

基于改进 CNN 特征的场景识别^①

薄康虎, 李菲菲, 陈虬

(上海理工大学 光电信息与计算机学院, 上海 200093)

通讯作者: 陈虬, E-mail: qchen1701@163.com

摘要: 随着人工智能的发展, 场景识别作为计算机视觉研究的重要方向之一, 吸引着越来越多研究者的关注. 由于传统的手工特征无法充分描述场景图像的信息导致效果不理想, 而卷积神经网络 (CNN) 提取的特征能够包含丰富的场景语义和结构信息, 因此就常见的体系结构而言, 本文选取 AlexNet 网络模型进行场景识别的研究, 分别从网络模型的深度、宽度、多尺度化提取以及多层融合考虑进行改进, 改进后在两个数据集上的识别率分别可达 92.0% 和 94.5%, 通过对比结果表明了本文方法的有效性.

关键词: 场景识别; 计算机视觉; 卷积神经网络; AlexNet

引用格式: 薄康虎, 李菲菲, 陈虬. 基于改进 CNN 特征的场景识别. 计算机系统应用, 2018, 27(12): 25-32. <http://www.c-s-a.org.cn/1003-3254/6684.html>

Scene Recognition Algorithm Using Advanced CNN Features

BO Kang-Hu, LEE Fei-Fei, CHEN Qiu

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: With the development of artificial intelligence, scene recognition has attracted more and more researchers' attention, which is one of the important directions of computer vision research. The traditional manual features cannot sufficiently describe the characteristics of the scene images, which leading to unsatisfied performance. On the contrary, the features extracted from Convolutional Neural Networks (CNN) contain rich semantics and structural information of the scene images. As one of the most common architectures, AlexNet network model is chosen in this study. By improving the following 4 aspects of the network: depth, width, multi-scale extraction, and multilayer fusion, the proposed approach achieves high accuracies of 92.0% and 94.5% on two publicly available datasets respectively, showing the superiority compared with other methods.

Key words: scene recognition; computer vision; Convolutional Neural Networks (CNN); AlexNet

随着信息时代的发展, 场景识别在很多图像 (视频) 处理任务中扮演着重要的角色. 然而, 由于同类间的相似性和异类间的差异性导致场景识别在计算机视觉领域的发展具有挑战性, 强有力的特征提取和准确的分类器成为攻克难关的重要关键所在.

传统的方法主要是提取低级特征用于贝叶斯或支持向量机 (SVM) 分类. 但是, 将低级特征直接进行映

射可能会引起更大的语义鸿沟^[1], 后来又将低级特征编码为中层语义信息进行解决上述问题. 其中广泛使用的编码方法有词袋模型 (Bag of Words)^[2]、空间金字塔模型^[3]、概率生成模型^[4]、稀疏编码^[5]等.

近年来由于计算硬件的发展, 卷积神经网络 (CNN) 在计算机视觉领域带来革命性的改革, 推动了场景识别的发展. 与传统的统计学方法相比, 神经网络

① 基金项目: 上海市高校特聘教授 (东方学者) 岗位计划 (ES2015XX)

Foundation item: Distinguished Professor (Oriental Scholar) Plan in Universities of Shanghai (ES2015XX)

收稿时间: 2018-05-12; 修改时间: 2018-06-04; 采用时间: 2018-06-19; csa 在线出版时间: 2018-12-03

无需对概率模型进行假设,具有极强的学习能力和容错能力,所以卷积神经网络不仅可以提高场景识别的准确性,而且还可以作为各种通用特征提取器进行识别任务,如目标检测、语义分割以及图像检索等。只要简单的预处理图像就可以直接作为输入,省去了复杂的特征工程,提高 CNN 的传输能力,因此卷积神经网络逐渐成为图像识别问题的重要工具,在图像理解领域中获得了广泛应用。

基于上述研究方法的考虑,本文采用卷积神经网络进行场景识别的研究,由于不同模型之间的特性和运行效率,故选择 AlexNet 模型作为基础模型,分别进行不同方式的改进提高场景识别率。

1 卷积神经网络

卷积神经网络模型是利用空间结构数据的特殊神经网络架构。一个标准的 CNN 包含三种特征操作层,即卷积层、池化层和完全连接层。当这些层被堆叠时,如图 1 所描述的那样就形成了 CNN 架构。

