



改进MobileNet的图像分类方法研究

高淑萍, 赵清源, 齐小刚, 程孟菲

引用本文:

高淑萍, 赵清源, 齐小刚, 等. 改进MobileNet的图像分类方法研究[J]. 智能系统学报, 2021, 16(1): 11–20.

GAO Shuping, ZHAO Qingyuan, QI Xiaogang, et al. Research on the improved image classification method of MobileNet[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(1): 11–20.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202012034>

您可能感兴趣的其他文章

一种改进的深度学习的路面交通标识识别算法

An improved deep learning algorithm for road traffic identification

智能系统学报. 2020, 15(6): 1121–1130 <https://dx.doi.org/10.11992/tis.201811009>

融合整体与局部信息的武夷岩茶叶片分类方法

Classification of Wuyi rock tealeaves by integrating global and local information

智能系统学报. 2020, 15(5): 919–924 <https://dx.doi.org/10.11992/tis.202003018>

基于增强AlexNet的音乐流派识别研究

Music genre recognition research based on enhanced AlexNet

智能系统学报. 2020, 15(4): 750–757 <https://dx.doi.org/10.11992/tis.201909032>

基于小样本学习的LCD产品缺陷自动检测方法

An automatic small sample learning-based detection method for LCD product defects

智能系统学报. 2020, 15(3): 560–567 <https://dx.doi.org/10.11992/tis.201904020>

基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566–574 <https://dx.doi.org/10.11992/tis.201804056>

基于卷积特征和贝叶斯分类器的人脸识别

Face recognition based on convolution feature and Bayes classifier

智能系统学报. 2018, 13(5): 769–775 <https://dx.doi.org/10.11992/tis.201706052>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202012034

改进 MobileNet 的图像分类方法研究

高淑萍, 赵清源, 齐小刚, 程孟菲

(西安电子科技大学 数学与统计学院, 陕西 西安 710126)

摘 要: 针对神经网络结构的特征提取能力不足以及在包含复杂图像特征的数据集上分类准确率不高的问题, 本文提出了一种对 MobileNet 神经网络的改进策略 (L-MobileNet)。将原标准卷积形式替换为深度可分离卷积形式, 并将深度卷积层得到的特征图执行取反操作, 通过深度卷积融合层传递至下一层; 采用 Leaky ReLU 激活函数代替原 ReLU 激活函数来保留图像中更多的正负特征信息, 并加入类残差结构避免梯度弥散现象。与 6 种方法进行对比, 实验结果表明: L-MobileNet 在数据集 Cifar-10、Cifar-100(coarse)、Cifar-100(fine) 和 Dogs vs Cats 上平均准确率和最高准确率都取得了最佳结果。

关键词: 卷积神经网络; 图像分类; 特征提取; MobileNet; 深度可分离卷积; 激活函数; Leaky ReLU; 残差结构
中图分类号: TP391; TP181 **文献标志码:** A **文章编号:** 1673-4785(2021)01-0011-10

中文引用格式: 高淑萍, 赵清源, 齐小刚, 等. 改进 MobileNet 的图像分类方法研究 [J]. 智能系统学报, 2021, 16(1): 11-20.

英文引用格式: GAO Shuping, ZHAO Qingyuan, QI Xiaogang, et al. Research on the improved image classification method of MobileNet[J]. CAAI transactions on intelligent systems, 2021, 16(1): 11-20.

Research on the improved image classification method of MobileNet

GAO Shuping, ZHAO Qingyuan, QI Xiaogang, CHENG Mengfei

(School of Mathematics and Statistics, Xidian University, Xi'an 710126, China)

Abstract: This paper proposes an improved strategy for the MobileNet neural network (L-MobileNet) because the feature extraction ability of a neural network structure is insufficient, and the classification accuracy is not high on the dataset containing complex image features. First, the original standard convolution form is replaced by the depth separable convolution form, and the feature map obtained from the deep convolution layer is reversed and transferred to the next layer through the deep convolution fusion layer. Second, the leaky ReLU activation function is used to replace the original ReLU activation function to retain more positive and negative feature information in the image, and residual structure is added to avoid the gradient diffusion phenomenon. Finally, the experimental results showed that when compared with six methods, L-MobileNet achieved the best results in the datasets of Cifar-10, Cifar-100 (coarse), Cifar-100 (fine), and Dogs vs Cats.

Keywords: convolutional neural network; image classification; feature extraction; MobileNet; depth separable convolution; activation function; Leaky ReLU; residual structure

图像分类是通过涵盖在图像里的不同特征信息来对图像集进行归类, 它是目前计算机视觉领域的基本问题之一, 也是行为检测、目标识别、图像修复、图像切割等复杂视觉任务的基础。图像分类广泛应用于诸多领域, 包括医学影像识别等。早期图像分类技术以人工提取特征为主, 而

后发展为特征表达的学习, 首先是利用尺度不变特征转换 (scale-invariant feature transform, SIFT)、方向梯度直方图 (histogram of oriented gradient, HOG) 等方法进行底层特征提取; 其次是利用核词典编码、Fisher 向量编码等进行特征编码; 然后将编码后的特征进行汇聚; 最后利用支持向量机等分类器进行图像分类。虽然特征表达的学习方式能够提取一部分图像特征, 但这种方式容易出现特征丢失、泛化性能差的问题, 最终难以达

收稿日期: 2020-12-31.

基金项目: 国家自然科学基金项目 (91338115); 高等学校学科创新引智基地“111”计划 (B08038).

通信作者: 赵清源. E-mail: zqy353364144@163.com.

