



综述

基于深度学习的图像分类研究综述

苏赋¹, 吕沁¹, 罗仁泽²

(1. 西南石油大学电气信息学院, 四川 成都 610500;

2. 西南石油大学地球科学与技术学院, 四川 成都 610500)

摘要: 近年来, 深度学习在计算机视觉领域中的表现优于传统的机器学习技术, 而图像分类问题是其中最突出的研究课题之一。传统的图像分类方法难以处理庞大的图像数据, 且无法满足人们对图像分类精度和速度的要求, 而基于深度学习的图像分类方法突破了此瓶颈, 成为目前图像分类的主流方法。从图像分类的研究意义出发, 介绍了其发展现状。其次, 具体分析了图像分类中最重要的深度学习方法(即自动编码器、深度信念网络与深度玻尔兹曼机)以及卷积神经网络的结构、优点和局限性。再次, 对比分析了方法之间的差异及其在常用数据集上的性能表现。最后, 探讨了深度学习方法在图像分类领域的不足及未来可能的研究方向。

关键词: 深度学习; 图像分类; 自编码器; 深度信念网络; 卷积神经网络

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019268

Review of image classification based on deep learning

SU Fu¹, LV Qin¹, LUO Renze²

1. School of Electrical and Information Engineering, Southwest Petroleum University, Chengdu 610500, China

2. School of Earth Sciences and Technology, Southwest Petroleum University, Chengdu 610500, China

Abstract: In recent years, deep learning performed superior in the field of computer vision to traditional machine learning technology. Indeed, image classification issue drew great attention as a prominent research topic. For traditional image classification method, huge volume of image data was of difficulty to process and the requirements for the operation accuracy and speed of image classification could not be met. However, deep learning-based image classification method broke through the bottleneck and became the mainstream method to finish these classification tasks. The research significance and current development status of image classification was introduced in detail. Also, besides the structure, advantages and limitations of the convolutional neural networks, the most important deep learning methods, such as auto-encoders, deep belief networks and deep Boltzmann machines image classification were concretely analyzed. Furthermore, the differences and performance on common datasets of these methods were compared and analyzed. In the end, the shortcomings of deep learning methods in the field of image classification and the possible future research directions were discussed.

Key words: deep learning, image classification, auto-encoders, deep belief networks, CNN

收稿日期: 2019-05-23; 修回日期: 2019-11-10

基金项目: 国家重点研发计划基金资助项目 (No.2016YFC0601100); 四川省科技计划基金资助项目 (No.2019CXRC0027)

Foundation Items: The National Key Research and Development Program(No.2016YFC0601100), Sichuan Science and Technology Project (No.2019CXRC0027)

1 引言

图像分类是计算机视觉领域的热门研究方向之一,也是实现物体检测^[1-2]、人脸识别^[3-4]、姿态估计^[5-6]等应用的重要基础,因此图像分类技术有很高的学术研究和科技应用价值。图像分类,即给定一幅输入图像,通过某种分类算法来判断该图像所属类别。图像分类的主要过程包括图像预处理、特征提取和分类器设计。图像预处理包括图像滤波,如中值滤波^[7]、均值滤波^[8]、高斯滤波^[9]以及图像归一化等操作,其主要作用是过滤图像中的一些无关信息,在简化数据的前提下最大限度地保留有用信息,增强特征提取的可靠性。特征提取是图像分类任务中最为关键的一部分,其将输入图像按照一定的规则变换生成另一种具有某些特性的特征表示,新的特征往往具有低维度、低冗余、低噪声、结构化等优点,从而降低了对分类器复杂度的要求,提高了模型性能。最后通过训练分类器对提取的特征进行分类,从而实现图像的分类。

传统的图像分类研究中,多数为基于图像特征的分类,即根据不同类别图像的差异,利用图像处理算法提取相应的经过定性或定量表达的特征,对这些特征进行数学统计分析或使用分类器输出分类结果。在特征提取方面,主要包括纹理、颜色、形状等底层视觉特征,尺度不变特征变换^[10]、局部二值模式^[11]、方向梯度直方图^[12]等局部不变性特征,这些人工设计特征缺乏良好的泛化性能,且依赖于设计者的先验知识和对分类任务的认知理解。目前,海量、高维的数据也使得人工设计特征的难度呈指数级增加。在分类器方面,主要包括 k NN (k -nearest neighbor, k 最近邻)^[13]、决策树^[14]、SVM (support vector machine, 支持向量机)^[15]、人工神经网络^[16]等方法。这些分类器大大地提升了图像分类的效果,但对于处理庞大的图像数据、图像干扰严重等问题,其分类精度

无法满足实际需求,故传统分类器不适合复杂图像的分类。深度学习^[17]是机器学习的一种新兴算法,因其在图像特征学习方面具有显著效果而受到研究者的广泛关注。相较于传统的图像分类方法,其不需要对目标图像进行人工特征描述和提取,而是通过神经网络自主地从训练样本中学习特征,提取出更高维、抽象的特征,并且这些特征与分类器关系紧密,很好地解决了人工提取特征和分类器选择的难题,是一种端到端的模型。

2 图像分类数据库介绍

目前常用的图像分类数据库主要包括以下5个,且数据库在数据体量及复杂程度上依次递增。

(1) MNIST

MNIST^[18]是图像分类领域最经典的一个数据库,包含 70 000 张 28 dpi×28 dpi 的灰度图像,由数字(0~9)构成,共 10 个类别。训练集包含 60 000 个样本和 60 000 个标签,测试集包含 10 000 个样本和 10 000 个标签。

(2) fashion-MNIST

fashion-MNIST^[19]是一个类似 MNIST 数据库的时尚产品数据库。涵盖了来自 10 种类别的共 70 000 个不同商品的正面图片。fashion-MNIST 的大小、格式和训练集、测试集划分与 MNIST 完全一致,60 000 和 10 000 的训练、测试数据划分,28 dpi×28 dpi 的灰度图像。

(3) CIFAR-10

CIFAR-10^[20]数据集包含 60 000 张 32 dpi×32 dpi 的彩色图像,由飞机、马、狗等 10 个类别构成,10 类之间相互独立,无任何重叠的情况,训练集包含 50 000 个样本和 50 000 个标签,测试集包含 10 000 个样本和 10 000 个标签。

(4) CIFAR-100

CIFAR-100^[21]数据集同样包含 60 000 张 32 dpi×32 dpi 的彩色图像,共 100 个类别,每类有 600 幅图像,包括 500 幅训练图像和 100 幅测试图像。



不同于 CIFAR-10, 该数据集又将 100 个类划分为 20 个超类。

(5) ImageNet

ImageNet^[22] 是一个计算机视觉系统识别项目, 是目前世界上图像识别最大的数据库, 也是最常用的数据库, 包含 1 400 多万幅图像, 涵盖 2 万多个类别, 其中有超过 100 万幅图像有明确的类别标注和主要物体的定位边框, 图像分类、定位、检测等研究工作大部分基于此数据集展开。

3 基于深度学习的图像分类

基于深度学习的图像分类方法相比于传统的图像分类方法的关键优势在于, 其能通过深层架构自动学习更多抽象层次的数据特征, 无需针对特定的图像数据或分类方式设计具体的人工特征, 显著地提升了图像分类的效果。深度学习按照学习模型可分为生成模型、判别模型和混合模型^[23]。生成模型主要包括: 自编码器^[24]、深度信念网络^[25]和深度玻尔兹曼机^[26]等。判别模型主要包括深度前馈网络^[27]、卷积神经网络^[18]等, 混合模型由生成模型和判别模型两部分组成。1957 年, 第一代神经网络单层感知器由 Resenblatt^[28]提出, 可区分三角形、正方形等基本形状, 但无法解决异或问题。1986 年, Rumelhart 等^[24]则提出将原始单一的特征提取层扩展为多个隐藏层, 第二代神经网络诞生。神经网络由多个神经元按照一定的层次结构连接组成, 如图 1 所示, 神经网络包含输入层、隐藏层和输出层, 而隐藏层大于两层的神经网络被称为深度神经网络。图 1 中, 神经元被分成了多层, 层与层之间的神经元有连接, 层内的神经元之间无连接, 且每条连接线表示神经元之间的连接权重, 而神经网络的学习过程, 就是利用 BP (back propagation, 反向传播) 算法^[29]和梯度下降算法^[30]来最小化损失函数, 从而调整连接权重及每个神经元的偏置。

神经网络的前向传播计算式可表示如下:

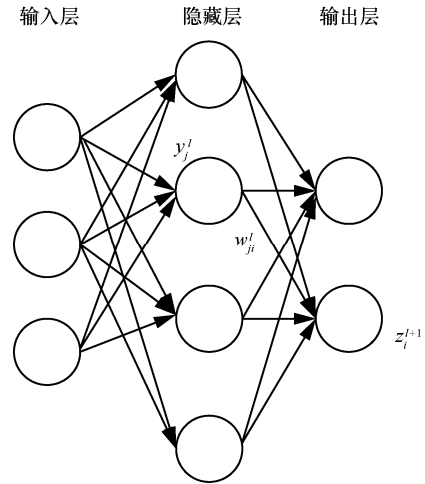


图 1 神经网络的前向传播

$$\begin{aligned} z_i^{l+1} &= \sum_j W_{ji}^l y_j^l + b_i^l \\ y_i^{l+1} &= f(z_i^{l+1}) \end{aligned} \quad (1)$$

其中, y_j^l 是第 l 层第 j 个神经元的输出, W_{ji}^l 是第 l 层第 j 个神经元与第 $l+1$ 层第 i 个神经元的连接权重, b_i^l 是偏置, $f(\cdot)$ 是非线性激活函数, 常用的激活函数有 Sigmoid、Tanh、ReLU、PReLU 等。神经网络的连接权重和偏置, 通过最小化损失函数学习获得, 若采用均方误差, 则损失函数可表示如下:

$$J(W, b) = \sum (L(x, y)) = \frac{1}{2} \sum_{i=1}^n (y_i - y_i^L)^2 \quad (2)$$

其中, y_i 是第 i 个神经元的真实值, y_i^L 是输出层第 i 个神经元的预测值, L 表示网络层数的最后一层 (即输出层), $l \leq L$ 。而在图像分类任务中更常用的是交叉熵, 则损失函数可表示如下:

$$\begin{aligned} J(W, b) &= -\frac{1}{n} \sum_{i=1}^n \left(y_i \ln(y_i^L) + (1 - y_i) \ln(1 - y_i^L) \right) \end{aligned} \quad (3)$$

3.1 自编码器

自编码器 (auto-encoders, AE) 由 Rumelhart 等^[24]提出, 是一种基于无监督学习的生成模型, 借助稀疏编码^[31]的思想, 使用稀疏的一些高阶特

征重新组合来重构输入数据,可以有效编码输入数据,主要用于数据降维或特征提取。自编码器包括编码阶段和解码阶段,其结构对称,输入、输出向量维度相同。自编码器的网络结构如图 2 所示,编码和解码过程可表示如下:

编码过程:

$$h^1 = f_1(W^1x + b^1) \quad (4)$$

解码过程:

$$y = f_2(W^2h^1 + b^2) \quad (5)$$

其中, W^1 、 b^1 为编码权重和偏置, W^2 、 b^2 为解码权重和偏置, f_1 、 f_2 为非线性变换, y 为重构输入, 损失函数常用均方误差, h^1 表示编码层输出, x 表示原始输入。

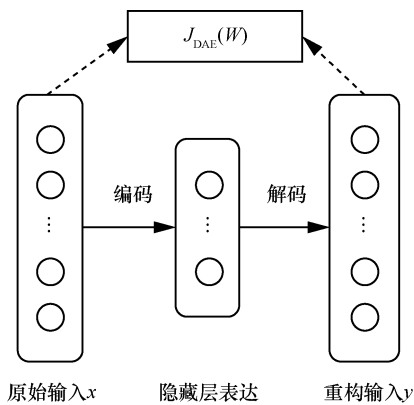


图2 自编码器的网络结构

若自编码器的隐藏层只有一层,网络中连接都为线性连接时,那么其原理类似于 PCA (principal component analysis, 主成分分析)^[32], 但单个隐藏层通常不能获得原始数据的代表性特征, Hinton 等^[17]提出了深度自编码器,并在参考文献[33-34]中进行了广泛的研究。同时,自编码器的隐藏层数不能太多,常将单个自编码器逐个训练,再堆叠多个自编码器的编码层,以完成深度学习的训练过程。单独的自编码器并不能对数据进行分类,它仅仅是去重构输入数据的误差,通常在图像分类任务中会在自编码器的顶层添加一个分类器(如 Softmax^[35]等),通过反向传播算

法对网络进行有监督的训练,最终获得分类结果。Wang 等^[36]利用自编码器的降维特性提取图像特征,将无标签的 MNIST 图片数据输入到自编码器中,采用无监督学习方法对输入图片的特征进行学习;然后,通过编码器生成的特征,对网络进行逐层训练;最后,使用带标签的图片数据进行监督训练微调网络权重。参考文献[36]中将 MNIST 图片维度从 784 减小到 s , 记录维度 s 变化时 Softmax 分类器性能的变化,探究隐藏层节点数量对自编码器学习性能的影响。实验结果显示,当隐藏层节点数量围绕数据的本质维度设置时,在 MNIST 数据集上可达到 93% 以上的分类精度,但在图像结构较为复杂时(如人脸图像),此参数设置方式并不适用。传统自编码器的编码、解码器能力过强时,仅实现了对训练样本的记忆,却难以发现数据的内在规律,针对传统自编码器出现的问题,接下来介绍以下典型的改进自编码器。

(1) 去噪自编码器

去噪自编码器(denoising auto-encoders, DAE)由 Vincent 等^[37]提出,从稳健性着手,针对隐藏层表达增加一定的约束。DAE 的原始输入数据被人为地添加一些噪声,使隐藏层表达与输入不同,从而迫使模型学习原始输入数据的分布结构,打破传统自编码器难以发现数据内在规律的限制,提高了重构数据的稳健性,在 MNIST 数据集上表现出优于传统自动编码器的分类性能。去噪自编码器网络结构如图 3 所示,原始输入 x , 加噪声后输入 $\tilde{x}$, 重构输入 y , 加噪声输入通过对原始输入随机映射得到 $\tilde{x} \sim q_D(\tilde{x}|x)$, 损失函数可表示如下:

$$J_{\text{DAE}}(W, b) = \sum_{x \sim q_D(\tilde{x}|x)} E_{\tilde{x}} [L(x, y)] \quad (6)$$

(2) 堆叠去噪自编码器

堆叠去噪自编码器(stacked denoising auto-encoders, SDAE)是 Vincent 等^[38]受深度信念

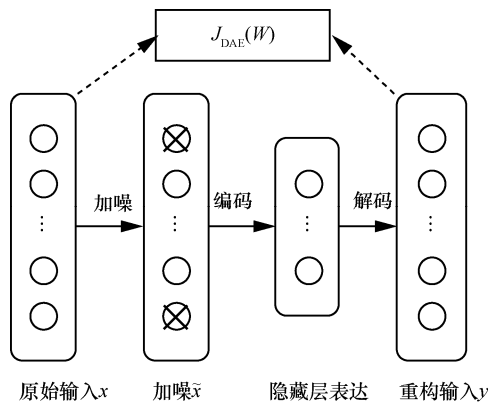


图3 去噪自编码器网络结构图

网络的启发，将去噪自编码器进行堆叠构造的，获得了更多输入数据的深层特征，并在 MNIST 数据集上取得了 1.28% 的分类错误率，相较于传统自编码器降低了 0.5%。虽然去噪自编码器对输入数据中的噪声表现出一定的稳健性，但其人为添加噪声的阶段增加了模型的处理时间。堆叠去噪自编码器的网络结构如图 4 所示。

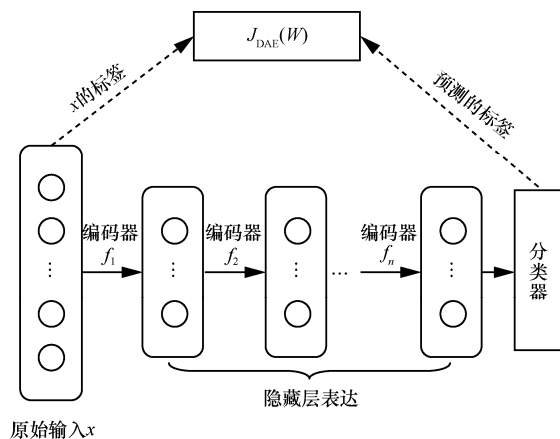


图4 堆叠去噪自编码器网络结构

(3) 其他

与去噪自编码器出发点相同，稀疏自编码器^[39]对隐藏层节点进行稀疏性限制，可获得高维而稀疏的特征表达，其在保证模型重构精度的基础上，极大地降低了数据维度，但无法获得一个准确的稀疏度。基于无监督学习的图像特征表示常导致模型需要学习大量的特征，致使参数量大幅上升。Li 等^[40]提出了一种快速有效的策略，先使用稀疏

