

Deep Convolutional Neural Network Model Based on Multi-Region Feature

Ya-mei Wang*, Zhen-you Wang#

Department of Applied Mathematics, Guangdong University of Technology, Guangzhou Guangdong
Email: 913329040@qq.com, #zywang@gdut.edu.cn

Received: Oct. 23rd, 2019; accepted: Nov. 11th, 2019; published: Nov. 18th, 2019

Abstract

Aiming at the problems of complex network structure and large computational complexity of existing deep convolutional neural network models, it cannot be widely used in practice. This paper proposes a deep convolutional neural network model based on multi-region features. Firstly, the model divides the image into multiple regions and then uses the standard convolution operation to get the image semantic context information. Then, the multi-region input is used to learn the context interaction feature. By cascading and inputting the spatial information of the global region and the multiple sub-regions into the convolution layer, the context feature information of the image is extracted in an information supplement manner. Finally, the images are classified by the Softmax function. The experimental results show that the model has simple structure and few parameters, and the modeling with multi-region feature and context information fusion has better robustness and higher classification accuracy than that with single-region feature.

Keywords

Multiregional Feature, Convolutional Neural Network, Image Classification

基于多区域特征的深度卷积神经网络模型

王雅湄*, 王振友#

广东工业大学应用数学学院, 广东 广州
Email: 913329040@qq.com, #zywang@gdut.edu.cn

收稿日期: 2019年10月23日; 录用日期: 2019年11月11日; 发布日期: 2019年11月18日

*第一作者。

#通讯作者。

摘要

针对现有的深度卷积神经网络模型存在的网络结构复杂、计算量较大等问题, 从而无法在实际中有较广泛的应用, 该文提出一种基于多区域特征深度卷积神经网络模型。模型首先对图像进行多区域划分, 然后用标准卷积操作得到图像语义上下文信息, 接着利用多区域的输入来学习上下文交互特征, 通过把全局区域和多个分区域的空间信息级联再输入卷积层, 以一种信息补充的方式提取图像的上下文特征信息, 最后通过Softmax函数对图像进行分类。实验结果表明, 该模型结构简单, 参数量较少, 且多区域特征融合上下文信息建模比单区域特征建模具有更好的鲁棒性和更高的分类精度。

关键词

多区域特征, 卷积神经网络, 图像分类

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近十年来, 深度学习在计算机视觉领域取得了巨大的成功。目前计算机视觉技术在图像分类[1]-[11]、人脸识别[12]、物体检测[13]、目标跟踪[14]等领域取得了突飞猛进的发展。传统基于特征的图像分类方法存在着泛化能力弱、移植性差、准确度较低的缺点。卷积神经网络作为深度学习的三大重要模型之一, 不仅拥有深度学习技术自动提取特征的优点, 而且通过权值共享大大减少了所需要训练的参数, 还具备了一定的平移不变性, 使得卷积神经网络能快速准确地处理图像, 因此在图像分类领域取得显著的效果。

2012年10月, Hinton教授以及他的学生采用卷积神经网络模型 AlexNet [1], 在著名 ImageNet [15] 问题上取得了当时世界上最好的成果, 使得对于图像分类的研究工作前进了一大步。AlexNet 卷积神经网络模型大获成功后, 掀起了深度卷积神经网络的研究热潮。在这之后, 新的卷积神经网络模型不断被提出, 比如牛津大学的 VGGNet [2]、谷歌的 GoogleNet [3]和微软提出的 ResNet [4]等不断刷新 AlexNet 在 ImageNet 竞赛上的记录。在 2017 年, ResNet [4]作为近年来 CNN 结构发展中最为关键的一个结构, 一些基于 ResNet 的变体不断被提出, 比如 ResNeXt [5]、DenseNet [6]、NPN [7]、SENet [8]等。在 2018 年 5 月 11 日微软亚洲研究所进行的 CVPR2018 中国论文宣讲研讨会中, 张婷团队延续 IGCV 1 [9]的工作, 提出了 IGCV 2 [10]和 IGCV 3 [11]模型。

随着卷积神经网络的层数不断增加, 模型拟合能力逐渐增强, 识别效果也越好。但模型越深, 训练越不易收敛, 并且训练效率低下; 反之, 当网络模型层数比较少的时候, 模型拟合能力降低, 识别率也会降低。另外, 对于复杂的模型而言, 除了训练效率极低外, 模型理论分析较为困难, 训练时往往依靠丰富的调参技巧和经验进行多次实验得出结果。