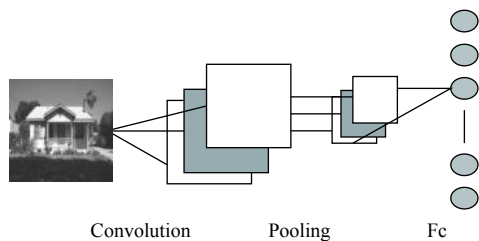


图 1 标准的卷积神经网络框架

1.1 卷积层

卷积层首先执行卷积操作来完成一系列的线性激活,然后每个线性输入到非线性激活函数中,例如 ReLU、tanh。在卷积层中,输入与一系列的卷积核进行卷积学习,卷积是一种在两种信号上的线性操作,例如:有两个函数 $x(t)$ 和 $\omega(t)$, t 为连续变量,移位值 a ,卷积操作可以定义为:

$$h(t) = (x * \omega)(t) = \sum_{a=-\infty}^{\infty} x(a)\omega(t-a) \quad (1)$$

其中, x 表示信号输入, ω 表示卷积核(滤波器),输出 h 表示特征映射, t 表示时间, a 表示移动的步长。

传统的神经网络每个输出都连接每个输入,而卷积神经网络拥有局部感受野,这意味着每个输出单元只连接到输入的一个子集,利用空间局部相邻单元之间的相关性进行卷积运算。CNN 的另一个显著特性是

参数共享,卷积层中使用的参数共享意味着每个位置共享相同的参数(权重和偏差),减少了整个网络的参数数量,提高了计算的效率。

1.2 池化层

池化层一般存在于每个卷积层之后来降低上一层卷积计算输出的维数,即将特征图像区域的一部分求个均值或者最大值,用来代表这部分区域。如果是求均值就是 mean pooling,求最大值就是 max pooling。常用的是最大值池化,最大值池化输出的是矩形邻域内的最大值,图 2 描述了具有 2×2 滤波器和步长为 2 的最大值池化,下采样表示空间大小,此外,池化操作可以保证图像的转移不变性。

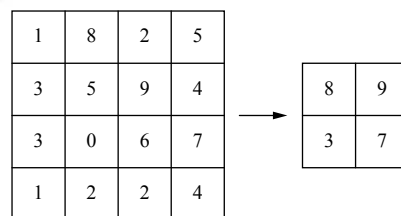


图 2 最大池化示例

1.3 全连接层

在完成多个卷积层和池化层之后,卷积神经网络通常采用全连接层结束学习,完全连接层中的每个神经元都完全连接到前一层中的所有神经元,在整个卷积神经网络中起到“分类器”的作用。

2 改进 CNN 网络模型

传统的卷积神经网络(CNN)稀疏的网络结构无法保留全连接网络密集计算的高效性和实验过程中激活函数的经验性选择造成结果不准确或计算量大等一系列待优化的问题,本章节以 AlexNet 作为基本的网络模型,分别从模型深度、宽度、多尺度提取以及多层特征融合方面进行改进提高场景识别的有效性。

2.1 基本原则

(1) 避免早期网络阶段的表达瓶颈问题

一般来说,在最终的表示之前从输入到输出表示的尺寸(特征映射分辨率)会缓慢降低,步幅的减少可能会减缓特征映射分辨率的下降并减少图像信息的丢失,集成多分辨率特征来学习更多空间信息从而克服表达的瓶颈问题。

(2) 平衡模型的深度、宽度以及卷积核大小的关系

随着深度的加深,宽度和卷积核大小需要适当的调整.表1显示了卷积核大小和步长之间的几个经验值.深度和宽度是调节网络参数的两个关键因素.尽管模型的深度越来越深,模型性能的显著提高以及计算成本的加剧,但是深度依然被认为是设计 CNNs 架构时的最高优先级^[6-8],另外,减少滤波器(卷积核)的数量只能部分地补偿参数过多的问题.过多的参数最终会导致过度拟合并限制分类精度的提高,因此深度是考虑优化的第一要务,但不是唯一因素.

表1 卷积核大小和步长的经验值

Kernel size	Stride
11×11	4
7×7	2
3×3	1

(3) 降低模型复杂度

从神经元数量、模型参数尺度以及所有卷积层的时间复杂度三个方面考虑深度卷积神经网络的模型复杂度,综合评估一个表现良好的网络模型需要考虑这三个方面.