自编码器提取图像的初始特征，再利用多重归一化差分方法对特征进行扩展，简化了模型复杂度，并在 UCM 数据集上的准确率达到 91.29%，但模型无法有效学习图像信息的空间关系。对于图像分类任务而言，模型对输入数据在一定程度上扰动具有不变性非常重要。Rifai 等^[41]提出了一种收缩自编码器，来抑制训练样本在所有方向上的扰动，提高隐藏层表达的稳健性。合成孔径雷达图像^[42]的分类，因其存在斑点噪声且缺乏有效的特征表示，是一项重要且具有挑战性的任务。

Geng 等^[42]利用灰度共生矩阵和 Gabor 滤波器去提取雷达图像的初始特征，然后将初始特征送入稀疏自编码器进行学习。接着，Geng 等^[43]在此基础上提出使用灰度梯度共生矩阵，通过 Gabor 和 HOG 滤波器提取图像的初步特征，再利用收缩自编码器对初步提取的特征进行学习，相比前者^[42]，该方法在相同雷达图像上的准确率最大提升了约 9%。同时，收缩自编码器对隐藏层表达的稳健性提升了模型整体性能。

3.2 概率生成模型

深度信念网络（deep belief networks, DBN）和深度玻尔兹曼机（deep Boltzmann machines, DBM）以受限玻尔兹曼机（restricted Boltzmann machines, RBM）为学习模块。RBM 由 Smolensky^[44]提出，是一种具有双层结构的无向图模型，也称为随机神经网络。RBM 具有一个可见层和一个隐藏层，且网络层间全连接，层内无连接。RBM 网络结构如图 5 所示， v 为可见层（输入层）， h 为隐藏层（输出层）， W 为两层间的连接权重。Welling 等^[45]证明，RBM 中的隐藏单元和可见单元的分布可以是任意的指数族分布，如高斯分布、泊松分布等。RBM 是基于能量的模型，以伯努利—伯努利 RBM 为例，假设其含有 m 个可见单元和 n 个隐藏单元，对任意的 i, j ， $v_i \in \{0, 1\}$ ， $h_j \in \{0, 1\}$ ， v_i 表示第 i 个可见单元的状态， h_j 表示第 j 个隐藏单元的状态。

态, W_{ij} 表示 v_i 和 h_j 之间的连接权重, v_i 和 h_j 分别表示第 i 个可见节点和第 j 个隐藏节点的偏置, 可得能量函数:

$$E(v, h | \theta) = - \sum_{i=1}^m b_i v_i - \sum_{j=1}^n c_j h_j - \sum_{i=1}^m \sum_{j=1}^n v_i w_{ij} h_j \quad (7)$$

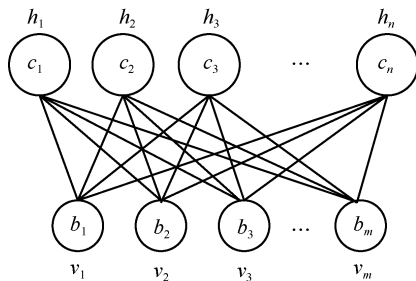


图5 RBM 网络结构

基于该能量函数可得 (v, h) 的联合概率分布:

$$P(v, h | \theta) = \frac{e^{-E(v, h | \theta)}}{Z(\theta)}, Z(\theta) = \sum_{v, h} e^{-E(v, h | \theta)} \quad (8)$$

因此, 可得似然函数:

$$P(v | \theta) = \frac{1}{Z(\theta)} \sum_h e^{-E(v, h | \theta)} \quad (9)$$

其概率越大, 表明训练得到的网络模型对观测数据的还原效果越好, 其中, $Z(\theta)$ 为归一化因子。由于 $Z(\theta)$ 的存在, (v, h) 的联合概率分布难以获取, 只能利用一些采样方法, 如 Gibbs 采样^[46] 获取其近似值, 但通常要使用较大的采样步数, 使得 RBM 的训练不高。为此, Hinton 等^[46] 提出了一种 RBM 的快速学习算法, 即 CD (contrastive divergence, 对比散度) 算法, 指出只需要 $k=1$ 步采样, 就可以获得很好的近似结果。Roux 等^[47] 证明了如果 RBM 的隐藏单元足够多, 则 RBM 可以拟合任意离散分布。利用 RBM 作为学习模块可以组成深度信念网络、深度玻尔兹曼机等深层模型。

(1) 深度信念网络

由 Hinton 等^[25] 提出, 它由若干 RBM 堆叠而成, 顶部两层是无向图结构的 RBM, 其余层自上

而下构成有向图的概率生成模型, 结构如图 6 (a) 所示。DBN 的输入层是底层 RBM 的可见层, 顶层 RBM 的隐藏层可与逻辑回归等分类器相连, 进行图像分类识别。DBN 模型通过训练连接权重, 网络可获得最大概率生成的训练数据。假设 DBN 有 l 层隐藏层, 建立的可视单元 v 和隐藏单元 h 之间的联合概率分布可表示如下:

$$P(v, h^1, \dots, h^l) = \left(\prod_{k=0}^{l-2} P(h^k | h^{k+1}) \right) P(h^{l-1}, h^l) \quad (10)$$

其中, $v = h^0$, 即输入的观测数据; $P(h^k | h^{k+1})$ 是第 k 层的隐藏单元在 $k+1$ 层隐藏单元上的条件概率分布; $P(h^{l-1}, h^l)$ 是顶层 RBM 的联合概率分布。

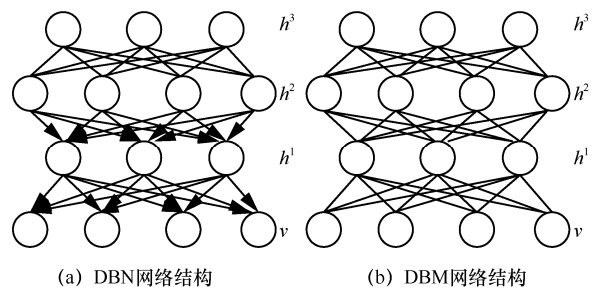


图6 DBN 和 DBM 网络结构

徐丽坤等^[48] 采用 DBN 算法进行高分辨率遥感图像的物分类, 网络输入是一个 784 维的特征向量, 训练分为无监督预训练和有监督微调两个阶段。预训练阶段通过无监督的贪心算法^[49] 单独训练每层 RBM, 当前训练好的 RBM 隐藏层作为下一层的 RBM 的可见层继续训练, 依次对每一层进行参数调整, 将网络的权重初始化到当前层特征向量映射最优, 直至所有层都训练完成。微调阶段在 DBN 的顶层连接 Softmax 分类层, 采用带标签的图像数据对网络进行微调, 弥补贪心算法只针对当前训练层最优的缺点, 得到全局最优, 最终输出分类结果。DBN 可以根据标签反向生成样本数据, 预训练和微调整阶段的训练时间、算法效率与传统神经网络相比都有较大改进, 更好地拟合了样本的内在结构, 遥感图像分类精度达



到 92%左右,但由于 DBN 预训练时层间单向接收数据信息,限制了模型学习数据间依赖关系的能力。

(2) 深度玻尔兹曼机

由 Salakhutdinov 等^[26]提出,不同于 DBN 顶部两层是无向图模型,下层是有向图模型,DBM 的结构均是无向连接。如图 6(b)所示,每一层隐藏层节点同时接收来自相邻两侧隐藏层的节点信息,使模型在相同数量隐藏层节点条件下,学习到更高维、抽象的特征。DBM 具有多层隐藏单元,其中,奇数层中的单元有条件地独立于偶数层,反之亦然。网络预训练时,联合训练无监督模型的所有层,而不是直接最大化似然函数,采用随机最大似然算法(SML)^[50]最大化可能性下限,但容易陷入局部最优^[51]。因此,在预训练 DBM 网络时,也会采用无监督逐层训练的贪心算法,这与 DBN 相同^[52]。同时,DBM 在训练时引入 Mean Field 方法^[26],使模型收敛更快。故在 DBM 模型中,各层节点条件概率可表示如下:

$$\begin{aligned} P(v_i = 1 | h^1) &= f \left(\sum_j W_{ij}^1 h_j^1 + a_i \right) \\ P(h_j^1 = 1 | v, h^2) &= f \left(\sum_i W_{ij}^1 v_i + \sum_q W_{jq}^2 h_q^2 + b_j \right) \\ P(h_q^2 = 1 | h^1) &= f \left(\sum_j W_{jq}^2 h_j^1 + c_q \right) \end{aligned} \quad (11)$$