在用卷积神经网络处理图像分类的任务中, 常规操作都是先对图像进行预处理后整幅图像输入到卷积神经网络模型再输出分类。然而, 由于不同图像类别目标可能存在区域上的差异, 不同的类别可能具有相似性极高的特征信息, 因此仅通过全局特征去进行图像分类的准确率不高。本文设计一个基于多区域特征深度卷积神经网络模型, 该模型设计了不同的输入模式, 以保证特征信息的完整性。输入模

式,即不同的区域或全局区域(例如,中心区域、原始区域和左上右上左下右下四个方向的子区域),支持每个像素的联合表示,以确保得到更丰富和更有鲁棒性的特征信息。本文模型不仅网络层数较少、参数量较少,而且图像分类的准确率也较高。

2. 多区域深度卷积神经网络模型

2.1. MR-CNN 模型介绍

深度卷积神经网络(DCNN)一般由卷积层、池化层、全连接层三种神经网络层以及一个输出层组成。图像一般先经过卷积层进行卷积,得到若干特征图后通过池化层对特征图进行压缩,最后通过一个全连接层输出以识别图像特征,进而用分类器对图像进行分类。传统的卷积神经网络在卷积和池化各特征图之间虽然采用并行计算,但是整个网络是在单个通路内完成的,网络只在深度上拓展,容易造成梯度消失和过拟合现象。本文通过把原始图像划分成多个区域,设计简单的多通路卷积层,进而把多区域图像分别卷积后进行级联卷积再输出分类。

(1) 区域截取的指导原则

原始图像平均分成四块,分别标记为原始图像的左上方分块(LT)、右上方分块(RT)、左下方分块(LB)和右下方分块(RB)。把原始图像平均分成四块,可以专注于图像的局部区域,进而更好地提取原始图像的局部特征,提高对图像局部细节信息的识别。

(2) 中心区域截取标准

由于一般图像的大部分信息主要集中在图像的中心,因此,除了把原始图像平均裁剪成四个部分去提取更多图像的局部细节信息外,再提取一个图像的中心区域,使得图像的局部信息结合得更加完整。同时中心区域与四个方向的子区域具有同样大小。

(3) 模型设计

先把原始图像输入 MR-CNN 模块一与不同尺寸的滤波器进行卷积操作。各分块图像分别输入 MR-CNN 模块二进行卷积操作,其中滤波器的尺寸先采用 3×3 是因为分块图像的尺寸较小,使用尺寸较小的滤波器可以更好提取各分块图像的上下文特征信息,再使用 1×1 的滤波器可以保持特征图尺度不变(即不损失分辨率)的前提下大幅增加非线性特性,使网络能提取到更有鲁棒性的特征。然后对原始图像和各分块卷积后的特征图级联后再输入 MR-CNN 模块三,以一种信息补充的方式学习到图像的上下文交互特征,其中模块三也是采用两层 3×3 滤波器加一层 1×1 一层滤波器。这样设计是因为级联后的特征图较小,而且通道数较多,需要使用较小尺寸的滤波器卷积提取特征,用 1×1 的滤波器除了可以增加特征的非线性还可以起到降维的作用。最后输入全局平均池化层再输出分类。多区域深度卷积神经网络模型(MR-CNN)如图 1 所示,图 2 为 MR-CNN 三个模块的具体结构,其中 $[5 \times 5]/1:96$ 表示使用 96 个步长为 1 的 5×5 的卷积核进行卷积。

MR-CNN 通过把原始图像和各个分区域图像分别卷积可以获得更加精细的全局特征和局部特征,卷积后级联的步骤把图像的全局特征和局部特征进行融合,使图像的上下文信息更好地结合,图像特征更有表达力。特征融合后的再次卷积用于提取更深层的特征进行分类。卷积的具体步骤与传统 CNN 相同,原始输入图像与滤波器进行卷积后通过一个激活函数,得到输出特征图,公式为:

$$x_j^{l+1} = f \left(\sum_{i \in M_j} x_i^{(l)} * k_{ij}^{(l+1)} + b_j^{(l+1)} \right) \quad (1)$$

其中:上标表示所在的层数;*为卷积操作; $x_j^{(l+1)}$ 表示卷积后第 j 个神经元的输出; $x_i^{(l)}$ 表示第 l 层第 i 个神经元,即输入数据; $k_{ij}^{(l+1)}$ 表示滤波器; $b_j^{(l+1)}$ 表示偏置; M_j 表示选择的输入特征图的集合。