到理想图像分类精度的要求。

深度学习中卷积神经网络 (convolutional neural network, CNN) 能够完成输入信息从浅层学习到深度学习的逐级表达, 提取更加准确的特征, 降低在图像分类技术上的操作难度并取得显著效果。在图像分类领域, 最初是由 Yann LeCun^[1] 将 CNN 部署在手写体识别上, 取得了良好的识别效果; 2012 年后, 相继出现了 AlexNet、GoogLeNet^[2]、ResNet^[3] 等具有代表性的深度网络结构; 近几年, 国内外提出轻量化版本的网络结构, 如 GhostNet^[4]、ShuffleNetV1^[5]、MobileNetV1^[6]、MobileNetV2^[7]、ShuffleNetV2^[8]、SqueezeNet^[9] 和 IGCv3^[10], 其中 MobileNet 系列更具代表性。在 MobileNetV1 中, 以深度可分离卷积替换标准卷积的方式降低了网络结构的大小和计算复杂度, 而 MobileNetV2 在保持深度可分离卷积的基础上, 加入多个逐点卷积, 进一步提升网络结构性能和降低计算复杂度。但 MobileNet 系列的网络结构在应用于图像分类任务时, 因其网络内部设计、优化策略和激活函数的原因, 导致会不断地忽略输入信息中的负值特征信息, 而这些负值特征信息却可能是图像分类的关键点。近 3 年对 MobileNet 网络结构在图像分类领域的改进方法被相继提出。如 2019 年, 黄跃珍等^[11] 利用 Squeeze-and-Excitation 模块对 MobileNet 网络结构进行改进, 提高了车型识别精度; 刘鸿智^[12] 在 MobileNet 网络结构的基础上, 采用新的混合链接结构, 提高了在数据集 Cifar-10 和 Cifar-100 的分类精度; 2020 年, Guo 等^[13] 提出在 MobileNet 网络结构中插入八度卷积和中心损失函数的方法, 提升了多国人脸分类识别率。

本文在前人工作的基础上, 提出一种改进的 MobileNet 网络结构, 即 L-MobileNet。其主要思想为

1) 用深度可分离卷积形式代替标准卷积形式, 并将深度卷积得到的特征图执行取反操作, 通过深度卷积融合层传递至下一层;

2) 卷积层中使用 Leaky ReLU 激活函数以保留图像更多的正负特征信息, 并加入类残差结构^[3] 避免梯度弥散现象, 降低网络参数计算量。

为了提高网络结构的泛化能力, 在数据输入时采用数据预处理技术。通过将本文 L-MobileNet 与经典的 CNN、MobileNetV1、MobileNetV2、经典的 CNN+Leaky ReLU^[14]、MobileNetV1+Leaky ReLU^[15] 和 MobileNetV2+Leaky ReLU^[15] 进行对比实验, 结果表明 L-MobileNet 在 Cifar-10^[16]、Cifar-100 (coarse)^[16]、Cifar-100 (fine)^[16] 和 Dogs vs Cats^[17] 数据集上具有更好的图像分类效果。

1 相关预备知识

1.1 经典的 CNN 网络结构

经典的 CNN 属于前馈神经网络类型, 其结构组件有标准卷积层、池化层、全连接层、激活函数和分类器, 如图 1 所示。该网络主要通过标准卷积层进行图像特征提取, 在该层中通常包括若干特征图, 每个特征图由神经元节点排列成矩形形状, 相同特征图中神经元节点之间进行权值共享。通过权值共享可以减少神经元节点之间的连接数量, 避免过拟合现象。池化层能够降低计算维度, 进而提升网络结构的泛化性能。激活函数用来增强网络结构的非线性学习能力。经过多次特征提取过程后, 把最后的特征图进行拼接, 得到一维矩阵。最后由全连接层输入到分类器进而输出分类结果。

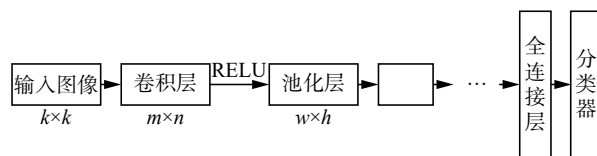


图 1 经典的 CNN 网络结构

Fig. 1 Classical CNN network structure

经典的 CNN 网络结构利用卷积核提取图像中的特征, 对局部进行感知并且通过权值共享降低参数计算量。但经典的 CNN 反向传播过程中容易出现参数不更新现象, 和池化层丢失图像中特征信息的问题。

1.2 MobileNet 网络结构

MobileNet 网络系列在近几年被广泛应用, 更是轻量级网络的典型代表。MobileNet 网络结构在经典的 CNN 网络结构基础上, 将池化层、部分全连接层用卷积层来代替, 其突出特点是引入了深度可分离卷积的卷积形式。标准卷积是将一组卷积核与输入数据作用后组合成单通道特征的输出, 而深度可分离卷积是因式分解的卷积形式, 将标准卷积分解为深度卷积和 1×1 逐点卷积 2 部分, 深度卷积是把单个固定大小的卷积核作用到每个输入通道, 然后逐点卷积通过 1×1 的卷积核来进行通道信息融合并输出。图 2 显示标准卷积 (图 2(a)) 分解为深度卷积 (图 2(b)) 和 1×1 逐点卷积 (图 2(c)) 的过程。

图 2 中 M 表示输入通道数量, N 表示输出通道数量, $D_K \times D_K$ 表示卷积核大小的乘积, $D_F \times D_F$ 表示特征映射图大小的乘积。