其中, W^1 、 W^2 分别是可见层与隐藏层 h^1 、隐藏层 h^1 与 h^2 间的连接权重, $\{a_i\}$ 、 $\{b_j\}$ 、 $\{c_q\}$ 分别为可见层、第一层、第二层的偏置。杨建功等^[53]提出了一种基于 DBM 模型的融合谱—空域信息的高光谱图像分类方法。网络输入为每个像元谱—空域综合信息,对高光谱图像数据进行主成分分析法白化处理,并提取像元的空域信息,与像元的光谱信息组合为谱—空域信息。然后,将像元的谱—空域信息输入 DBM 模型提取出更具判别能力的深层次类别特征。最后,利用逻辑回归模型

对所提取特征进行分类,分别在 Indian Pines 和 Pavia University 两个数据集上获得 96.25%和 96.81%的分类精度。DBM 可以学习获得输入数据的多层复杂表示,但学习的时间复杂度远高于 DBN,这使得 DBM 参数的联合优化对于大型数据集来说不切实际。为提高 DBM 的效率,已经提出了几种改善方法,包括利用单独的模型来初始化所有层中的隐藏单元的值^[54-55],在预训练阶段^[56-57]或在训练阶段^[58-59]的其他改进,从而有效地加速学习过程。

(3) 其他

由于 DBN 不考虑输入图像的二维结构,影响了其在图像分类任务上的精度。为此, Lee 等^[60]提出了一种卷积深度信念网络(convolutional deep belief networks, CDBN),通过引入卷积 RBM 来利用相邻像素的空间信息,生成平移不变的生成模型,该模型在参考文献[61]中进一步扩展,在人脸识别中表现出优异的性能。针对 DBN 底层结构层间单向传递信息,难以提取深层特征的缺点,耿志强等^[62]提出了一种基于胶质细胞链的改进 DBN 模型及其学习算法,其中胶质细胞组成的链式结构与 DBN 的隐含层相连,调整隐藏层单元的输出并向其他胶质细胞传递相关信息,可提取更多数据特征,其在 MNIST 数据库上的表现相比于传统 DBN 模型,分类错误率下降了 1.06%,但模型引入胶质细胞机制增加了参数量,增大了模型训练时寻找最优参数的难度。针对深度学习方法严重依赖大规模带标签训练数据的问题, Goodfellow 等^[63]提出了一种更好的概率生成模型,即生成对抗网络(generative adversarial network, GAN)。GAN 避免了马尔科夫链式的学习机制,直接进行采样和推断,提高了 GAN 的应用效率,但由于无需预先建模,模型过于自由不可控。针对 GAN 本身不可控的缺点,唐贤伦等^[64]提出了一种条件深度卷积生成对抗网络(conditional-DCGAN, C-DCGAN),其结合深度

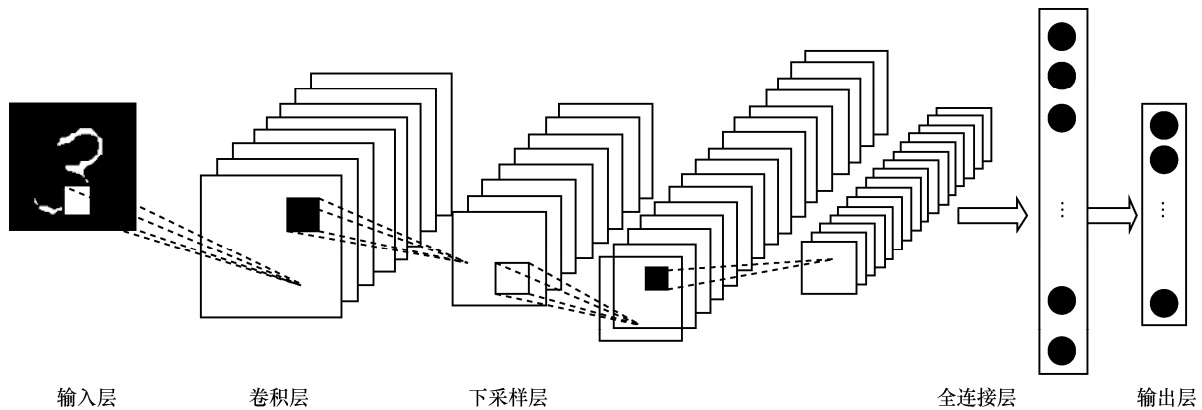


图7 CNN 基本结构

卷积生成对抗网络和条件生成对抗网络的优点，分别在 MNIST 和 CIFAR-10 数据集上的分类准确率达到 99.45% 和 84%，但生成模型与判别模型在对抗训练过程中速度较慢。

3.3 卷积神经网络

卷积神经网络(convolutional neural networks, CNN)的概念最早出自 19 世纪科学家提出的“感受野”^[65]，其是深度前馈网络的一种，也是当前图像分类领域的研究热点。卷积神经网络是一种监督学习的判别模型，其具有局部连接、权值共享、下采样的特点，善于挖掘数据局部特征，对图像的平移、缩放、旋转等表现出较高的稳健性^[66]。它以原始数据作为输入，通过卷积、池化和非线性激活函数映射等一系列操作，将原始数据逐层抽象为目标任务的特征表示。CNN 基本结构如图 7 所示，主要由输入层、卷积层、池化层、全连接层和输出层组成，图像分类任务中输出层为分类器，常用的分类器有 Softmax、SVM 等。

卷积层由多个滤波器组成，学习输入图像的特征表示，不同大小的卷积核可提取出不同的特征信息，低层卷积层可学习边缘和曲线等底层特征，而高层卷积层可学习更多的抽象特征，权值共享的方式降低了模型复杂度，减少了过拟合^[67]的风险，提高了模型的泛化性能^[68]。卷积过程如图 8 所示。

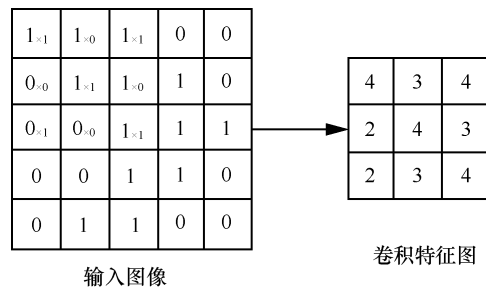


图8 卷积操作示意

池化层一般连接在连续卷积层之后，对特征进行降维，在一定程度上保持特征的尺度不变性、平移不变性和旋转不变性^[69]，常用方法包括最大池化与平均池化。池化过程如图 9 所示。在多次卷积和池化操作之后，网络连接若干全连接层，获得全局语义信息，在分类任务中对提取的特征进行分类，得到基于输入图像的概率分布，最后在输出层连接分类器，输出输入图像属于某一类别的概率。卷积神经网络通过残差的反向传播对网络中的参数进行训练来最小化损失函数，图像分类任务常用的损失函数是交叉熵函数。卷积神经网络训练时的过拟合、梯度消失^[70]和梯度爆炸^[71]问题会严重影响网络的收敛性能。为此，已提出一些有效的改善方法，包括：利用随机失活(dropout)^[72]技术增加网络的稀疏性和随机性，减轻过拟合问题；在网络训练过程中采用批量归一化(batch normalization, BN)^[73]技术，每一次梯度下降后权重因子都会得到改变，从而保证下一



层网络得到的数据拥有合适的分布；除此之外，利用经过预训练的网络进行参数初始化来加速学习过程，可以增强网络的泛化能力。

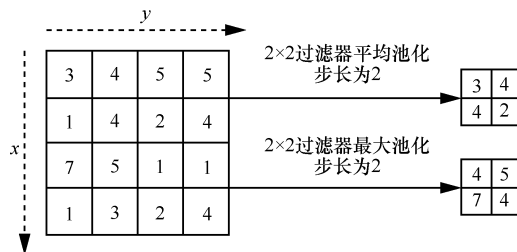


图9 池化操作示意

LeCun 等^[18]提出了 LeNet-5 网络，网络共 7 层，输入 MNIST 图片尺寸归一化为 32 像素×32 像素，经两次 5×5 卷积和 2×2 池化后，得到 5 像素×5 像素的图片，再依次连接含 120、84、10 个神经元的全连接层，最后采用 Sigmoid 函数进行激活，基于梯度的反向传播算法对卷积神经网络进行有监督的训练，输出数字 0~9 的分类概率。最终，LeNet-5 网络在手写数字识别任务中取得 0.8% 的分类错误率，获得了非常好的效果，证实了卷积神经网络在图像分类中的优越性。LeNet-5 网络虽然在手写字符识别任务上取得了成功，但存在训练数据集规模小、泛化能力弱、训练开销大的缺点。为此，Krizhevsky 等^[74]提出了具有 5 层卷积层的 AlexNet 网络，使用了 ReLU 激活函数，解决了 Sigmoid 函数在网络较深时的梯度弥散问题，同时加快了梯度下降的速度；引入 Dropout 技术，减轻网络过拟合，降低训练模型计算量；在大型图像数据库 ImageNet 的图像分类竞赛中首次实现了 15.4% 的 Top-5 错误率并夺得冠军，促使卷积神经网络的研究成为学术界的焦点。在 AlexNet 网络之后，将卷积神经网络的发展分为两类，一类是网络深度的增加，另一类是网络结构的改进。