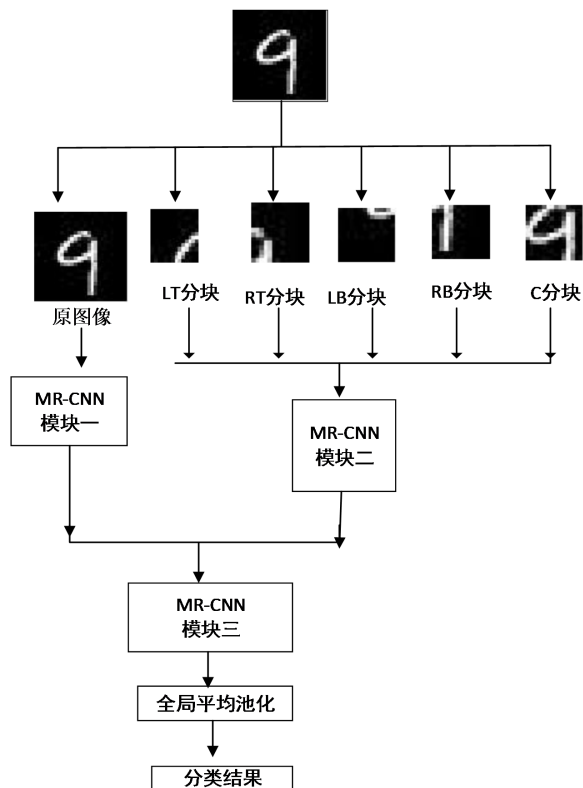


Figure 1. MR-CNN model
图 1. MR-CNN 模型



Figure 2. The specific structure of the three modules of MR-CNN
图 2. MR-CNN 三个模块的具体结构

2.2. 批量归一化

在深度卷积神经网络的训练中, 由于网络中前一层的参数调整会导致后一层输入值的分布发生变化, 通常需要仔细选取初始参数并采取较小的学习率。但是, 这不但导致模型训练的效率低下, 而且使得含

饱和非线性函数模型的训练极为困难。因此, MR-CNN 采用了批量归一化[16]加快网络的训练和提高分类效果。批量归一化允许使用更高的学习率并且减少对参数初始化的依赖, 它在一定情况下也可以起到正则化的作用, 故 MR-CNN 去掉了 Dropout 层也不会造成过拟合的现象, 同时使网络的泛化能力更强。MR-CNN 在每个卷积层之后都加上一个批量归一化层, 减少了初始化的强依赖性的同时, 改善了整个网络的梯度流, 使网络具有更好的鲁棒性。批量归一化公式如下:

$$\hat{x}_i = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \xi}} + \beta \quad (2)$$

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (3)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (4)$$

其中: $B = \{x_1, x_2, \dots, x_m\}$ 表示 m 个小批量输入, μ_B 表示小批量均值, σ_B^2 表示小批量方差, γ, β 表示需要学习的超参数, ξ 表示小批量方差的一个常量。

2.3. Swish 激活函数

MR-CNN 与现有的 CNN 的另一个不同之处是激活函数把 Relu 激活函数替换成了 Swish 激活函数[17]。激活函数通过引入非线性激活因素, 对于提高深度卷积神经网络模型的表达能力具有十分重要的作用。Swish 函数是 2017 年提出的新型激活函数, 和 Relu 函数一样, Swish 函数无上界有下界。与 Relu 函数不同的是 Swish 函数是平滑且非单调的函数。大量实验表明, 在应用于各种图像分类和机器翻译等领域的深度神经网络上, Swish 函数可以获得与 Relu 函数相匹配的性能甚至优于 Relu 函数。因此, MR-CNN 选择 Swish 函数作为网络模型的激活函数以提高图像的分类精确率。它的数学表达式为:

$$f(x) = x \cdot \sigma(x) \quad (5)$$

$$\sigma(x) = 1 / (1 + \exp(-x)) \quad (6)$$

其中 $\sigma(x)$ 为 Sigmoid 函数。

2.4. 全局平均池化

MR-CNN 通过去掉传统的池化层, 避免因模型中的池化操作压缩了特征图而使得图像部分有用的信息被丢弃, 从而保证图像特征信息的完整性。全连接层的参数量巨大, 容易造成网络模型的过拟合, 从而阻碍了整个网络模型的泛化能力。MR-CNN 把全连接层用全局平均池化层替换掉, 在保证一定准确率的前提下, 减少网络的参数量, 进而使网络的复杂度降低, 加快网络模型的训练速度, 节省计算资源的同时减轻网络模型的过拟合现象, 提高网络模型的泛化能力。如图 3 左为一般 CNN 用到的全连接层, 图 3 右为 NIN (Network in Network [18])提出的全局平均池化层。全局平均池化对整个网络在结构上做正则化防止过拟合, 其直接剔除全连接中的黑箱特征, 直接赋予每个通道实际的类别意义(即直接分别对每个通道的整个特征图进行平均池化然后输入 Softmax 进行分类)。