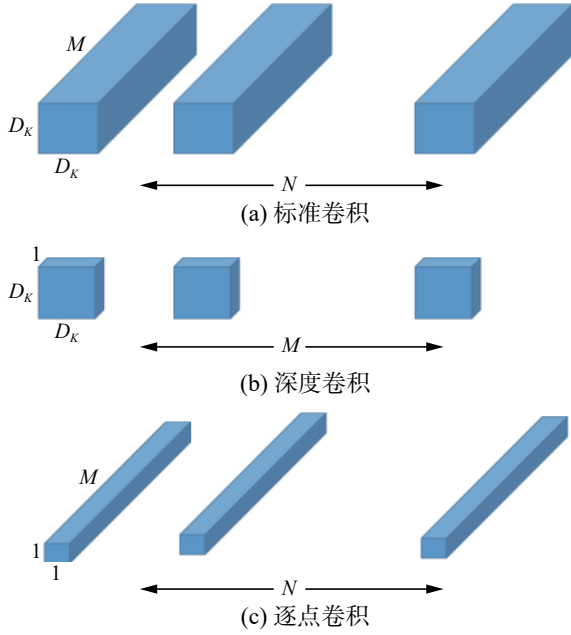


图 2 标准卷积分解过程

Fig. 2 Standard convolution decomposition process

深度卷积计算成本为

$$D_K \times D_K \times M \times D_F \times D_F \quad (1)$$

逐点卷积计算成本为

$$M \times N \times D_F \times D_F \quad (2)$$

深度可分离卷积计算成本占标准卷积计算成本的比例为

$$\frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} \quad (3)$$

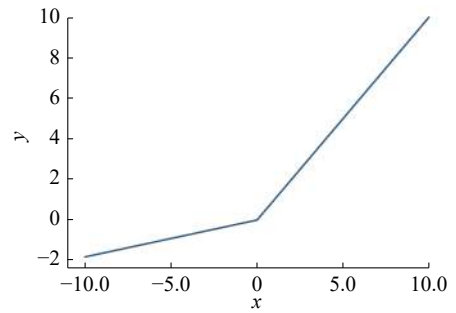
实验表明^[6], 当使用 3×3 的深度可分离卷积时, 其计算量比标准卷积少 8~9 倍, 而精度只有很小的损失。MobileNet 网络系列在尽可能保证图像分类精度稳定的前提下, 极大缩短网络训练的时间与减少参数更新过程中的计算成本, 也为后续网络结构的优化提供了方向。但 MobileNet 网络结构中仍存在着特征信息提取能力不足导致分类精度不高和网络层中激活函数丢失负值特征信息的现象。

1.3 Leaky ReLU 激活函数^[18]

激活函数是神经网络结构中非线性学习能力的关键点, ReLU 激活函数可以有效缓解梯度弥散现象, 成为大多数激活函数的主流选择。但随着网络训练轮数的增加, 部分神经元对应权重得不到更新, 出现神经元死亡现象, 并且 ReLU 在特征提取过程中会丢失特征图内的负值特征信息。ReLU 输出值的均值始终大于零, 也不利于网络学习能力的表达。Leaky ReLU 激活函数的特性解决了 ReLU 的问题, 有效提取负值特征信息, 数学形式为

$$y_i = \begin{cases} x_i, & x_i \geq 0 \\ \frac{x_i}{a_i}, & x_i < 0 \end{cases} \quad (4)$$

式中: x_i 代表第 i 层的输出; y_i 代表第 i 层经过非线性变换后的输出; a_i 是第 i 层中的固定参数, 其范围是 $(1, +\infty)$ 。Leaky ReLU 激活函数如图 3 所示。由于 Leaky ReLU 激活函数负半轴是一个较小斜率的函数, 可以用来初始化神经元, 避免神经元死亡的同时增加了负值特征信息的提取。由文献 [18] 可知, 当 Leaky ReLU 激活函数中 $a_i = 5.5$ 时, Leaky ReLU 激活函数的分类效果要优于 ReLU 激活函数。

图 3 Leaky ReLU 激活函数 ($a_i = 5.5$)Fig. 3 Leaky ReLU activation function ($a_i = 5.5$)

2 L-MobileNet 网络结构

针对经典的 CNN 网络结构和 MobileNet 网络结构特征提取能力不足和易忽略负值特征信息的问题, 本文提出一种改进的 MobileNet 网络结构的图像分类方法, 即 L-MobileNet。首先, 研究了基于 MobileNet 网络中存在特征信息的各个局部区域; 其次, 设计了深度卷积取反层和深度卷积融合层, 将原图像特征和其取反特征进行融合提取; 最后, 加入类残差结构, 组合卷积块, 通过损失函数优化整个网络进行分类。

L-MobileNet 网络结构如图 4 所示, 参数设置如表 1 所示。深度卷积层得到的特征图通过深度卷积取反层进行取反操作, 并将其与原特征图的特征信息一起通过深度卷积融合层进行特征相加, 传递至下一个逐点卷积层, 然后传递至下一个卷积块, 以此类推。其中的卷积块中采用 Leaky ReLU 激活函数, 将其作用于每一个卷积块中的每一个卷积层。为了防止出现梯度弥散现象, 在卷积块中加入了类残差结构, 先进行信息通道的扩展再进行信息通道的压缩。通过交叉熵损失函数优化整个网络, 直至收敛。最终经过全局平均池化层和逐点卷积层来达到分类数目要求, 通过 Softmax 分类器进行输入样本的分类。