(1) VGG 网络

Simonyan 等^[75]在 AlexNet 网络的基础上，探索卷积神经网络的深度与性能之间的关系，提出

了 VGG 网络。VGG 通过反复堆叠 3×3 的小型卷积核和 2×2 的最大池化层，实现了 16~19 层深的卷积神经网络，采用多尺度训练策略增加了数据量，证明了神经网络越深，效果越好。VGG 网络的参数虽然比 AlexNet 网络多，但其更深的网络和更小的卷积核带来的隐式的正则化效果使其只需要较少的迭代次数就可以收敛。VGG 网络在图像分类和物体定位任务上都取得了很好的效果，且 VGG 网络的泛化性很好，但网络深度的增加会带来训练误差增大的网络退化问题，因此 VGG 网络的最佳深度受限于 16~19 层。

(2) ResNet 网络

He 等^[76]提出了 ResNet 网络，即 ResNet V1，解决了深层网络训练的退化问题。ResNet 通过 shortcut 结构，将低层的特征 X 直接映射到高层的网络中，即增加了一个恒等映射。假设某段神经网络的输入是 X ，期望输出是 $H(X)$ ，shortcut 将原始学习目标 $H(X)$ 转换成 $H(X)-X$ （即残差），使整个网络只需要学习输入、输出差别的一部分，简化了网络学习目标和难度，残差模块的基本结构如图 10 所示。同时，ResNet 的结构可以极快地加速超深神经网络的训练，模型的准确率也有非常大的提升。参考文献[76]通过使用残差学习模块成功训练了 152 层深的 ResNet 网络，获得 ILSVRC 2015 比赛的冠军，取得 3.57% 的 Top-5 错误率，同时参数量却比 VGG 网络低。He 等^[77]在 ResNet V1 的基础上提出了 ResNet V2 模型，区别于 ResNet V1，该模型 shortcut 的非线性激活函数替换为恒等映射 $y=x$ 。同时，ResNet V2 每一层都使用了 BN 技术，进一步提升了网络性能。ResNet 网络的出现，使得构建超深层网络成为现实，目前可达到 1 000 层以上，对卷积神经网络的后续发展产生了深远的意义。

(3) GoogLeNet 网络

Szegedy 等^[78]从优化卷积神经网络结构，降低网络复杂度的方向，提出了 GoogLeNet 网络，

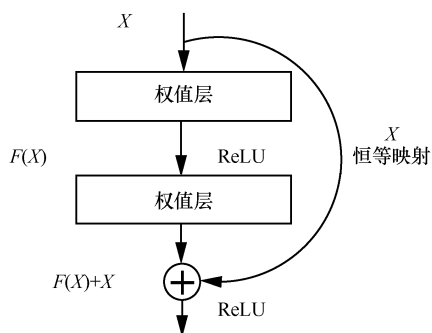


图 10 残差模块基本结构

即 Inception V1, 网络有 22 层深, 由若干 Inception 模块级联而成, 采用全局平均池化层替代最后的全连接层, 增加辅助分类节点, 最终以 6.67% 的 Top-5 错误率获得 ILSVRC 2015 比赛分类任务的冠军。Inception 模块的基本结构如图 11 所示, 其中有 4 个分支, 包含 1×1 、 3×3 、 5×5 的 3 种不同尺寸卷积核和 1 个 3×3 的最大池化, 与多尺度的思想类似, 增加了网络对不同尺度的适应性。同时, 1×1 的卷积可以用很小的计算量增加一层特征变换和非线性化, 它跨通道组织信息的功能提高了网络的表达能力, 同时可以对输出通道升维和降维。参考文献[78]指出 Inception 模块可以让卷积神经网络的深度和宽度都高效率地扩充, 提升模型准确率且不至于过拟合。在 Inception V1 的基础上, Ioffe 等^[73]提出了 Inception V2 网络, 该网络借鉴了 VGG 网络, 采用两个 3×3 的卷积替代 5×5 的卷积, 用以降低参数量并减轻过拟合, 训练时使用 BN 方法加速网络训练, 最终在 ImageNet 上获得 4.8% 的 Top-5 错误率。2015 年, Szegedy 等^[79]提出了 Inception V3 网络, 该网络引入了分解卷积的思想, 将一个较大的二维卷积拆分成两个较小的一维卷积, 如将 7×7 的卷积拆分成 1×7 和 7×1 的卷积。同时 Inception V3 优化了 Inception 模块, 在分支中嵌入分支, 最终在 ImageNet 上获得 3.5% 的 Top-5 错误率。

(4) 其他

在 ResNet 网络的基础上, Huang 等^[80]提出了一种 DenseNet 模型, 该模型以前馈的方式将网络的每一层和前面的所有层相连, 同时把每一层设

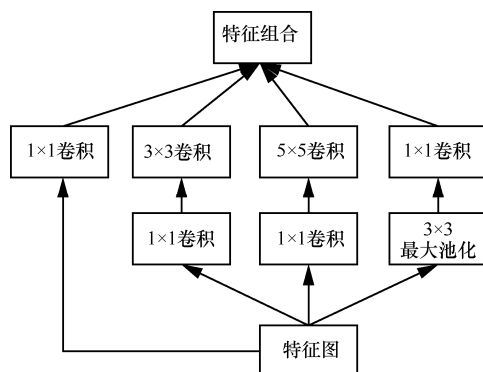


图 11 Inception 模块基本结构

计得特别窄, 学习很少的特征图以此降低冗余性, 在 ImageNet 上获得与 ResNet 相当的准确性, 但所需参数明显更少。虽然网络深度的加深使得网络性能得到显著提高, 但计算效率成了超深层网络不可忽视的问题。为此, Iandola 等^[81]提出了一种轻量化模型 SqueezeNet, 在 ImageNet 上实现了 AlexNet 级精度, 且参数减少到 $1/50$ 。2017 年, Howard 等^[82]提出了一种 MobileNet 的高效模型, 用于移动和嵌入式视觉应用。该模型使用深度可分离卷积来构建轻量级深度神经网络, 与 GoogLeNet 参数相当, 且在运算量上小于 GoogLeNet 一个量级, 与其他流行模型相比, 在 ImageNet 上显示出了强大的性能。Attention 机制^[83]能够让深度学习模型集中关注输入数据中最为重要的一部分, Wang 等^[84]提出了一种基于注意力机制的卷积神经网络, 即残差注意力网络, 由多个注意力模块堆叠而成, 随着层越来越深入, 来自不同模块的注意力感知特征会自适应地改变。同时, 参考文献[84]提出的注意力残差训练方法使得网络能够扩展到数百层, 最终在 ImageNet 上获得 4.8% 的 Top-5 错误率。Hu 等^[85]提出了 SENet (squeeze-and-excitation network) 模型, 该模型根据建模特征通道之间的相互依赖关系的方向, 采用了一种全新的特征重标定策略, 即通过学习的方式来自动获取每个特征通道的重要程度, 然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征, 最终以 2.25% 的 Top-5 错误率获得 ILSVRC



2017 分类任务冠军。基于 CNN 在探索特征之间相对位置关系和特征方向关系上的局限性，Sabour 等^[86]提出了 CapsNets (capsule network, 胶囊网络)，网络由胶囊构建，是目前深度学习方法的最新突破之一，也是图像分类领域最前沿的技术之一，其在 MNIST 数据库上可以达到当前 CNN 最高性能表现，在识别高度重叠的数字时表现出优于卷积神经网络的性能。

4 算法对比

本文分析了近年来基于深度学习的图像分类方法的发展现状，图像分类效果的提升主要得益于深度学习算法的不断改进，其发展历程如图 12 所示，其中箭头指向即深度学习算法改进的时间推进方向，大括号包含内容即某一算法原型不同改进方向的算法。以下将从参考文献中提取部分代表性深度学习算法，以其在 MNIST 和 ImageNet 数据库上的实验结果作为参考，进行对比分析。同时，