① 时间复杂度

$$Times \sim O(M^2 \cdot K^2 \cdot C_{in} \cdot C_{out}) \quad (2)$$

其中, M 代表每个卷积核输出特征图 (Feature Map) 的边长; K 代表每个卷积核 (Kernel) 的边长; C_{in} 代表每个卷积核的通道数; C_{out} 代表本卷积层具有的卷积核个数,即输出通道数.

由此可见,每个卷积层的时间复杂度由输出特征图面积 M^2 、卷积核面积 K^2 , 输入 C_{in} 和输出通道数 C_{out} 完全决定.

输出特征图尺寸本身由输入矩形尺寸 X , 卷积核尺寸 K , $Padding$ 以及 $Stride$ 四个参数所决定,表示如下:

$$M = (X - K + 2 * padding) / Stride + 1 \quad (3)$$

卷积神经网络整体的时间复杂度可以表示为:

$$Times \sim O(\sum_{l=1}^D M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l) \quad (4)$$

其中, D 表示神经网络所具有的卷积层数; l 代表神经网络第 l 个卷积层; C_l 表示神经网络第 l 个卷积层的输出通道数 C_{out} .

② 空间复杂度 (模型的参数尺度)

空间复杂度即模型的参数数量,体现为模型本身的体积,可表示为:

$$Space \sim O(\sum_{l=1}^D K_l^2 \cdot C_{l-1} \cdot C_l) \quad (5)$$

可见,网络的空间复杂度只与卷积核的尺寸 K 、通道数 C 、网络的深度 D 相关.而与输入数据的大小无关.

时间复杂度一方面决定了模型的训练/预测时间,如果复杂度过高,则会导致模型训练和预测耗费大量时间,既无法快速的验证想法和改善模型,也无法做到快速的预测;另一方面决定了模型的参数数量,由于维度限制,模型的参数越多,训练模型所需的数据量就越大,而现实生活中的数据集通常不会太大,这会导致模型的训练更容易过拟合.

③ 神经元数量

神经元是神经网络进行计算的重要单元之一,计算机模拟人的大脑去感知和认知世界,在某种程度上来说,神经元越多,学习能力必然会越强.但是,计算机不会像人脑一样同时可以进行多样任务的学习,庞大的神经元系统自然也就导致了一系列的问题,如计算效率下降、难以优化等.

2.2 增强的 AlexNet 网络模型

基于卷积神经网络的场景识别主要分三个主要部分,即图像预处理、卷积操作提取特征以及全连接层的分类.以 AlexNet 网络为基础模型进行改进, AlexNet 网络模型如图3所示,具体算法如下:

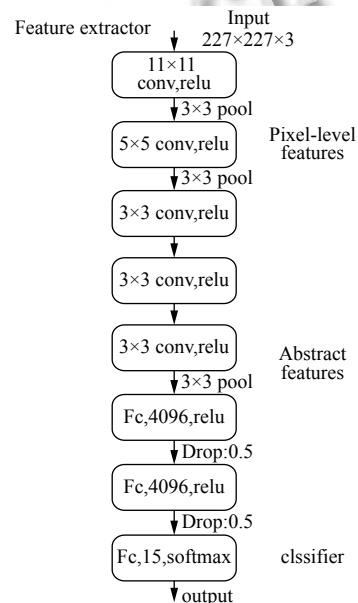


图3 AlexNet 学习框架

对于场景识别任务而言,空间布局信息是至关重要

要的, 卷积神经网络提取特征的过程中特别容易丢失空间布局的细节信息. 现实中一张场景图像包含多个目标场景, 其次大部分场景对象相对于其他对象保持在水平或者垂直方向, 如图4所示.

考虑到不同目标场景图像之间的冗余信息在特征提取过程中可能会有或多或少的影响, 本实验尝试采用图5所示的划分方式, 和原数据集组合为一个数据集, 这样一方面可以保留独立的个体目标的表现力, 另一方面可以达到数据增强的效果.