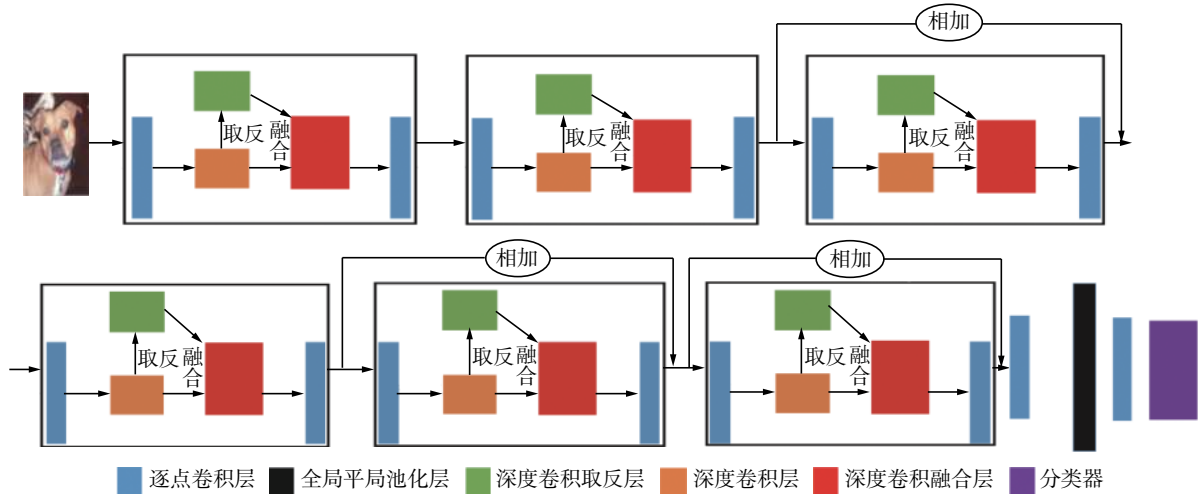


图 4 L-MobileNet 网络结构

Fig. 4 L-MobileNet network structure

表 1 L-MobileNet 参数设置

Table 1 L-MobileNet parameters setting

层名称	卷积核		步长	填充	层名称	卷积核		步长	填充
	大小	个数				大小	个数		
输入层	—	—	—	—	逐点卷积层8	1×1	64	1	有
逐点卷积层1	1×1	32	1	有	逐点卷积层9	1×1	384	1	有
深度卷积1	3×3	32	1	有	深度卷积5	3×3	384	1	有
逐点卷积层2	1×1	24	1	有	逐点卷积层10	1×1	64	1	有
逐点卷积层3	1×1	144	1	有	逐点卷积层11	1×1	384	1	有
深度卷积2	3×3	144	2	有	深度卷积6	3×3	384	1	有
逐点卷积层4	1×1	32	1	有	逐点卷积层12	1×1	64	1	有
逐点卷积层5	1×1	196	1	有	逐点卷积层13	1×1	384	1	有
深度卷积3	3×3	196	1	有	全局平均池化	—	—	—	—
逐点卷积层6	1×1	32	1	有	逐点卷积层14	1×1	—	1	有
逐点卷积层7	1×1	196	1	有	—	—	—	—	—
深度卷积4	3×3	196	2	有	输出层	—	—	—	—

2.1 特征信息的融合

如图 4 所示, 本文对特征信息的融合采用的是在深度卷积层提取了原始图像局部特征后, 同步进行深度卷积取反层, 然后利用深度卷积融合层对 2 层特征信息进行处理。符号说明如下: 对于给定的图像, 其中 l 表示当前层, $l-1$ 表示前一层, x_j^l 表示当前层第 j 个特征图, \circ 代表深度卷积运算, x_i^{l-1} 表示前一层第 i 个特征图, k_{ij}^l 表示该层第 i 个特征图和前一层第 j 个特征图所作用的卷积核, b_j^l 表示偏置值, M_j 表示产生该层第 j 个特征图相关的前一层特征图累计值, $f(\cdot)$ 表示 Leaky ReLU 激活函数。

深度卷积层对输入的处理为

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} \circ k_{ij}^l + b_j^l\right) \quad (5)$$

深度卷积取反层同步对输入进行处理为

$$x_j^l = f\left[-\left(\sum_{i \in M_j} x_i^{l-1} \circ k_{ij}^l + b_j^l\right)\right] \quad (6)$$

深度卷积融合层将前 2 步提取的信息进行融合为

$$x_j^l = f\left[x_i^{l-1} \circ k_{ij}^l + b_j^l\right] + f\left[-\left(x_i^{l-1} \circ k_{ij}^l + b_j^l\right)\right], \quad i \in M_j \quad (7)$$

深度卷积融合层所提取出的特征, 不仅包含了原始深度卷积层中的特征信息, 而且对潜在的图像局部区域的特征进行提取, 融合后的特征随着网络训练的进行, 不断学习出识别效果优秀的高维特征信息。

2.2 负值特征信息的提取

如图 4 可知, L-MobileNet 网络结构中有 6 个大卷积块, 为了提高 MobileNet 网络中对负值特征信息的敏感度和负值特征信息传播过程中的完

整度, 本文将原始的 ReLU 激活函数替换为 Leaky ReLU 激活函数, 并引入类残差结构。经典的 CNN 受限于固定的卷积核大小和激活函数的选取, 影响了负值特征信息的提取和传播, 而 MobileNet 网络虽对卷积核形式有所改变但对负值特征信息量仍提取不足。由于 Leaky ReLU 激活函数自身特性对负值敏感, 能够保障神经元不死亡, 并且类残差结构可以防止梯度弥散, 利于负值特征信息的传播, 所以本文在每一个卷积层后使用 Leaky ReLU($a_i = 5.5$) 激活函数和在第 3、5 及第 6 卷积块使用类残差结构。

由于前向传播过程中前面网络层中梯度的计算尚可保持较大值, 而越靠后的网络层中梯度值会越来越小, 导致反向传播过程中极易出现梯度弥散现象, 所以本文将类残差结构放置在网络中后端位置。以其中一个卷积块为例, 如果特征向量的维度越低, 卷积层中的计算量越小, 导致网络结构整体计算速度越快, 但这并不能够提取出输入图像全面且有效的特征信息。为了达到网络结构平衡的目的, 本文采取含有高维特征向量的设计: 卷积块的输入端采用 1×1 的逐点卷积结构, 目的是将特征信息从低维空间映射到高维空间, 同时受文献 [7] 的启发, 将其中负责维度扩展的超参数的默认值设置为 6。卷积块的输出端同样采用 1×1 的逐点卷积结构, 在没有超参数的设定下, 可以将特征信息从高维空间映射到低维空间。

逐点卷积计算为

$$x'_j = f\left(\sum_{i \in M_j} x_i^{l-1} \cdot k_{ij}^l + b_j^l\right) \quad (8)$$

式中: (\cdot) 代表逐点卷积运算; 其余符号与式 (5) 相同。为了保证负值特征信息的完整性, 在卷积块中每一个卷积层后, 摒弃掉可能会造成信息丢失的 ReLU 激活函数, 使用 Leaky ReLU 激活函数。卷积块以逐点卷积、深度卷积、逐点卷积交错的方式排列。

2.3 传播过程中的参数更新原理

在反向传播的参数更新中, L-MobileNet 网络结构使用 Tensorflow 框架下 tf.average_pooling2d 进行全局平均池化, 根据输入信息的类别数目由逐点卷积层输出。

用 Softmax 分类器来进行输出分类为

$$Y(p) = P(y = p | \mathbf{x}, \theta_p) = \frac{e^{\mathbf{x}^T \theta_p}}{\sum_{p=1}^C e^{\mathbf{x}^T \theta_p}} \quad (9)$$