3. MR-CNN 算法

图 4 为本文提出的 MR-CNN 算法步骤。

算法步骤如下:

(1) 将训练集中的原图像(G)平均裁剪成四个方向的子区域, 同时也在原图像上取与四个子区域图像尺寸相同的中心区域, 分别命名为左上方区域(LT)、右上方区域(RT)、左下方区域(LB)、右下方区域(RB)

和中心区域(C)。

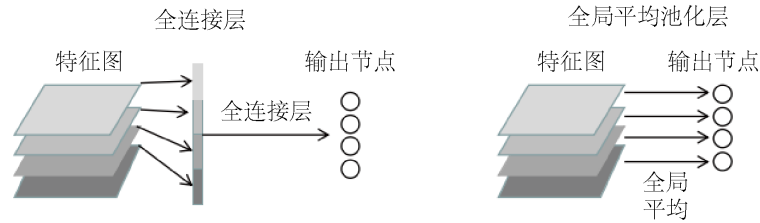


Figure 3. Fully connected layer and global average pooling layer
图 3. 全连接层和全局平均池化层

(2) 对原图像(G)按图 2 的 MR-CNN 模块一用公式(1)计算卷积, 用公式(2)、(3)、(4)计算批量归一化后得到特征图 R_G 。

(3) 对五个小分块按图 2 的 MR-CNN 模块二分别用公式(1)计算卷积, 用公式(2)、(3)、(4)计算批量归一化后得到特征图 R_{LT} 、 R_{RT} 、 R_{LB} 、 R_{RB} 和 R_C 。

(4) 将特征图 R_G 、 R_{LT} 、 R_{RT} 、 R_{LB} 、 R_{RB} 和 R_C 进行级联(concat)。

(5) 把级联后输出的特征图作为下一个卷积层的输入, 按图 2 的 MR-CNN 模块三用公式(1)计算卷积, 用公式(2)、(3)、(4)计算批量归一化后得到特征图。

(6) 将上一步得到的特征图作为全局平均池化层的输入, 利用 Softmax 分类器得到图像的分类结果。

(7) 用交叉熵计算分类结果和标签的差异, 通过卷积神经网络模型专用的反向传播算法调节并更新参数 k_{ij} , 直到损失函数收敛于一个较小的值, 训练完毕。

(8) 输入测试集, 利用训练得到的网络对测试图像进行分类。

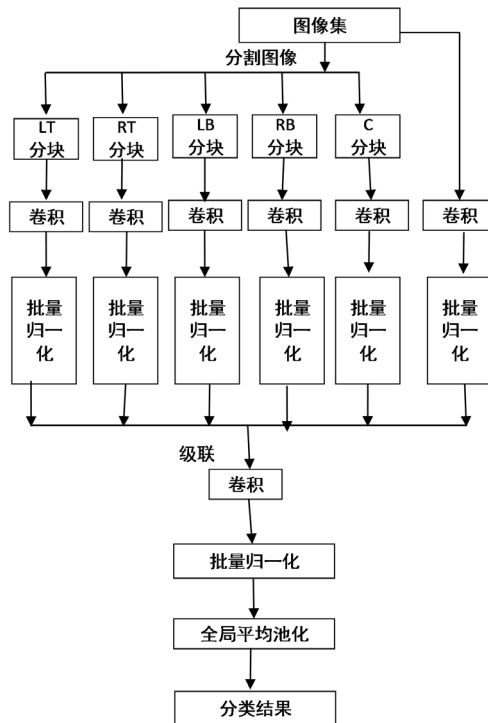


Figure 4. MR-CNN algorithm flow
图 4. MR-CNN 算法步骤

4. 实验与分析

为验证本文算法的分类性能,我们在 MNIST 数据集和 Cifar-10 数据集上进行实验,并与不同的分类算法进行比较,以验证本文的网络模型结构简单,参数量较少且基于多区域上下文特征信息建模比单区域特征信息建模可以得到更高的分类准确率。实验硬件平台 Intel(R)Core(TM)i7-8700 CPU@3.20 GHz 3.19 GHz, 8 G 内存。实验的软件平台: Windows10 64 位操作系统 Python3.5, Tensorflow1.12。