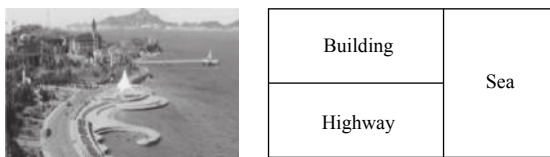


图4 场景图像的布局信息示例

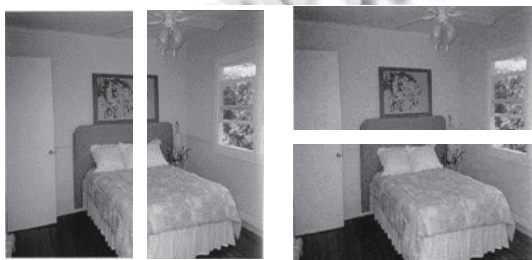


图5 场景图像分割示例

(1) 算法 1. 改变滤波器的数量.

有研究显示 AlexNet 网络模型第三、四层的特征提取能力最强, 故本文选择高层次 (第四层) 的卷积核进行改进, 对于特定的分类任务, 在一定范围内, 卷积核数量过大提取过多的冗余信息会降低最终的识别效率, 相反卷积核数量过小提取特征会表现得不充分, 故本文通过多次实验, 选择合适的卷积核数量进行改进, 实验结果如图6所示, 第四层滤波器数量取 388 识别率可达 91.5%, 效果最佳.

(2) 算法 2. 加深网络的深度.

网络模型的深度直接影响特征提取的表现能力, 网络模型越深, 提取的特征就会越抽象. 因此, 本文中尝试将 AlexNet 网络模型进行加深, 分别在第二层、第四层、第五层卷积层后面添加卷积层 (卷积核为 3×3), 其中 2-1 代表的是在第二层卷积层后面添加一组卷积核为 3×3 的卷积层, 如图7表示对最终识别任务的影响, 最高精度可达 90.0%, 前三种加深方法具有一定的提升, 后三种的深度增强了提取特征的冗余性, 降

低了最终的识别率.

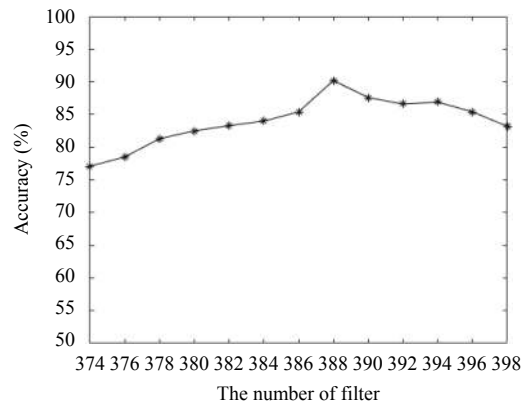


图6 滤波器数量对识别精度的影响

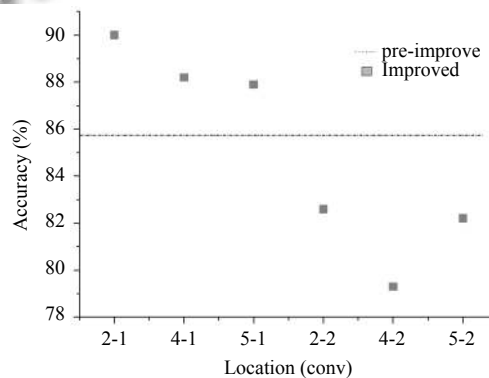


图7 网络模型深度对识别精度的影响

(3) 算法 3. 多尺度化特征提取.

获得高质量模型最保险的做法一般是增加模型的深度 (层数) 或者其宽度 (层核或者神经元数), 但是这样做法会引起三大共同的问题: ① 如果训练集有限, 参数过多会容易过拟合; ② 网络越大计算复杂度越大, 难以应用; ③ 网络越深, 梯度越往后传播越容易消失, 难以优化模型. 鉴于此, 本文通过不同的尺度卷积实现多尺度特征的利用, 多尺度化提取模块如图8所示.

此外本文还采用两个 3×3 的卷积核代替 5×5 的卷积核进行比较实验, 经过多次实验可知多尺度提取结构所在层数越高, 通道数越多, 该方法越能获得更高的效率. 因此在兼顾特征冗余性影响的同时, 分别在第五个卷积层 (conv5) 后加入 1, 2, 3 个多尺度化提取结构, 改进前后的识别率分别为: 88.6%, 82.4%, 73.6% 和 91.5%, 89.6%, 88.9%.

(4) 算法 4. 多层特征融合.

