内的数值称为异常值，并用“*”表示。

表 5-1 UcMerced 数据集遥感影像场景分类结果

Table 5-1 The Results of the Scene Classification on UcMerced Dataset

对象	LLC1	LLC2	LLC3	LLC4	AlexNet (fc6)	AlexNet (fc7)	VGG16 (fc6)	VGG16 (fc7)	VGG16 (fc6)
农田	83.1	93.4	93.4	93.4	93.4	93.4	93.4	93.4	93.4
机场	75.4	75.4	75.4	75.4	75.4	75.4	75.4	75.4	75.4
棒球场	77.8	77.8	77.8	77.8	77.8	77.8	77.8	77.8	77.8
海滩	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6	77.6
建筑物	74.3	74.3	74.3	74.3	74.3	74.3	74.3	74.3	74.3
丛林	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4
密集住宅	83.6	83.6	83.6	83.6	83.6	83.6	83.6	83.6	83.6
森林	88.5	88.5	88.5	88.5	88.5	88.5	88.5	88.5	88.5
高速公路	78.5	78.5	78.5	78.5	78.5	78.5	78.5	78.5	78.5
高尔夫球场	77.1	77.1	77.1	77.1	77.1	77.1	77.1	77.1	77.1
海港	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1
十字路	83.3	83.3	83.3	83.3	83.3	83.3	83.3	83.3	83.3
中型住宅	81.0	81.0	81.0	81.0	81.0	81.0	81.0	81.0	81.0

区									
安置房	77. 4	77. 4	77. 4	77. 4	77. 4	77. 4	77. 4	77. 4	77. 4
立交桥	76.4	76.4	76.4	76.4	76.4	76.4	76.4	76.4	76.4
停车场	78.1	78.1	78.1	78.1	78.1	78.1	78.1	78.1	78.1
河流	79.5	79.5	79.5	79.5	79.5	79.5	79.5	79.5	79.5
机场跑道	80.2	80.2	80.2	80.2	80.2	80.2	80.2	80.2	80.2
稀疏住宅区	83.1	83.1	83.1	83.1	83.1	83.1	83.1	83.1	83.1
储油罐和网球场	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1	84.1

表 5-2 UcMerced 数据集遥感影像场景分类结果

Table 5-2

对象	LLC1	LLC2	LLC3	LLC4	AlexNet (fc6)	AlexNet (fc7)	VGG16 (fc6)	VGG16 (fc7)	VGG16 (fc6)
农田	66.3	66.3	66.3	66.3	66.3	66.3	66.3	66.3	66.3
商业区	72.1	72.1	72.1	72.1	72.1	72.1	72.1	72.1	72.1
港湾	73.1	73.1	73.1	73.1	73.1	73.1	73.1	73.1	73.1
闲置土地	76.4	76.4	76.4	76.4	76.4	76.4	76.4	76.4	76.4
工业区	66.3	66.3	66.3	66.3	66.3	66.3	66.3	66.3	66.3

草地	77.2	77.2	77.2	77.2	77.2	77.2	77.2	77.2	77.2
立交桥	69.3	69.3	69.3	69.3	69.3	69.3	69.3	69.3	69.3
公园	67.4	67.4	67.4	67.4	67.4	67.4	67.4	67.4	67.4
池塘	74.2	74.2	74.2	74.2	74.2	74.2	74.2	74.2	74.2
住宅区	75. 2	75. 2	75. 2	75. 2	75. 2	75. 2	75. 2	75. 2	75. 2
河流和水	79.2	79.2	79.2	79.2	79.2	79.2	79.2	79.2	79.2

图 5-1 UC Merced 分类结果

Figure5-1 The Results of the Scene Classification on UcMerced Dataset

图 5-2 SIRI WHU 分类结果

从表 5-1 和 5-2 可以看出，基于卷积神经网络的C－SMI框架比基于词袋模型的 B-SMI 框架分类的准确度高3%左右。其中 VGG-16 取得了最好的结果，准确率分别为 94.72%，95.12%和 94.56%。