4.1. 实验数据集介绍

MNIST 数据集是一个手写识别数据集,图片大小为 28×28 ,其中包含 60000 个训练样本以及 10000 个测试样本。像素值为 0 表示白色,像素值为 255 表示黑色。

Cifar-10 数据集包含 50000 张训练图片以及 10000 张测试图片,数据集图片包含 airplane、automobile 等 10 个类别的物体。其中,每张图片都是 RGB 三通道的彩色图像,且均为 32×32 大小。

4.2. 实验结果及分析

4.2.1. 本文模型算法与其他模型算法对比

为验证本文网络模型的结构简单且能得到较高的分类准确率,我们在 MNIST 数据集和 Cifar-10 数据集上与不同的分类算法进行比较。

表 1 给出在 MNIST 数据集上,本文方法与其他不同的分类算法进行图像分类的错误率的对比。本文算法的图像分类识别准确率在 MNIST 数据集上优于其他算法,相比 PCANet [19]、Stochastic Pooling ConvNet [20]、Conv + Maxout + Dropout [21]分别提高了 0.27%、0.12%、0.10%,达到了 99.65%。其中,实验训练时,把测试集设置成 10 个 Batch-Size 输入网络进行测试验证,每个实验结果是在做了 20 次实验后取平均值所得。

Table 1. Comparison of classification error rates between model algorithm and other model algorithms in MNIST datasets

表 1. 本文模型算法与其他模型算法在 MNIST 数据集的分类错误率的比较

Methods	Error Rates (%)
PCANet [19]	0.62
Stochastic Pooling ConvNet [20]	0.47
Conv + Maxout + Dropout [21]	0.45
MR-CNN	0.35

表 2 给出在 Cifar-10 数据集上,本文方法与不同的分类算法进行图像识别的分类精确度的对比。本文算法的图像识别准确率在 Cifar-10 数据集上优于其他算法,比 PCANet [19]、Stochastic Pooling ConvNet [20]、Conv + Maxout + Dropout [21]分别提高了 11.48%、3.75%、0.30%,达到了 88.62%。同样地,实验训练时,把测试集设置成 10 个 Batch-Size 入网络进行测试验证,每个实验结果是在做了 20 次实验后取平均值所得。

4.2.2. 本文模型与其他模型复杂度对比

为了分析模型的复杂度,同样对四种经典的神经网络模型复杂度进行比较分析。五种模型的网络层数和网络整体参数量如表 3。

Table 2. Comparison of classification accuracy between model algorithm and other Model Algorithms in Cifar-10 data set**表 2.** 本文模型算法与其他模型算法在 Cifar-10 数据集的分类精度的比较

Methods	Accuracy (%)
PCANet [19]	77.14
Stochastic Pooling ConvNet [20]	84.87
Conv + Maxout + Dropout [21]	88.32
MR-CNN	88.62

Table 3. Comparison of five model network layers and network parameters**表 3.** 五种模型网络层数及网络整体参数对比

Model	Number of Layer	Parameters (M)
AlexNet [1]	8	60
VggNet16 [2]	16	138
GoogleNet [3]	22	5
ResNet50 [4]	50	5.5
MR-CNN	6	3

从表 3 可以看出, 本文所提出的模型网络层数是最少的, 只有 6 层, 对比其他经典模型的网络结构简单得多。虽然 AlexNet 的网络层数跟本文模型的网络相差不多, 但它的网络里存在全连接层, 模型整体的网络参数量较多, 使得网络复杂度增加, 加大网络的计算量, 网络的工作效率下降。本文所提出的模型网络参数量也是最少的, 对比 GoogleNet, 网络参数量虽然相差不多, 但它的网络层数较多, 结构也比较复杂。本文所提模型的网络参数主要消耗在多通道的级联部分, 但网络模型整体结构简单, 层数最少, 整体网络复杂度较低。所以, 本文所提模型权衡了网络复杂度和算法准确率之间的关系, 具有较好的鲁棒性的同时有较高的准确率。

4.2.3. 多个区域级联对比

为了分析级联区域个数对图像分类效果的影响, 在 Cifar-10 数据集上对各个分区域级联个数各种组合进行比较分析。

从表 4 可以看出, 级联任意两个分区域的分类精度相差不多, 本文所提的 MR-CNN 模型的分类精度明显优于它们。这表明, 级联两个分区域的特征信息不够完整, 本文的 MR-CNN 模型更加全面地把局部特征和全局特征结合, 得到更高的分类精度。