采用交叉熵优化的损失函数为

$$\min_{\{W, b, \theta\}} \varepsilon(W, b; \theta) = \frac{1}{N} \sum_{n=1}^N \left[- \sum_{p=1}^C \gamma(y_n = p) \log P(y = p | \mathbf{x}, \theta_p) \right] + \lambda R(\theta) \quad (10)$$

式中: C 表示类别; N 表示样本数; γ 表示 Dirichlet 函数; 参数 $\theta = (\theta_1, \theta_2, \dots, \theta_k)$; $R(\cdot)$ 表示正则化约束项; λ 表示正则化系数。

损失函数 ε 对输入 c 的导数为

$$\frac{\partial \varepsilon}{\partial c} = \frac{\partial \varepsilon}{\partial f(c)} \frac{\partial f(c)}{\partial c} \quad (11)$$

$$\frac{\partial f(c)}{\partial c} = \begin{cases} 1, & c \geq 0 \\ \frac{1}{a_i}, & c < 0 \end{cases} \quad (12)$$

式中: ε 表示损失函数; c 表示前一网络层的输出; $f(c)$ 表示前一网络层经过非线性变换后的输出; a_i 是 $(1, +\infty)$ 的固定参数; $\frac{\partial \varepsilon}{\partial f(c)}$ 表示从网络深层传播来的梯度; $\frac{\partial f(c)}{\partial c}$ 表示当前层输入值的梯度。当增大前向传播中对特征信息的存储后, 将减小训练样本的误差, 再通过反向传播算法进一步影响了网络结构中的参数更新。

参数的更新规则为

$$\Delta c \leftarrow \mu \Delta c + \omega \alpha c + \alpha \frac{\partial \varepsilon}{\partial c} \quad (13)$$

$$c \leftarrow c + \Delta c \quad (14)$$

式中: c 表示前一网络层的输出; ε 表示损失函数; μ 表示动量系数; α 表示学习率; ω 表示权重衰减系数。本文 L-MobileNet 通过影响网络中参数 c 和 $\frac{\partial \varepsilon}{\partial c}$ 更新, 最终影响到网络结构的分类结果。

网络模型的参数量 (空间复杂度) 的计算为

$$C_{\text{Space}} \sim O\left(\sum_{l=1}^D k_l^3 \times C_{l-1} \times C_l + \sum_{l=1}^D M^3 \times C_l\right) \quad (15)$$

网络模型的计算量 (时间复杂度) 的计算为

$$\text{Time} \sim O\left(\sum_{l=1}^B D_F^3 \times k_l^3 \times C_{l-1} + D_F^3 \times C_l \times C_{l-1}\right) \quad (16)$$

式中: C_l 为第 l 层卷积核个数; C_{l-1} 为上一层的输出通道数; k_l 为卷积核尺寸; M 为输出特征映射的边长; D 为所有卷积层数; $O(\cdot)$ 为所需要的空间资源量。由于本文使用的激活函数是 Leaky ReLU 函数, 因此可以通过原位运算完成从输入到输出的非线性映射, 排除输出特征映射的参数量。根据式 (15)、(16) 计算表明, L-MobileNet 参数总量为 280 万、计算量为 320 百万次, 与 MobileNetV1^[6] 的 420 万个参数、575 百万次计算量和 MobileNetV2^[7] 的 340 万个参数、300 百万次计算量相比, 本文方法参数量更少。

2.4 算法步骤

L-MobileNet 网络结构的初始学习率设置为 0.1, 随着训练轮数增加和训练误差变小, 学习率呈自适应状态且衰减系数为 0.96, 直到达到最大训练轮数时停止训练。对于数据集 $D = \{(x_n, y_n)\}_{n=1}^N$, L-MobileNet 网络算法步骤为

输入 训练数据集 $D = \{(x_n, y_n)\}_{n=1}^N$, L-MobileNet 网络结构, 图像分类数, 每一批的数量, 自适应学习率的初始值和衰减系数。

1) 利用预处理技术将数据集 $D = \{(x_n, y_n)\}_{n=1}^N$ 中的训练集与测试集进行处理, 转换为符合 L-MobileNet 网络结构输入的四维数组即每一批输入的图片数和每一批次输入的标签数。

2) 将每一批输入的图片数和每一批次输入的标签数导入网络架构文件中, 并初始化 L-MobileNet 网络中所有的权值 k 和偏置 b 。

3) 令 $l = 1$ 。

4) 用 L-MobileNet 网络对输入进行训练, 计算训练样本的实际输出值及每一网络层所产生的误差总和。

5) 用 RMSProp 算法优化器计算网络中权值和偏置的偏导数以此更新相应参数值。

6) 判断 $l > \text{Epoch}$ 训练轮数是否成立, 成立则停止训练; 不成立, 则返回 4)。

输出 参数 (权值 k 和偏置 b) 已知的 L-MobileNet 网络结构。

3 实验与结果分析

3.1 实验数据集

实验数据取自 Cifar-10、Cifar-100 和 Dogs vs Cats。Cifar-10 数据集中有 10 种类别, 由 60 000 张 32×32 彩色图片组成, 每种类别包含 6 000 张图片, 总计有 10 000 张测试图片和 50 000 张训练图片。Cifar-10 数据集分为 1 个测试批次和 5 个训练批次, 测试批次是由每种类别中随机选择的 1 000 张图片组成, 其余 5 个训练批次是由每种类别中随机的剩余图片组成 (某个训练批次可能包含某个类别的图片数量多于另一个类别的图片数量)。图 5 显示 Cifar-10 数据集中部分样本。Cifar-100 数据集与 Cifar-10 数据集结构类似, 但有 coarse 和 fine 2 种数据构成。图 6 显示 Dogs vs Cats 数据集中部分样本。实验采用的 Cifar-10、Cifar-100(coarse) 和 Cifar-100(fine) 数据集划分采取 5:1 的比例, 在 60 000 张图片中, 50 000 张图片用于训练, 10 000 张图片用于测试验证。Dogs vs Cats 数据集各类取前 2 000 张图片用于训练, 各