传统的 CNN 将输入图像层层映射, 最终得到特征提取的结果. 通过对 AlexNet 网络模型进行反卷积可

视觉可得,末端卷积层 Layer5 提取的特征具有完整的辨别能力,而卷积层 Layer2 主要映射的是具有颜色和边缘属性的特征,卷积层 Layer3 提取的主要是比较有区别性的纹理特征. CNN 主要采用的是最终提取的特征,卷积层 Layer2、卷积层 Layer3 对图像也具有一定的表现能力,故本算法将 Layer2、Layer3 卷积层进行多层再学习融合,具体学习框架如图 9 所示.

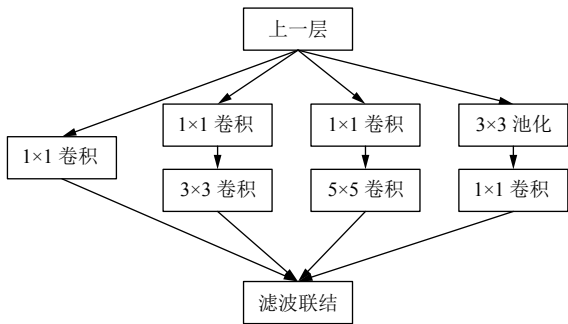


图 8 多尺度化提取结构

图 9 中,虚线框中对应的多层特征融合模块,实验步骤可以总结如下:(1) 首先采用 Fine-tuning 的方式训

练学习实线部分的框架;(2) 固定学习到的实线部分参数权重不变,如图 10,Block 块分别为无卷积层、Block1、Block2、Block3 四种情况进行学习;(3) 融合第 (1)、(2) 步骤学习到的特征,进行全连接然后送入分类器中. 经过多次实验,对应的四种情况识别精度分别为 82%, 85%, 87%, 92%.

综合上述四种算法,算法 1 和算法 2 为常见的 CNN 改进算法,本文主要就场景识别任务进行了验证,且都达到了一定的提升效果;算法 3 和算法 4 在此基础上对网络结构进行了改进,其中算法 3 对单层网络结构进行多尺度化特征学习,使得最终的表达能力更为丰富,识别精度可以达到与算法 1 等高的 91.5%,但此算法在泛化能力上明显优于前者;算法 4 使用多层特征融合的方式,考虑到不同特征之间的冗余性,采用不同的 Block 再学习方式,识别效果和算法 3 差距比较小,而且它充分利用了不同层次的特征,使得表现力更加多元化. 所以,下文将采用算法 3 和算法 4 的网络模型作为改进的 AlexNet 网络模型与原始 AlexNet 网络模型进行对比,并作相应的评测.