Table 4. Comparison of classification accuracy between proposed model and two sub-regional cascaded models**表 4.** 所提模型与两个分区域级联模型的分类精度对比

Region	Accuracy (%)
$R_G + R_C$	85.85
$R_G + R_{LT}$	86.13
$R_G + R_{RT}$	86.16
$R_G + R_{LB}$	85.99
$R_G + R_{RB}$	86.17
MR-CNN	88.62

从表 5 可以看出, 级联任意三个分区域的分类精度相差不大, 本文所提的 MR-CNN 模型的分类精度明显优于它们。这表明, 级联三个分区域的特征信息也不够完整, 但对比表 4 级联任意两个分区域的分类精度要高一些。因此, 级联多一个分区域有利于获得更多的图像特征信息, 提高分类精度, 其中本文的 MR-CNN 模型有最高的分类精度。

Table 5. Comparison of classification accuracy between proposed model and three sub-regional cascaded models

表 5. 所提模型与三个分区域级联模型分类精度对比

Region	Accuracy (%)
$R_G + R_{LT} + R_{RT}$	86.53
$R_G + R_{LT} + R_{LB}$	86.50
$R_G + R_{LT} + R_{RB}$	86.33
$R_G + R_{LT} + R_C$	86.35
$R_G + R_{RT} + R_{LB}$	86.42
$R_G + R_{RT} + R_{RB}$	86.33
$R_G + R_{RT} + R_C$	86.43
$R_G + R_{LB} + R_{RB}$	86.35
$R_G + R_{LB} + R_C$	86.28
$R_G + R_{RB} + R_C$	86.30
MR-CNN	88.62

从表 6 可以看出, 级联任意四个分区域的分类精度同样相差不大, 同样与本文所提的 MR-CNN 模型的分类精度有一些差距。这表明, 级联四个分区域的特征信息还不够完整, 但对比表 4 和表 5, 级联任意两个分区域或者级联任意三个分区域的分类精度都要高一些。因此, 更加能证明级联多一个分区域有利于获得更多的图像特征信息, 提高图像的分类精度。MR-CNN 把五个分区域全部级联, 把图像的各个部分局部特征和全局特征的有用信息提取和融合, 进而获得最高的分类精度。

Table 6. Comparison of classification accuracy between proposed model and four sub-regional cascade models

表 6. 所提模型与四个分区域级联模型分类准确率对比

Region	Accuracy (%)
$R_G + R_{LT} + R_{RT} + R_{LB}$	87.37
$R_G + R_{LT} + R_{RT} + R_{RB}$	87.34
$R_G + R_{LT} + R_{RT} + R_C$	86.82
$R_G + R_{RT} + R_{LB} + R_{RB}$	87.20
$R_G + R_{RT} + R_{LB} + R_C$	86.80
$R_G + R_{LB} + R_{RB} + R_C$	86.95
MR-CNN	88.62

4.2.4. 用于本文模型算法内部参数对比

为了分析本文模型算法内部参数设置对图像分类准确率的影响,我们在 MNIST 数据集和 Cifar-10 数据集上分别做了有无批量归一化、不同激活函数、全连接和全局平均池化对图像分类效果影响的实验。

表 7 和表 8 给出了在 MNIST 数据集上和 Cifar-10 数据集上级联两个分区域和本文模型在有无批量归一化情况下的分类性能。通过表 7 和表 8 我们可以看出无论在本文模型还是级联两个分区域的情况下,添加了批量归一化都明显优于没有添加批量归一化的分类效果,从而可以证明在本文所提模型中,批量归一化对于图像分类性能的确有所提高。在实验中,我们也可以发现加了批量归一化十分明显地加快了网络的训练速度,网络收敛速度的加快为我们训练网络节省了很多时间成本,提高了我们的网络模型的工作效率。通过本文设计的网络模型,把全局区域特征与各个分区域特征级联的方法可以把图像的分类精度达到最好的效果。

Table 7. Classification error rates (%) with or without batch normalization on MNIST data

表 7. 在 MNIST 数据集上, 有无批量归一化的分类错误率

Region	With BN	Without BN
R_G	0.42	0.65
$R_G + R_C$	0.39	0.62
$R_G + R_{LT}$	0.43	0.68
$R_G + R_{RT}$	0.44	0.69
$R_G + R_{LB}$	0.43	0.67
$R_G + R_{RB}$	0.44	0.66
MR-CNN	0.35	0.58