类另取 400 张图片用于测试验证。在实验前将 4 个数据集进行了左右翻转、上下翻转、调整图片亮度、调整图片对比度和白化的数据预处理技术。



图 5 Cifar-10 数据集的部分图例
Fig. 5 Partial legend of Cifar-10 dataset



图 6 Dogs vs Cats 数据集的部分图例
Fig. 6 Partial legend of Dogs vs Cats dataset

3.2 实验仿真环境

为验证 L-MobileNet 网络结构有效性, 本文将其与经典的 CNN、MobileNetV1、MobileNetV2、经典的 CNN+Leaky ReLU, MobileNetV1+Leaky ReLU 和 MobileNetV2+Leaky ReLU 进行对比实验, 并评价各网络结构, 所有对比实验均采用相同卷积层设置, 且每层参数设置也相同。

实验仿真环境采用 tensorflow1.14 框架, 编程语言为 Python3.7, 框架平台为 PyCharm2019.1.1, 使用 Windos10, 64 位操作系统, CPU 使用 Intel Core i7-7700K, GPU 使用 NVIDIA GTX1080 8 G, 内存为 32 G 以实现所有实验。

3.3 训练过程及实验对比

在 L-MobileNet 网络输入层中, 按照数据集中图片大小输入即可, 训练及测试样本 batch_size 为 128。卷积层中, 逐点卷积使用 1×1 卷积核, 深度卷积使用 3×3 卷积核, 并对每次卷积操作进行 Leaky ReLU 激活函数的非线性特征提取, 利用深度卷积融合层来产生更多的特征, 使得模型增强数据分类学习过程中的控制力。训练过程中使用 RMSPropOptimizer 优化策略, decay 为 0.9, momentum 为 0.9。学习率采用自适应学习率, decay_rate 为 0.96。对于 Cifar-10、Cifar-100(fine)、Cifar-100(coarse) 和 Dogs vs Cats 数据集, 经典的 CNN、MobileNetV1、MobileNetV2 以及 L-MobileNet 4 种

模型分别通过 50、50、50 和 200 个 Epoch 训练, 每经过一个 Epoch, 用测试集进行准确率的验证和

记录, 所有网络模型均达到收敛状态, 4 种模型随 Epoch 的准确率变化曲线如图 7 所示。

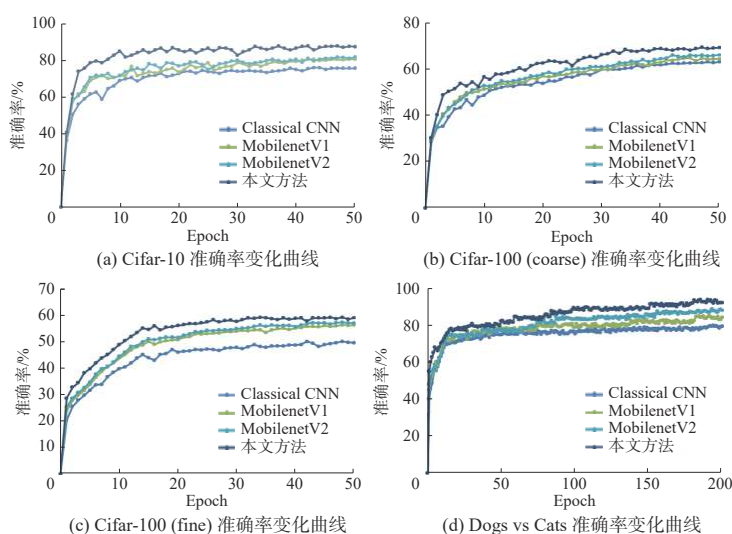


图 7 不同数据集下准确率变化曲线

Fig. 7 Accuracy curve under different datasets

在相同的参数设置和卷积层设置下, 本文还将经典的 CNN、MobileNetV1 以及 MobileNetV2 分别作用于 Leaky ReLU 激活函数同 L-MobileNet 进行对比实验, 在 Cifar-10、Cifar-100(fine)、Ci-

far-100(coarse) 和 Dogs vs Cats 数据集上, 分别经过 50、50、50 和 200 个 Epoch 训练, 所有网络模型均达到收敛状态, 其准确率随 Epoch 的变化曲线如图 8 所示。

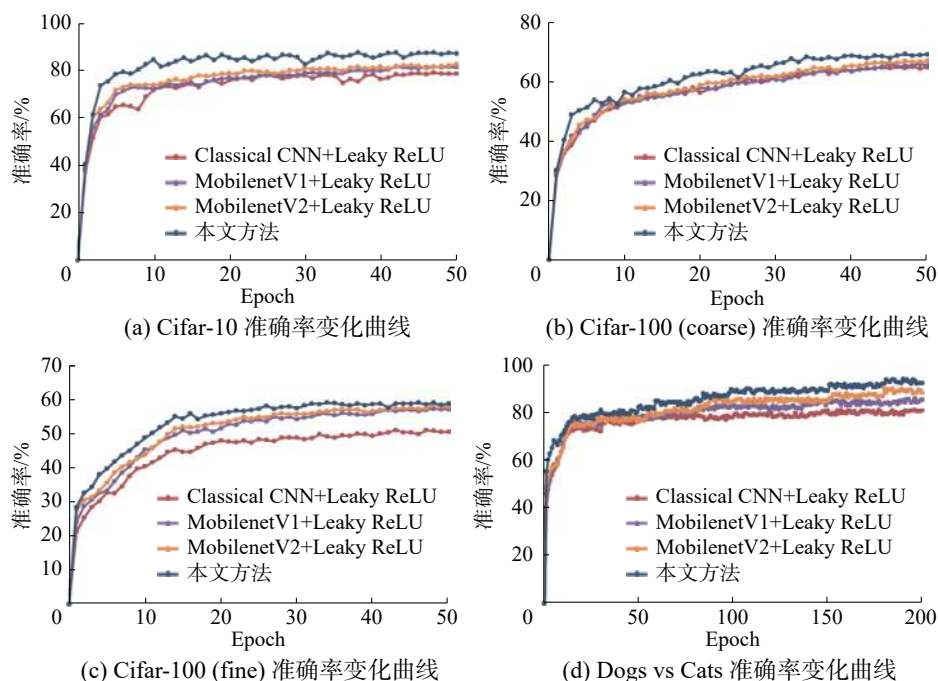


图 8 不同数据集下准确率变化曲线

Fig. 8 Accuracy curve under different datasets

可以看出, 对于经典的 CNN、MobileNetV1 以及 MobileNetV2 网络结构中使用 Leaky ReLU 激活函数时, 虽然分类准确率有所提升, 但仍然低于 L-MobileNet 的分类准确率, 这表明本文 L-Mo-

bileNet 是有效的。

3.4 实验结果及分析

本文使用的评价指标是最高准确率和平均准确率。最高准确率是在所有训练轮数中正确分类

的样本数占样本总数的比例最大值,平均准确率是所有训练轮数的准确率平均值。在 Cifar-10、Cifar-100(fine)、Cifar-100(coarse) 和 Dogs vs Cats 4 个数据集中所有网络结构的平均准确率数据和最高准确率数据如表 2 所示。