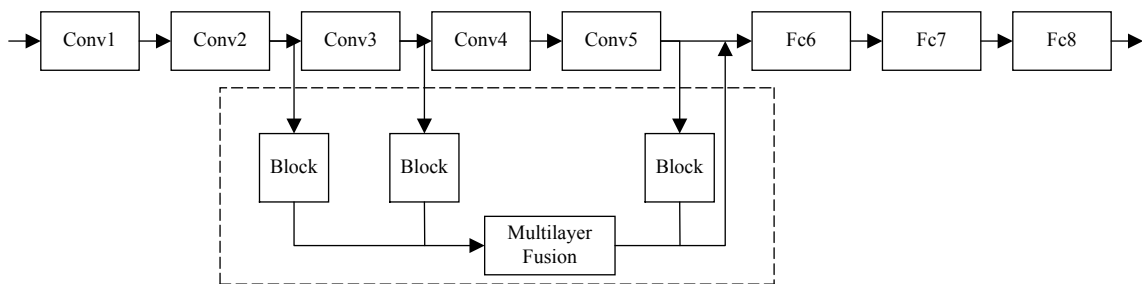


图 9 多层特征融合学习框架

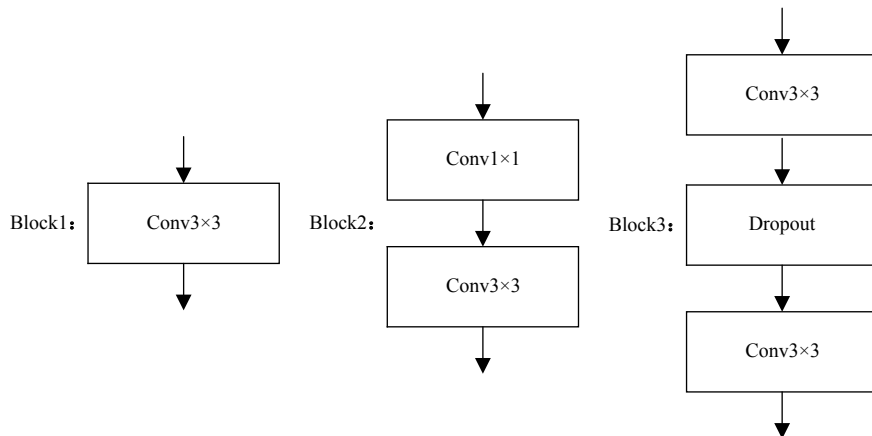


图 10 多层特征融合学习中 Block 块的选取

3 实验与分析

3.1 数据集

改进 CNN 识别能力分析需要选择合适的数据集. 深度卷积神经网络在训练过程中需要输入大量的图片进行数据处理, 因此实验所用的数据集应该包含丰富的场景信息. 其次, 所选的数据集应该具有一定权威性和公开性, 方便与之前研究方法进行比较.

(1) 数据集的选取

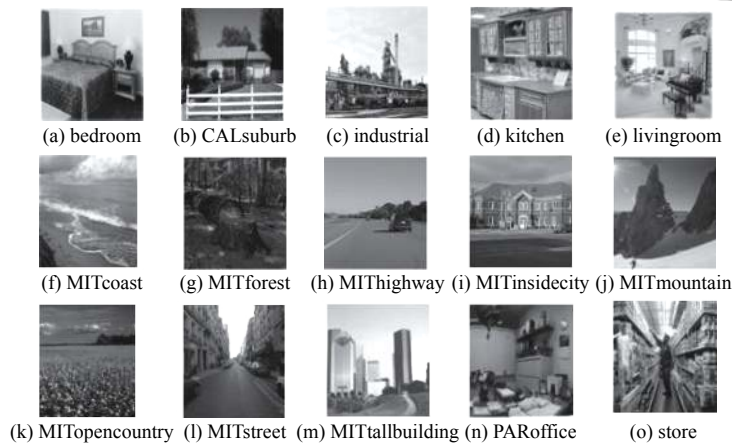


图 11 Scene15 数据集图像示例

(2) 图像预处理

改进 CNN 分类性能分析需要大量的数据集进行实验评估, 但是现实中数据集一般都比较少, 无法满足训练网络的要求, 所以一般进行实验前都对数据进行必要的增强操作, 比如旋转、翻转、移动、裁剪等一系列随机变换.

3.2 实验设置

在进行场景识别实验之前, 需要构建合适的网络模型学习框架, 现如今流行的深度学习框架主要有 Caffe、Torch、MXNet 以及 TensorFlow, 本文实验中采用的是 Caffe^[9], 作为清晰而高效的深度学习框架, 其核心语言采用 C++, 支持命令行、Python 和 MATLAB 接口, 且可以在 GPU 上运行, 直接集成了卷积神经网络神经层, 并且提供了如 LeNet, AlexNet 以及 GoogLeNet 等示例模型.

(1) 实验平台

本文的实验均在 ubuntu16.04 系统下进行, 采用的是 Caffe 学习框架, 其中使用编程语言 python, 且在图像预处理过程中采用 MATLAB.

本次实验采用数据集为 Fifteen Scene Categories 进行前期改进的测试, 采用数据集为 eight sports event categories dataset 进行后期性能的评估, 前者数据集包括 15 类场景图像, 每类含有 216~400 张图像, 本实验中每类随机选取 100 张图像作为训练集, 其余剩下的作为测试集, 如图 8 所示部分图像. 后者包含 8 种体育场景图像, 分别为划船, 羽毛球, 马球, 地滚球, 滑雪板, 槌球, 帆船以及攀岩. 根据人体判断, 图像被分为简单和中等, 还为每个图像提供前景对象距离的信息.

(2) 参数设置

使用 caffe 进行深度学习时, 为了使得更快收敛并获得比较高的识别精度, 调参是必不可少的工作, 表 2 为本次实验时根据以往实验的经验进行的参数设置.