Table 8. Classification accuracy (%) with or without batch normalization on Cifar-10 data

表 8. 在 Cifar-10 数据集上, 有无批量归一化的分类精确率

Region	With BN	Without BN
R_G	84.94	82.64
$R_G + R_C$	85.85	82.73
$R_G + R_{LT}$	86.13	83.01
$R_G + R_{RT}$	86.16	83.15
$R_G + R_{LB}$	85.99	82.88
$R_G + R_{RB}$	86.17	83.07
MR-CNN	88.62	84.24

表 9 和表 10 给出了在 MNIST 数据集上和 Cifar-10 数据集上, 本文所提出的模型和级联两个分区域模型采用不同激活函数的分类性能。将本文所提模型用到的 Swish 激活函数与现在最常用的 Relu 函数作了对比。通过实验结果表明, 本文网络模型用到的 Swish 函数无论在全局区域还是在级联两个分区域中都比 Relu 函数的分类效果好一些。其中, 在本文所提的模型中, Swish 函数比 Relu 函数的图像分类效果提高得更显著。

Table 9. Classification error rates (%) of different activation functions on MNIST data
表 9. 在 MNIST 数据上, 不同激活函数的分类错误率

Region	Relu	Swish
R_G	0.51	0.42
$R_G + R_C$	0.47	0.39
$R_G + R_{LT}$	0.49	0.43
$R_G + R_{RT}$	0.48	0.44
$R_G + R_{LB}$	0.50	0.43
$R_G + R_{RB}$	0.49	0.44
MR-CNN	0.45	0.35

Table 10. Classification accuracy (%) of different activation functions on Cifar-10 data
表 10. 在 Cifar-10 数据集上, 不同激活函数的分类精度

Region	Relu	Swish
R_G	85.44	84.94
$R_G + R_C$	85.48	85.85
$R_G + R_{LT}$	85.46	86.13
$R_G + R_{RT}$	85.82	86.16
$R_G + R_{LB}$	86.03	85.99
$R_G + R_{RB}$	86.05	86.17
MR-CNN	87.18	88.62

表 11 和表 12 给出了在 MNIST 和 Cifar-10 数据集上, 本文所提出的模型分别在使用全连接层和全局平均池化层的分类性能。实验结果表明, 本文通过把全局平均池化层替换掉了传统的全连接层, 网络模型除了大大地减少了参数量, 使网络模型更加简化, 加快网络模型的训练速度外, 无论在全局区域还是在级联两个分区域中准确率都有很大的提升。其中, 本文所提的 MR-CNN 模型分类效果是最好的。

Table 11. Classification error rates (%) of fully connected and global average pooling on MNIST data
表 11. 在 MNIST 数据集上, 全连接和全局平均池化的分类错误率

Region	FC	GAP
R_G	0.78	0.42
$R_G + R_C$	0.76	0.39
$R_G + R_{LT}$	0.83	0.43
$R_G + R_{RT}$	0.86	0.44
$R_G + R_{LB}$	0.85	0.43
$R_G + R_{RB}$	0.85	0.44
MR-CNN	0.72	0.35

Table 12. Classification accuracy (%) of fully connected and global average pooling on Cifar-10 data
表 12. 在 Cifar-10 数据集上, 全连接和全局平均池化的分类精度

Region	FC	GAP
R_G	84.72	84.94
$R_G + R_C$	85.09	85.85
$R_G + R_{LT}$	85.37	86.13
$R_G + R_{RT}$	85.40	86.16
$R_G + R_{LB}$	85.73	85.99
$R_G + R_{RB}$	85.77	86.17
MR-CNN	87.04	88.62

表 7~12 这六个实验结果表明, 本文的 MR-CNN 模型在分类精度上都明显优于仅有全局区域和级联两个分区域的分类精度。很明显无论是在全局区域还是在级联两个分区域中, 批量归一化、Swish 函数和全局平均池化层的分类操作都对分类任务起着积极的作用。

从每个实验数据的结果来看, 基于全局区域与多个局部区域级联(MR-CNN)的分类性能要优于单个区域、两个区域级联、三个区域级联、四个区域级联。本文提出的 MR-CNN 算法在考虑多个分布区域的情况下, 具有更强的鲁棒性特征表示, 由于区域分块卷积和多区域级联卷积的结合, MR-CNN 达到了最佳的分类精度, 并且所提出的 MR-CNN 模型明显优于其它分类器。