从 CNN 模型的时间复杂度和空间复杂度的角度，对比分析参考文献中具有代表性的 CNN 模型。

4.1 分类精度

深度学习算法在 MNIST、ImageNet 数据库上的性能表现分别见表 1、表 2，对于有多个实验结果的方法，只选取了最优结果。通过观察表 1 可以发现，MLP 算法相较于传统的 kNN 算法，可以学习到图像更多的特征信息，在图像分类任务上有明显的优势。AE 算法则使用稀疏的高阶特征重新组合来重构输入数据，有效地编码图像数据，数据降维的同时提升了特征的提取效果，DBN 算法在计算效率上跟传统神经网络相比有较大的改进。然而，卷积神经网络充分利用了图像二维空间邻域信息，使其在图像分类上性能遥遥领先于前述深度学习算法。表 2 中 Top-1 错误率代表所学习到的标签中预测概率最大的那一类不是正确类别的比率，Top-5 错误率则代表所学习到的标签中预测概率最大的 5 个类别中不包含正确类别

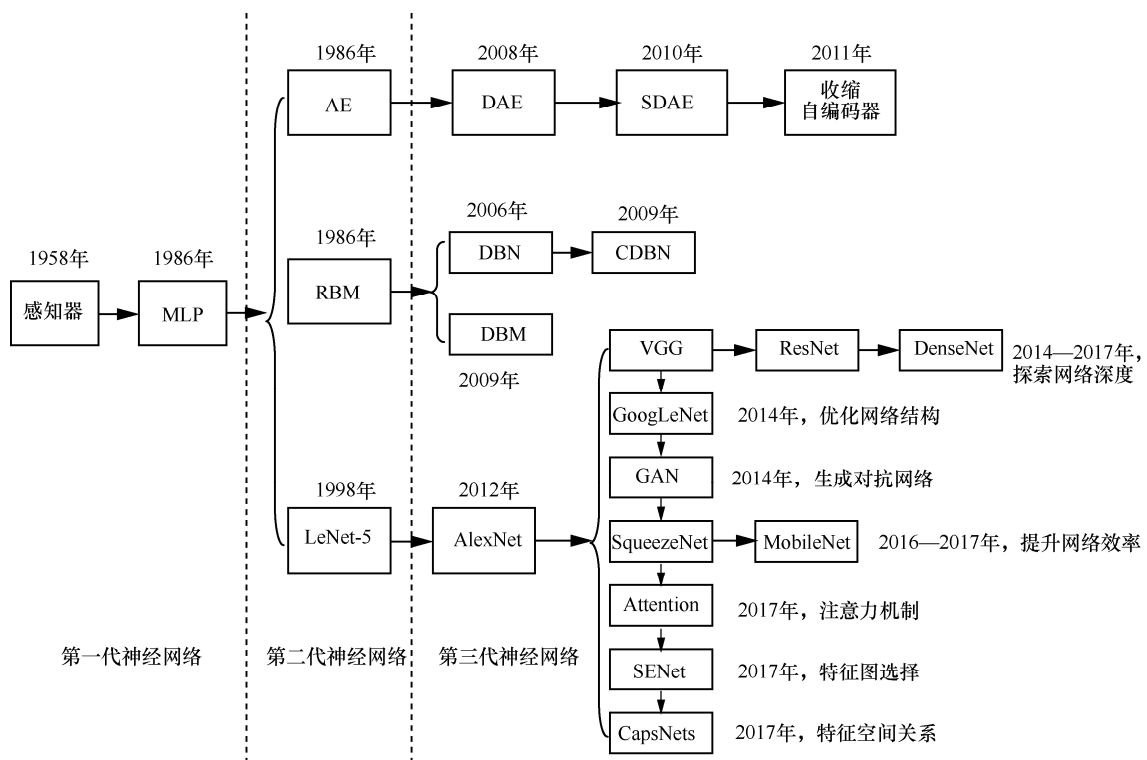


图 12 深度学习发展史

的比率,第4列测试集错误率,均为当年 ILSVRC 大赛上提交的集成网络测试集错误率,可见卷积神经网络在图像分类领域的发展尤为迅速,且尚有突破空间,故后续卷积神经网络的深入研究对图像分类领域非常重要。

表1 基于深度学习图像分类模型的 MNIST 数据集
错误率对比

深度学习模型	错误率
KNN ^[13]	5.0%
MLP ^[18]	2.45%
传统自编码器 ^[41]	1.78%
堆叠去噪自编码器 ^[38]	1.28%
DBN ^[25]	1.25%
LeNet-5 ^[18]	0.8%

表2 基于 CNN 图像分类模型的 ImageNet
数据集错误率对比

CNN 模型	Top-1 错误率 (val)	Top-5 错误率 (val)	Top-5 错误率 (test)
AlexNet ^[74]	36.7%	15.4%	15.3%
VGGNet ^[75]	—	8.43%	7.32%
GoogLeNet ^[78]	—	7.89%	6.66%
BN-Inception ^[73]	21.99%	5.82%	—
ResNet-152 ^[76]	19.38%	4.49%	3.57%
Inception-V3 ^[79]	18.77%	4.2%	—
DensNet-264 ^[80]	20.80%	5.29%	—
Attention-92 ^[84]	19.5%	4.8%	—
1.0 MobileNet-224 ^[82]	29.4%	—	—
SENet-154 ^[85]	18.68%	4.47%	2.25%

注: val 为 validation 的简写,表示验证集; test 表示测试集

4.2 FLOPS 和参数数量

时间复杂度决定了模型训练/预测需要运算的次数,以 FLOPS (floating-point operations per second, 每秒浮点运算次数) 衡量,空间复杂度决定了参数的数量,以模型的参数数量衡量。通过观察表2、表3可以发现,近几年 CNN 结构和深度的改变提升了分类性能,但其时间复杂度及

空间复杂度随网络深度的加深持续增大,引发了网络效率问题。模型的时间复杂度随着硬件设备的发展得到了一定改善,但空间复杂度的增大,一方面造成模型维度大幅增加且训练易陷入过拟合,另一方面会造成模型时间复杂度增加,故减少模型空间复杂度才能从整体上保证模型性能并提升模型效率。SqueezeNet、MobileNet 等高效模型的提出,给提升模型效率提供了一个新的方向,轻量化模型的研究逐渐展开。

表3 CNN 图像分类模型的时间复杂度和空间复杂度

CNN 模型	FLOPS (每秒浮点运算次数)	参数数量/个
LeNet-5 ^[18]	4×10^5	6×10^4
AlexNet ^[74]	7×10^9	6×10^7
VGGNet-16 ^[75]	1.5×10^{11}	1.38×10^8
GoogLeNet ^[78]	1.5×10^{10}	5×10^6
ResNet-50 ^[76]	3.86×10^9	2.5×10^7
Attention-92 ^[84]	1.04×10^{10}	5.13×10^7
MobileNet ^[82]	5.69×10^8	4.2×10^6
SENet-50 ^[85]	3.87×10^9	2.75×10^7

5 未来研究方向

尽管深度学习在图像分类领域取得了巨大的进步,但仍存在一些问题,本文针对数据集问题、模型效率问题和落地问题分别探讨了未来可能的研究方向。数据集问题指的是图像数据集不易获取的问题,模型效率问题指的是模型的存储问题和模型进行预测的速度问题,落地问题指的是基于深度学习的图像分类方法实际应用的问题。

5.1 数据集问题

(1) 基于半监督、无监督学习的图像分类

目前,深度学习在图像分类任务中大多数是基于监督学习的研究,而监督学习的数据集需要大量的标记数据,成本高且不易获得。目前基于半监督、无监督学习的研究相较于监督学习还有较大差距,因此还需继续深入研究基于半监督、



无监督学习的图像分类算法。

(2) 基于迁移学习的小样本图像分类

由于深度卷积神经网络的特征数量非常庞大,依赖于大规模的带标签样本数据集,而实际生活中的特定任务数据集规模往往比较小(如农作物病虫害数据集等),若直接对其建模训练,模型易过拟合。迁移学习旨在解决训练数据不足的问题,其将大型源数据域中学习的参数迁移到目标数据域,有效利用迁移学习对小样本目标数据域进行训练,可加快目标数据域的网络收敛并获得更好的泛化性能。

5.2 模型效率问题

移动平台部署。卷积神经网络在图像分类领域取得了明显优于其他深度学习算法的效果,但随着其网络层数的加深,计算成本也大幅上升,测试时需要占用大量的内存且极为耗时,使得它们不适合部署在资源有限的移动平台上。研究如何在保证网络性能的基础上提高模型效率、减少网络参数非常重要,而轻量化模型的研究才刚刚起步,如何将卷积神经网络更好地应用在图像分类领域中是未来的研究热点。