5.2 目标识别的准确率

表 5-3 给出了选择搜索算法在一幅影像上生成的感兴趣区域数量以及处理一幅影像所需要的时间。从表中可以看出，

表 5-4 给出了 Fast R-CNN 模型在测试集上的结果。从表中可以看出，Fast R－CNN识别飞机的准确率率为 94.21%，识别率为 85.14%。说明该方法在高分辨率遥感图像中表现良好，具有非常大的发展空间。

5.3 小结

本章主要验证了提出了基于 CNN 的遥感图像分类模型(C-SMI)分类的准确性。C-SMI 在高分辨率遥感图像的基础上，结合 CNN 算法流程进行图像的分类，首先利用已有数据训练卷积神经网络，然后将高分辨率遥感图像输入到训练好的模型中，得到特征向量集，最后利用支持向量机对目标进行分类。实验数据表明该模型可以高精度的对高分辨率遥感图像进行分类。验证提出了基于词袋模型的遥感图像识别模型(B-SMI)的精度：基于词袋模型在文本领域的作用原理，本文将高分辨率遥感图像作为特殊的“文本”，高分辨率遥感图像的特征作为词语。该模型首先提取视觉特征，然后生成视觉字典，最后表达视觉特征。实验结果表明该框架不如 C-SMI 的分类精度。采用经验研究验证所提模型对遥感图像分类的准确性以及 Fast R-CNN 卷积神经网络在识别单个遥感图像目标的有效性：利用 UcMerced 和 SIRI-WHU 数据集，随机选取数据训练 CNN 模型(包括：AlexNet 和 GooLeNet)与词袋模型，并利用测试用例集评估所提模型的正确性与准确性，以及训练模型需要的时间；利用 Google Earth 数据集，检验 Fast R-CNN 卷积神经网络识别遥感图像单一目标的准确性。

第6章　工作总结与展望

遥感图像目标识别作为当前遥感图像应用领域中的主要研究内容，具有重要的理论研究意义与广泛的应用价值。遥感图像作为各种信息的综合体，具有内容复杂、数量多等特点。人工识别遥感图像作为一种传统的、广泛应用的图像识别技术无法满足大数据背景下的分类精度与时效要求。

深度学习是机器学习的一个重要分支，在实际应用中表现良好甚至达到了人类的水平。卷积神经网络作为一种广为研究与应用的深度学习模型，被认为是最强大的图像识别模型。

本文旨在利用 CNN 强大的图像识别能力自动识别遥感图像中的目标，提出一种基于 CNN 的遥感图像识别框架，且通过经验研究的方式验证所提框架的正确性与有效性。具体来说，本文的主要研究内容包括：

1) **提出了基于 CNN 的遥感图像分类模型(C-SMI)**：C-SMI 在高分辨率遥感图像的基础上，结合 CNN 算法流程进行图像的分类，首先利用已有数据训练卷积神经网络，然后将高分辨率遥感图像输入到训练好的模型中，得到特征向量集，最后利用支持向量机对目标进行分类。
2) **提出了基于词袋模型的遥感图像识别模型(B-SMI)**：基于词袋模型在文本领域的作用原理，本文将高分辨率遥感图像作为特殊的“文本”，高分辨率遥感图像的特征作为词语。该模型首先提取视觉特征，然后生成视觉字典，最后表达视觉特征。
3) **采用经验研究验证所提模型对遥感图像分类的准确性以及 Fast R-CNN 卷积神经网络在识别单个遥感图像目标的有效性**：利用 UcMerced 和 SIRI-WHU 数据集，随机选取数据训练 CNN 模型(包括：AlexNet 和 GooLeNet)与词袋模型，并利用测试用例集评估所提模型的正确性与准确性，以及训练模型需要的时间；利用 Google Earth 数据集，检验 **Fast R-CNN** 卷积神经网络识别遥感图像单一目标的准确性。

本文的所提的基于 CNN 的遥感图像识别模型，改进了传统识别遥感图像的精度，并且 CNN 在遥感图像目标识别任务重表现突出，具有很大的研究潜力。