表 2 Caffe 训练参数设置

Parameter	Value
Batch_size	20/10
Momentum	0.9
Iteration	6000
Weight decay	0.0005
Initial learning rate	0.001

(3) 采取微调 (Fine-tuning) 的方式训练模型

在本次研究中, 采用的是一种微调预训练卷积神经网络 (AlexNet) 模型的方法, 这样可以更好的解决由于缺乏标记图像而导致的过拟合问题. 因为前几层提取的底层特征可以更好的适应不同的问题, 故在微调的过程中前几层应该被保持不动, 最后的完全连接层是唯一一个可以用于调整权重的特征图层, 应修改为合适的数据集标签. 另外, 在整个训练过程中使用多批

次的图像进行学习, 批量的大小由于 CPU 内存决定, 图 9 证实了微调预先训练好的模型是有效的实验方法.

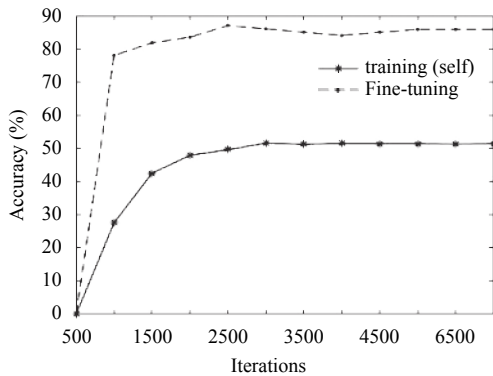


图 12 微调 AlexNet 模型效果图

3.3 实验与分析

综合上述四种算法的改进, 可知算法 4 和算法 3 表现得较为明显, 提升幅度达到 6.2% 和 5.7%, 最差的算法 4 提升幅度为 4.2%, 考虑到模型的泛化能力, 因此选择算法 4 和算法 3 作为最终的改进算法, 为了更好的验证改进算法的有效性, 将数据集 eight sports event categories 输入到训练好的改进模型中进行测试, 并取最好的实验结果进行比较和作合理的评价.

表 3 和表 4 为各种方法分别在两种实验数据集上的识别精度结果, 文献[10,11]采用的是传统的手工特征提取的方法, 实验表明使用 CNN 进行自动提取并进行改进远远高于普通手工提取的方法. 在 Scene15 数据集上, 改进的算法最终的结果比传统的手工算法高 8.1%~13.3%, 比改进前 AlexNet 高 6.2%; 同样的在 eight sports 数据集上, 改进的算法比传统算法高 7.2%~15.1%, 比改进前 AlexNet 高 2.9%. 此外, 由实验结果可以得知算法 4 的泛化能力略优于算法 3.

表 3 不同方法对 Scene15 数据集的平均识别精度

Methods	Accuracy(%)
VLAD + Aug. ^[10]	78.7
SIFT + SPM ^[2]	81.4
SIFT + ORSP ^[11]	83.9
AlexNet+Imagenet ^[12]	84.2
Convnets + Linear-SVM ^[12]	84.2
Convnets+RSP ^[13]	89.4
Ours(Algorithm3/Algorithm4)	91.5/92.0

图 13 为改进模型在 eight sports 数据集上的混淆矩阵, 分别表示每个运动场景类别的准确度. 图 13 表

明了 8 类运动场景中, 其中“bocce”、“sailing”场景均可以达到 100% 的识别率, 而“badminton”容易被“polo”和“rowing”混淆, 识别率相对表现得有点逊色, 其他运动场景类均可以达到 90% 以上的识别率.

表 4 不同方法对 eight sports 数据集的平均识别精度

Methods	Accuracy(%)
BOWL+SVM ^[14]	87.7
LLC+SPM ^[15]	83.0
ScSPM ^[16]	79.0
Wu and Rehg ^[13]	84.3
Convnets+ shared FC	91.8
Ours(算法 3/算法 4)	93.2/94.5

4 结论与展望

为了获得高质量的网络模型, 本文首先从常规的改进方法出发 (即增加模型的深度或者其宽度), 并在场景识别任务上进行了测试; 接着提出多尺度化特征提取和多层特征融合的改进方法, 分别测试其识别效果; 最终通过实验结果的比较选择最适合场景识别任务的改进模型, 并选择常用的数据集进行了有效性的验证. 实验表明, 本文的改进方法优于一般常见的识别算法.