5. 结语

本文提出一种新的基于多区域特征的深度卷积图像分类模型, 提出的 MR-CNN 模型通过精心设计的网络结构提取图像的语义上下文特征信息, 利用级联技术把全局特征和局部特征融合再设计简单的卷积层提取更具有表达力的上下文交互特征信息进行分类。MR-CNN 的优点来自于利用不同区域的输入和深层网络结构进行丰富的空间特征的探索。其中, MR-CNN 把图像的全局区域与多个分区域进行组合, 充分地提取了图像的全局与局部上下文信息, 同时把批量归一化、Swish 函数和全局平均池化三者结合, 共同加快网络模型的训练速度和提高网络模型分类精度, 从而达到更好的图像分类效果。实验结果表明, 本文提出的 MR-CNN 不仅网络模型的结构简单, 层数较少, 参数量较少, 而且多区域特征融合上下文信息建模比单区域特征建模具有更好的鲁棒性和更高的分类精度。

参考文献

- [1] Krizhevsky, A., Sutskever, I. and Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1097-1105.
- [2] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. *International Conference on Learning Representations (ICLR)*.
- [3] Szegedy, C., Liu, W., Jia, Y., et al. (2015) Going Deeper with Convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [4] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Deep Residual Learning for Image Recognition. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [5] Xie, S., Ross, G., Dollar, P., Tu, Z. and He, K. (2017) Aggregated Residual Transformations for Deep Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 21-26 July 2017, 1492-1500. <https://doi.org/10.1109/CVPR.2017.634>
- [6] Huang, G., Liu, Z., Maaten, L.V.D. and Weinberger, K.Q. (2017) Densely Connected Convolutional Networks. *2017*

-
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 21-26 July 2017, 4700-4708. <https://doi.org/10.1109/CVPR.2017.243>
- [7] Chen, Y., Li, J., Xiao, H., *et al.* (2017) Dual Path Networks. *Neural Information Processing Systems (NIPS)*.
- [8] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 18-23 June 2018, 7132-714. <https://doi.org/10.1109/CVPR.2018.00745>
- [9] Zhang, T., Qi, G.J., Xiao, B. and Wang, J. (2017) Interleaved Group Convolutions. 2017 *IEEE International Conference on Computer Vision*, Venice, Italy, 22-29 October 2017, 4373-4382. <https://doi.org/10.1109/ICCV.2017.469>
- [10] Xie, G., Wang, J., Zhang, T., *et al.* (2018) IGCV\$2\$: Interleaved Structured Sparse Convolutional Neural Networks. 2018 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 18-23 June 2018, 8847-8856. <https://doi.org/10.1109/CVPR.2018.00922>
- [11] Sun, K. (2018) IGCV3: Interleaved Low-Rank Group Convolutions for Efficient Deep Neural Networks. *The British Machine Vision Conference (BMVC)*.
- [12] Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. (2014) Deep Face: Closing the Gap to Human-Level Performance in Face Verification. 2014 *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, 23-28 June 2014, 1701-1708. <https://doi.org/10.1109/CVPR.2014.220>
- [13] Girshick, R., Donahue, J., Darrelland, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [14] Wang, N. and Yeung, D.Y. (2013) Learning a Deep Compact Image Representation for Visual Tracking. *International Conference on Neural Information Processing Systems (NIPS)*.
- [15] Deng, J., Dong, W., Socher, R., *et al.* (2009) ImageNet: A Large-Scale Hierarchical Image Database. 2009 *IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 20-25 June 2009, 248-255. <https://doi.org/10.1109/CVPR.2009.5206848>
- [16] Ioffe, S. and Szegedy, C. (2015) Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *International Conference on Machine Learning*, 1-11.
- [17] Ramachandran, P., Zoph, B. and Le, Q.V. (2017) Searching for Activation Functions. *Computer Science*, 1-13.
- [18] Lin, M., Chen, Q. and Yan, S. (2013) Network in Network. *Computer Science*, 1-10.
- [19] Chan, T.H., Jia, K., Gao, S., *et al.* (2015) PCANet: A Simple Deep Learning Baseline for Image Classification? *IEEE Transactions on Image Processing*, **24**, 5017-5032. <https://doi.org/10.1109/TIP.2015.2475625>
- [20] Zeiler, M.D. and Fergus, R. (2013) Stochastic Pooling for Regularization of Deep Convolutional Neural Networks. *International Conference on Learning Representations (ICLR)*, 1-9.
- [21] Goodfellow, I.J., Warde-Farley, D., Mirza, M., Courville, A. and Bengio, Y. (2013) Maxout Networks. *Computer Science*, 1319-1327.