从图 7 和表 2 可知,在 4 个数据集分类实验中,本文提出的 L-MobileNet 的分类准确率要明显高于经典的 CNN、MobileNetV1 以及 MobileNetV2。在 Cifar-10 数据集的实验结果中,经典的 CNN 的平均准确率为 70.31%, MobileNet 系列的准确率相比经典的 CNN 有较大提升,平均准确率分别达到了 74.31% 和 75.65%。在最高准确率方面, MobileNetV1 和 MobileNetV2 也比经典的 CNN 分别提升了 4.57% 和 5.67%,但仍然低于 L-MobileNet 中 83.11% 的平均准确率和 87.36% 的最高准确率。在 Cifar-100(coarse) 和 Cifar-100(fine) 数据集的实验结果中,可以看出 L-MobileNet 仍是达到最

好结果,在最高准确率方面,相较于 MobileNetV1 分别提升了 4.56% 和 2.84%,比 MobileNetV2 分别提升了 3.26% 和 1.84%。在 Dogs vs Cats 数据集的实验结果中,因类别数较少,整体分类精度较高。MobileNetV1 和 MobileNetV2 在平均准确率和最高准确率上的结果相较于经典的 CNN 提升明显, MobileNetV1 分别达到了 78.48% 和 84.96%,而 MobileNetV2 分别达到了 80.87% 和 88.81%,本文提出的 L-MobileNet 比经典的 CNN 在平均准确率上提升了 10.23%,最高准确率上提升了 14.08%,分别达到了 85.51% 和 93.94% 的分类精度。出现这样结果的原因是 MobileNet 系列使用了改进的深度可分离卷积结构,改善了经典的 CNN 网络结构中的池化等丢失特征信息的问题,而本文方法在 MobileNet 的基础上,利用深度卷积融合层和 Leaky ReLU 激活函数可以更全面的提取特征信息。

表 2 7 种网络结构在 4 种数据集中准确率统计
Table 2 Accuracy statistics of seven network structures in four datasets

模型结构	Cifar-10		Cifar-100(coarse)		Cifar-100(fine)		Dogs vs Cats	
	平均	最高	平均	最高	平均	最高	平均	最高
	准确率	准确率	准确率	准确率	准确率	准确率	准确率	准确率
经典 CNN	70.31	75.69	54.69	62.97	43.92	49.86	75.28	79.86
经典 CNN+Leaky ReLU	73.42	78.86	57.28	64.85	44.97	50.96	76.92	80.98
MobileNetV1	74.31	80.26	56.38	64.58	49.21	56.12	78.48	84.96
MobileNetV1+Leaky ReLU	75.54	81.54	57.51	65.86	50.17	57.19	79.88	85.87
MobileNetV2	75.65	81.36	57.47	65.88	50.24	57.12	80.87	88.81
MobileNetV2+Leaky ReLU	77.11	82.36	58.51	66.78	51.19	58.33	81.65	89.96
L-MobileNet	83.11	87.36	61.63	69.14	53.61	58.96	85.51	93.94

从图 8 和表 2 可知,将经典的 CNN、MobileNetV1 以及 MobileNetV2 分别作用于 Leaky ReLU 激活函数后,实验结果均不及本文提出的 L-MobileNet。经典的 CNN+Leaky ReLU 在 4 个数据集中都处于最差水平, L-MobileNet 在 4 个数据集的平均准确率和最高准确率都处于最好水平。在 Cifar-10、Cifar-100(fine)、Cifar-100(coarse) 和 Dogs vs Cats 4 个数据集的最高准确率方面, L-MobileNet 比 MobileNetV1+Leaky ReLU 分别提升了 5.82%、3.28%、1.77% 和 8.07%,并且改进的 MobileNet 比 MobileNetV2+Leaky ReLU 分别提升了 5%、2.36%、0.63% 和 3.98%。实验结果说明了网络结构与 Leaky ReLU 激活函数的结合的确可以提高图像分类准确率,原因是 Leaky ReLU 激活函

数对潜在局部区域的负值特征信息更敏感。

综合以上实验结果表明,本文对 MobileNet 的改进是有效的,所提出的 L-MobileNet 网络结构具有较好的分类效果。L-MobileNet 在不同的数据集上表现出优秀的泛化能力,具有良好的推广应用前景。

为了进一步验证本文方法的有效性,在 Cifar-10 数据集上将 L-MobileNet 网络结构与近几年图像分类上具有代表性的 8 种方法进行分类精度的对比,比较结果如表 3 所示。

由表 3 可以得出,与传统学习方法 (Mean-covariance RBM、Improved local coord、PCANet) 相比,本文提出的 L-MobileNet 优势明显,且有较大提升。但与其他基于深度学习的方法 (Highway、

Maxout+Dropout、NIN) 相比较, 分类精度尚有不足, 主要原因是 NIN 网络结构在卷积层后加入了 BN 和 Dropout 技术, 而 Maxout 网络结构和 Highway 网络结构使用了更多的网络层数, 属于深度网络。本文提出的 L-MobileNet 网络结构在低网络结构中具有一定的优势。

表 3 9 种不同方法在 Cifar-10 数据集上的分类精度
Table 3 Classification accuracy of 9 different methods on Cifar-10 dataset