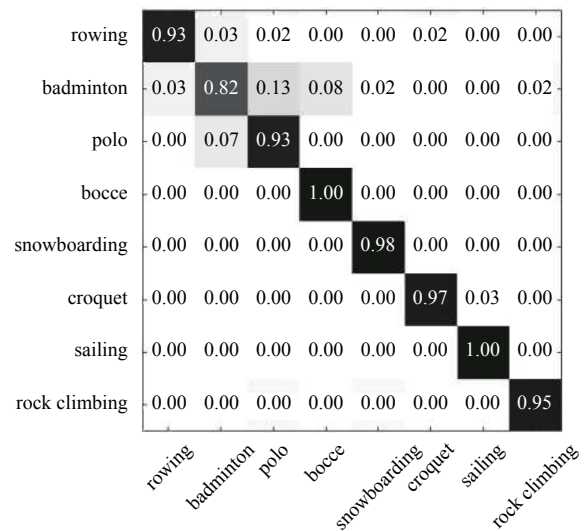


图 13 改进模型在 eight sports 数据集上的混淆矩阵

基于本文的研究, 后续的工作将从两方面开展: 首先选择其他比较深的网络模型进行改进, 比如 VGG 和 ResNet 网络, 探索出更加优化的网络模型; 其次保证高精度识别的同时, 提高其识别的效率, 此外结

合迁移学习的思想对模型的泛化能力进行优化。

参考文献

- 1 Koskela M, Laaksonen J. Convolutional network features for scene recognition. Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, FL, USA. 2014. 1169–1172.
- 2 Li T, Mei T, Kweon IS, *et al.* Contextual bag-of-words for visual categorization. IEEE Transactions on Circuits and Systems for Video Technology, 2011, 21(4): 381–392. [doi: [10.1109/TCSVT.2010.2041828](https://doi.org/10.1109/TCSVT.2010.2041828)]
- 3 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA. 2006. 2169–2178.
- 4 Menze BH, Van Leemput K, Lashkari D, *et al.* A generative probabilistic model and discriminative extensions for brain lesion segmentation—with application to tumor and stroke. IEEE Transactions on Medical Imaging, 2016, 35(4): 933–946. [doi: [10.1109/TMI.2015.2502596](https://doi.org/10.1109/TMI.2015.2502596)]
- 5 Yu J, Rui Y, Tao DC. Click prediction for web image reranking using multimodal sparse coding. IEEE Transactions on Image Processing, 2014, 23(5): 2019–2032. [doi: [10.1109/TIP.2014.2311377](https://doi.org/10.1109/TIP.2014.2311377)]
- 6 He KM, Sun J. Convolutional neural networks at constrained time cost. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 5353–5360.
- 7 Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland. 2014. 818–833.
- 8 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- 9 Jia YQ, Shelhamer E, Donahue J, *et al.* Caffe: Convolutional architecture for fast feature embedding. arXiv: 1408.5093, 2014.
- 10 Vedaldi A, Fulkerson B. VLFeat: An open and portable library of computer vision algorithms. Proceedings of the 18th ACM International Conference on Multimedia. Firenze, Italy. 2010. 1469–1472.
- 11 Jiang YN, Yuan JS, Yu G. Randomized spatial partition for scene recognition. Proceedings of the 12th European Conference on Computer Vision. Florence, Italy. 2012. 730–743.
- 12 Zhou BL, Lapedriza A, Xiao JX, *et al.* Learning deep features for scene recognition using places database. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada. 2014. 487–495.
- 13 Wu JX, Rehg JM. Beyond the Euclidean distance: Creating effective visual codebooks using the histogram intersection kernel. Proceedings of 2009 IEEE 12th International Conference on Computer Vision. Kyoto, Japan. 2009. 630–637.
- 14 Banerji S, Sinha A, Liu CJ. A new bag of words LBP (BoWL) descriptor for scene image classification. Proceedings of the 15th International Conference on Computer Analysis of Images and Patterns. York, UK. 2013. 490–497.
- 15 Wang JJ, Yang JC, Yu K, *et al.* Locality-constrained linear coding for image classification. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA. 2010. 3360–3367.
- 16 Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the 5th Annual Workshop on Computational Learning Theory. Pittsburgh, PA, USA. 1992. 144–152.