5.3 落地问题

自然场景下的图像分类。图像分类是一门与实际应用密切相关的研究课题,但目前学术研究中使用的数据库具有目标对象突出、背景单一的特点,而在实际生产工作中采集的图片通常还需要考虑遮挡、低分辨率及物体干扰等其他复杂场景的因素。因此,研究深度学习在自然场景下的图像分类有利于高效开展实际生活中的生产工作。

6 结束语

本文对基于深度学习的图像分类方法进行了介绍,首先回顾了传统图像分类方法及其存在的问题,主要对自动编码器、深度信念网络与深度玻尔兹曼机以及卷积神经网络4种最重要的深度学习方法在图像分类领域的发展现状进行了综述。

另外,对比分析了不同深度学习方法在图像分类任务上的性能表现,探讨了当前深度学习方法在图像分类领域的不足及未来可能的研究方向。

参考文献:

- [1] OUYANG W, ZENG X, WANG X, et al. DeepID-Net: object detection with deformable part based convolutional neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(7): 1320-1334.
- [2] DIBA A, SHARMA V, PAZANDEH A, et al. Weakly supervised cascaded convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: ACM Press, 2017: 5131-5139.
- [3] HU G, YANG Y X, YI D, et al. When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition[C]//International Conference on Computer Vision, December 11-18, 2015, Santiago, Chile. Piscataway: IEEE Press, 2015: 142-150.
- [4] LAWRENCE S, GILES C L, TSOI A C, et al. Face recognition: a convolutional neural-network approach[J]. IEEE Transactions on Neural Networks, 1997, 8(1): 98-113.
- [5] CAO Z, SIMON T, WEI S, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. EprintArxiv, 2017: 1302-1310.
- [6] TOSHEV A, SZEGEDY C. DeepPose: human pose estimation via deep neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: ACM Press, 2014: 1653-1660.
- [7] PERREAULT S, HEBERT P. Median filtering in constant time[J]. IEEE Transactions on Image Processing, 2007, 16(9): 2389-2394.
- [8] SLOT K, KOWALSKI J, NAPIERALSKI A, et al. Analogue median/average image filter based on cellular neural network paradigm[J]. Electronics Letters, 1999, 35(19): 1619-1620.
- [9] DIREKOGLU C, NIXON M S. Image-based multiscale shape description using Gaussian filter[C]//2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing, December 16-19, 2008, Bhubaneswar, India. Piscataway: IEEE Press, 2009: 673-678.
- [10] GRABNER M, GRABNER H, BISCHOF H. Fast approximated SIFT[C]//Asian Conference on Computer Vision, January 13-16, 2006, Hyderabad, India. Heidelberg: Springer, 2006: 918-927.
- [11] HE L, ZOU C, ZHAO L, et al. An enhanced LBP feature based on facial expression recognition[C]//IEEE Engineering in Med-

- icine and Biology 27th Annual Conference, September 1-4, 2005, Shanghai, China. Piscataway: IEEE Press, 2005: 3300-3303.
- [12] DENIZ O, BUENO G, SALIDO J, et al. Face recognition using histograms of oriented gradients[J]. Pattern Recognition Letters, 2011, 32(12): 1598-1603.
- [13] LECUN Y, JACKEL L, BOTTOU L, et al. Comparison of learning algorithms for handwritten digit recognition[C]//International Conference on Artificial Neural Networks, January, 1995, Nanterre, France. [S.l.:s.n.], 1995: 53-60.
- [14] BEUCHER A, MOLLER A B, GREVE M H. Artificial neural networks and decision tree classification for predicting soil drainage classes in Denmark[J]. Geoderma, 2017, 320: 30-42.
- [15] EBRAHIMI M A, KHOSHTAGHAZ M H, MINAEI S, et al. Vision-based pest detection based on SVM classification method[J]. Computers and Electronics In Agriculture, 2017, 137: 52-58.
- [16] 周建同, 杨海涛, 刘东, 等. 视频编码的技术基础及发展方向[J]. 电信科学, 2017, 33(8): 16-25.
ZHOU J T, YANG H T, LIU D, et al. Trends and technologies of video coding[J]. Telecommunications Science, 2017, 33(8): 16-25.
- [17] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504.
- [18] 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.
- [19] XIAO H, RASUL K, VOLLGRAF R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms[J]. Statistics, 2017(2).
- [20] LI H, LIU H, JI X, et al. CIFAR10-DVS: an event-stream dataset for object classification[J]. Frontiers in Neuroscience, 2017(11): 309.
- [21] MCCLURE P, KRIEGESKORTE N. Representational distance learning for deep neural networks[J]. Frontiers in Computational Neuroscience, 2016(10): 131.
- [22] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//The 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Washington, USA. Piscataway: IEEE Press, 2009: 248-255.
- [23] 郭丽丽, 丁世飞. 深度学习研究进展[J]. 计算机科学, 2015, 42(5): 28-33.
GUO L L, DING S F. Research progress on deep learning[J]. Computer Science, 2015, 42(5): 28-33.
- [24] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [25] HINTON G E, OSINDERO S, TEH Y. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [26] SALAKHUTDINOV R, HINTON G. Deep Boltzmann machines[C]//International Conference on Artificial Intelligence and Statistics, April 16-19, 2009, Florida, USA. [S.l.:s.n.], 2009: 448-455.
- [27] BABRI H A, TONG Y. Deep feedforward networks: application to pattern recognition[C]//International Conference on Neural Networks (ICNN'96), June 3-6, 1996, Washington, USA. Piscataway: IEEE Press, 1996: 1422-1426.
- [28] ROSENBLATT F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological Review, 1958, 65(6): 386-408.
- [29] 赵会敏, 雒江涛, 杨军超, 等. 集成 BP 神经网络预测模型的研究与应用[J]. 电信科学, 2016, 32(2): 60-67.
ZHAO H M, LUO J T, YANG J C, et al. Research and application of prediction model based on ensemble BP neural network[J]. Telecommunications Science, 2016, 32(2): 60-67.
- [30] 高雪鹏, 丛爽. BP 网络改进算法的性能对比研究[J]. 控制与决策, 2001(2): 167-171.
GAO X P, CONG S. Comparative study on fast learning algorithms of BP networks[J]. Control and Decision, 2001(2): 167-171.
- [31] OLSHAUSEN B A, FIELD D J. Sparse coding with an overcomplete basis set: A strategy employed by V1?[J]. Vision Research, 1997, 37(23): 3311-3325.
- [32] LIU Y, ZHAO S S, WANG Q Q, et al. Learning more distinctive representation by enhanced PCA network[J]. Neurocomputing, 2018(275): 924-931.
- [33] LIU T, LI Z R, YU C X, et al. NIRS feature extraction based on deep auto-encoder neural network[J]. Infrared Physics & Technology, 2017(87): 124-128.
- [34] HASSAIRI S, EJBALI R, ZAIED M. A deep stacked wavelet auto-encoders to supervised feature extraction to pattern classification[J]. Multimedia Tools and applications, 2018, 77(5): 5443-5459.
- [35] LIU Y, WU L Z. Geological disaster recognition on optical remote sensing images using deep learning[J]. Procedia Computer Science, 2016(91): 566-575.
- [36] WANG Y S, YAO H X, ZHAO S C. Auto-encoder based dimensionality reduction[J]. Neurocomputing, 2016, 184(S1): 232-242.
- [37] VINCENT P, LAROCHELLE H, BENGIO Y, et al. Extracting and composing robust features with denoising autoencoders[C]//the 25th International Conference on Machine Learning, July 5-9, 2008, Helsinki, Finland. New York: ACM Press, 2008: 1096-1103.