方法	分类精度/%
Mean-covariance RBM(3 layers) ^[19]	71.00
Improved Local Coord. Coding ^[20]	74.50
PCANet ^[21]	78.67
Stochastic pooling ConvNet ^[22]	84.87
Improved LeNet-5 ^[23]	86.78
Conv. Maxout+Dropout ^[24]	88.32
NIN ^[25]	89.59
Highway(19 layers) ^[26]	92.46
L-MobileNet	87.36

4 结束语

为了解决具有复杂特征图像分类中由于特征信息提取能力不足, 尤其是对负值特征信息不敏感而造成图像分类精度下降的问题, 本文提出了一个改进的 MobileNet 网络结构 (L-MobileNet)。通过在 4 个公开数据集上进行的实验结果表明, 与其他 6 种神经网络图像分类方法相比, L-MobileNet 分类精度高, 在轻量级网络中具有优势, 可复制能力强, 可方便移植到移动设备, 使相关设备在包含复杂特征图像场景下, 具有良好的分类能力。进一步的研究工作内容: 1) 验证本文改进方法在更深网络结构下的有效性, 进一步提升模型分类性能; 2) 进一步优化网络模型结构, 以适应没有标签样本的数据集。

参考文献:

- [1] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278–2324.
- [2] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with con-volutions[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1–9.
- [3] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770–778.
- [4] HAN K, WANG Y, TIAN Q, et al. GhostNet: More features from cheap operations[J]. *arXiv preprint arXiv: 1911.11907*, 2019.
- [5] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 6848–6856.
- [6] HOWARD A G, ZHU Menglong, CHEN Bo, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications[J]. *arXiv: 1704.04861*, 2017.
- [7] SANDLER M, HOWARD A, ZHU Menglong, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4510–4520.
- [8] MA Ningning, ZHANG Xiangyu, ZHENG Haitao, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany, 2018: 116–131.
- [9] IANDOLA F N, HAN Song, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[J]. *arXiv: 1602.07360*, 2016.
- [10] SUN Ke, LI Mingjie, LIU Dong, et al. IGCv3: interleaved low-rank group convolutions for efficient deep neural networks[J]. *arXiv: 1806.00178*, 2018.
- [11] 黄跃珍, 王乃洲, 梁添才, 等. 基于改进型 MobileNet 网络的车型识别方法 [J]. *电子技术与软件工程*, 2019(1): 22–24.
HUANG Yuezhen, WANG Naizhou, LIANG Tiancai, et al. Vehicle identification method based on improved mobilenet network[J]. *Electronic technology and software engineering*, 2019(1): 22–24.
- [12] 刘鸿智. 面向移动设备的轻型神经网络的改进与实现 [D]. 呼和浩特: 内蒙古大学, 2019.
LIU Hongzhi. Improvement and implementation of lightweight neural network for mobile devices[D]. Hohhot: Inner Mongolia University, 2019.
- [13] 郭奕君, 努尔毕亚·亚地卡尔, 朱亚俐, 等. 基于 MobileNet 网络多国人脸分类识别 [J]. *图像与信号处理*, 2020, 9(3): 146–155.

- GUO Yijun, ABUDIRIYIMU A, YADIKAR N, et al. Multinational face classification and recognition based on MobileNet network[J]. Journal of image and signal processing, 2020, 9(3): 146–155.
- [14] DUBEY A K, JAIN V. Comparative study of convolution neural network's relu and leaky-relu activation functions[M]. MISHRA S, SOOD Y R, TOMAR A. Applications of Computing, Automation and Wireless Systems in Electrical Engineering. Singapore: Springer, 2019: 873–880.
- [15] CHEN Yinpeng, DAI Xiyang, LIU Mengchen, et al. Dynamic ReLU[J]. arXiv: 2003.10027, 2020.
- [16] KRIZHEVSKY A. Learning multiple layers of features from tiny images[R]. Toronto: University of Toronto, 2009.
- [17] DUGAN P, CUKIERSKI W, SHIU Y, et al. Kaggle competition[J]. Cornell University, the ICML, 2013.
- [18] XU Bing, WANG Naiyan, CHEN Tianqi, et al. Empirical evaluation of rectified activations in convolutional network[J]. arXiv: 1505.00853, 2015.
- [19] RANZATO M A, HINTON G E. Modeling pixel means and covariances using factorized third-order Boltzmann machines[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA, 2010: 2551–2558.
- [20] YU Kai, ZHANG Tong. Improved local coordinate Coding using local tangents[C]//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel, 2010.
- [21] CHAN T H, JIA Kui, GAO Shenghua, et al. PCANet: a simple deep learning baseline for image classification?[J]. IEEE transactions on image processing, 2015, 24(12): 5017–5032.
- [22] ZEILER M D, FERGUS R. Stochastic pooling for regularization of deep convolutional neural networks[J]. arXiv: 1301.3557, 2013.
- [23] 刘金利, 张培玲. 改进 LeNet-5 网络在图像分类中的应用 [J]. 计算机工程与应用, 2019, 55(15): 32–37, 95.
- LIU Jinli, ZHANG Peiling. Application of improved LeNet-5 network in image classification[J]. Computer engineering and applications, 2019 (15): 5.
- [24] GOODFELLOW I J, WARDE FARLEY D, MIRZA M, et al. Maxout networks[J]. arXiv preprint arXiv: 1302.4389, 2013.
- [25] LIN Min, CHEN Qiang, YAN Shuicheng. Network in network[J]. arXiv: 1312.4400, 2013.
- [26] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway networks[J]. arXiv: 1505.00387, 2015.

作者简介:



高淑萍, 教授, 主要研究方向为多目标优化理论与应用、数学与信息科学交叉研究、大数据处理与分析。主持、参与国家级和省自然科学基金项目及横向项目多项。发表学术论文 30 余篇。



赵清源, 硕士研究生, 主要研究方向为深度学习、图像分类、算法优化。



齐小刚, 教授, 博士生导师, 主要研究方向为复杂系统建模与仿真、网络算法设计与应用。申请专利 47 项 (授权 19 项), 登记软件著作权 4 项。发表学术论文 100 余篇。