- [38] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion[V]. *Journal of Machine Learning Research*, 2010(11): 3371-3408.
- [39] PATHIRAGE C S N, LI J, LI L, et al. Development and application of a deep learning-based sparse autoencoder framework for structural damage identification[J]. *Structural Health Monitoring*, 2018, 18(1): 103-122.
- [40] LI E Z, DU P J, SAMAT A, et al. Mid-level feature representation via sparse autoencoder for remotely sensed scene classification[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(3): 1068-1081.
- [41] RIFAI S, VINCENT P, MULLER X, et al. Contractive auto-Encoders: explicit invariance during feature extraction [C]// *International Conference on Machine Learning*, June 28-July 2, 2011, Washington, USA. [S.l.:s.n.], 2011: 833-840.
- [42] GENG J, FAN J C, WANG H Y, et al. High-Resolution SAR image classification via deep convolutional autoencoders[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(11): 2351-2355.
- [43] GENG J, WANG H Y, FAN J C, et al. Deep supervised and contractive neural network for SAR image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(4): 2442-2459.
- [44] SMOLENSKY P. Information processing in dynamical systems: foundations of harmony theory[C]// *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, January 1-4, 1986, Cambridge, USA. Cambridge: MIT Press, 1986.
- [45] WELLING M, ROSEN-ZVI M, HINTON G. Exponential family harmoniums with an application to information retrieval[C]// *Advances in Neural Information Processing Systems 17*, December 13-16, 2004, Cambridge USA. Cambridge: MIT Press, 2005: 1481-1488.
- [46] HINTON G E. Training products of experts by minimizing contrastive divergence[J]. *Neural Computation*, 2002, 14(8): 1771-1800.
- [47] ROUX N L, BENGIO Y. Representational power of restricted Boltzmann machines and deep belief networks[J]. *Neural Computation*, 2008, 20(6): 1631-1649.
- [48] 徐丽坤, 刘晓东, 向小翠. 基于深度信念网络的遥感影像识别与分类[J]. *地质科技情报*, 2017, 36(4): 244-249.
- XU L K, LIU X D, XIANG X C. Recognition and classification for remote sensing image based on depth belief network[J]. *Geological Science and Technology Information*, 2017, 36(4): 244-249.
- [49] LIU Q, GAO Z Q, LIU B, et al. Automated rule selection for opinion target extraction[J]. *Knowledge-Based Systems*, 2016(104): 74-88.
- [50] YOUNES L. On the convergence of markovian stochastic algorithms with rapidly decreasing ergodicity rates[J]. *Stochastics and Stochastic Reports*, 1999, 65(3-4): 177-228.
- [51] ALJARAH I, FARIS H, MIRJALILI S. Optimizing connection weights in neural networks using the whale optimization algorithm[J]. *Soft Computing*, 2018, 22(1): 1-15.
- [52] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(8): 1798-1828.
- [53] 杨建功, 汪西莉, 刘侍刚. 融合谱-空域信息的 DBM 高光谱图像分类方法[J]. *西安电子科技大学学报*, 2019, 46(3): 109-115.
- YANG J G, WANG X L, LIU S G. Spectral-spatial classification of hyperspectral images using deep Boltzmann machines[J]. *Journal of Xidian University*, 2019, 46(3): 109-115.
- [54] SALAKHUTDINOV R, LAROCHELLE H. Efficient learning of deep Boltzmann machines[J]. *Journal of Machine Learning Research*, 2010(9): 693-700.
- [55] SALAKHUTDINOV R, HINTON G. An efficient learning procedure for deep Boltzmann machines[J]. *Neural Computation*, 2012, 24(8): 1967-2006.
- [56] SALAKHUTDINOV R, HINTON G. A better way to pretrain deep Boltzmann machines[C]// *The 26th Annual Conference on Neural Information Processing Systems*, December 3-6, 2012, Lake Tahoe, Nevada, USA. Red Hook: Curran Associates Inc, 2012: 2447-2455.
- [57] CHO K, RAIKO T, ILIN A, et al. A two-stage pretraining algorithm for deep Boltzmann machines[C]// *23rd International Conference on Artificial Neural Networks*, Sep 10-Oct 13, 2013, Techn Univ Sofia, Sofia, Bulgaria. Heidelberg: Springer, 2013: 106-113.
- [58] GOODFELLOW I, MIRZA M, COURVILLE A, et al. Multi-prediction deep Boltzmann machines[C]// *The 26th International Conference on Neural Information Processing Systems*, December 5-10, 2013, Lake Tahoe, Nevada, USA. Red Hook: Curran Associates Inc, 2013: 548-556.
- [59] BOURLARD H, KAMP Y. Auto-association by multilayer perceptrons and singular value decomposition[J]. *Biological Cybernetics*, 1988, 59(4): 291-294.
- [60] LEE H, GROSSE R, RANGANATH R, et al. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations[C]// *International Conference on Machine Learning*, June 14-18, 2009, Montreal, Canada. New York: ACM Press, 2009: 609-616.
- [61] HUANG G B, LEE H, LEARNED-MILLER E. Learning hier-

- archical representations for face verification with convolutional deep belief networks[C]//The IEEE Conference on Computer Vision and Pattern Recognition, Jun 16-21, 2012, Washington, USA. Piscataway: IEEE Press, 2012:2518-2525.
- [62] 耿志强, 张怡康. 一种基于胶质细胞链的改进深度信念网络模型[J]. 自动化学报, 2016, 42(6): 943-952.
- GENG Z Q, ZHANG Y K. An improved deep belief network inspired by glia chains[J]. *Acta Automatica Sinica*, 2016, 42(6): 943-952.
- [63] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Annual Conference on Neural Information Processing Systems, December 8-13, 2014, Cambridge, USA. Cambridge: MIT Press, 2014: 2672-2680.
- [64] 唐贤伦, 杜一铭, 刘雨微, 等. 基于条件深度卷积生成对抗网络的图像识别方法[J]. 自动化学报, 2018, 44(5): 855-864.
- TANG X L, DU Y M, LIU Y W, et al. Image recognition with conditional deep convolutional generative adversarial networks[J]. *Acta Automatica Sinica*, 2018, 44(5): 855-864.
- [65] SHERRINGTON C S. Observations on the scratch-reflex in the spinal dog[J]. *The Journal of Physiology*, 1906, 34(1-2):1-50.
- [66] AKHTAR S W, REHMAN S, AKHTAR M, et al. Improving the robustness of neural networks using k -support norm based adversarial training[J]. *IEEE Access*, 2016, 4: 9501-9511.
- [67] COOK J A, RANSTAM J. Overfitting[J]. *British Journal of Surgery*, 2016, 103(13): 1814.
- [68] ANTOL S, AGRAWAL A, LU J, et al. VQA: visual question answering[C]//The 2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. Piscataway: IEEE Press, 2015: 2425-2433.
- [69] TUYTELAARS T, MIKOLAJCZYK K. Local invariant feature detectors: a survey[J]. *Now Foundations and Trends*, 2007, 3(3): 177-280.
- [70] SQUARTINI S, PAOLINELLI S, PIAZZA F. Comparing different recurrent neural architectures on a specific task from vanishing gradient effect perspective[C]//2006 IEEE International Conference on Networking, Sensing and Control, April 23-25, 2006, FL, USA. Piscataway: IEEE Press, 2006:380-385.
- [71] PASCANU R, MIKOLOV T, BENGIO Y. Understanding the exploding gradient problem[J]. *Arxiv Preprint Arxiv*, 2012.
- [72] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. *Computer Science*, 2012, 3(4): 212-223.
- [73] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning, July 6-11, 2015, Lille, France. [S.l.: s.n.], 2015: 448-456.
- [74] KRIZHEVSKY A, SUTSKEVER I, E. HINTON G. ImageNet classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems, December 3-6, 2012, Lake Tahoe, Nevada. Red Hook: Curran Associates Inc, 2012: 1097-1105.
- [75] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference of Learning Representation, May 7-9, 2015, San Diego, CA. arXiv:1409.1556v6 [cs.CV], 2015.
- [76] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, Nevada. Los Alamitos: IEEE Computer Society, 2016: 770-778.
- [77] HE K M, ZHANG X Y, REN S Q, et al. Identity mappings in deep residual networks[C]//14th European Conference on Computer Vision, October 8-16, 2016, Amsterdam, Netherlands. Heidelberg: Springer, 2016: 630-645.
- [78] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. Piscataway: IEEE Press, 2015: 1-9.
- [79] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Seattle, WA, USA. Piscataway: IEEE Press, 2016: 2818-2826.
- [80] HUANG G, LIU Z, MAATEN L V D, et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. Piscataway: IEEE Press, 2017: 2261-2269.
- [81] IANDOLA F, HAN S, W. MOSKEWICZ M, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[C]//International Conference on Learning Representations, April 24-26, 2017, Toulon, France. arXiv:1602.07360v4 [cs.CV]. 2016.
- [82] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[J]. arXiv:1704.04861v1 [cs.CV]. 2017.
- [83] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//31st Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. [S.l.: s.n.], 2017.
- [84] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification[C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. Piscataway: IEEE Press, 2017: 6450-6458.
- [85] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, New York, USA. Piscataway:



IEEE Press, 2018: 7132-7141.

- [86] SABOUR S, FROSST N, E HINTON G. Dynamic routing between capsules[C]//31st Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. arXiv:1710.09829v2 [cs.CV]. 2017.

[作者简介]



苏赋（1973-），女，博士，西南石油大学副教授，主要研究方向为信号与信息处理。



吕沁（1995-），女，西南石油大学硕士生，主要研究方向为深度学习与图像处理。



罗仁泽（1973-），男，博士，西南石油大学教授、博士生导师，主要研究方向为信号处理与人工智能。