基于集成卷积神经网络的遥感影像场景分类

张晓男[1,2**]，钟兴[1,3*]，朱瑞飞[1,3]，高放[3]，张作省[1,2]，鲍松泽[1,2]，李竺强[3]

[1]中国科学院长春光学精密机械与物理研究所，吉林 长春 130033；

[2]中国科学院大学，北京 100049；

[3]长光卫星技术有限公司吉林省卫星遥感应用技术重点实验室，吉林 长春 130102

摘要 提出了一种基于集成卷积神经网络(CNN)的遥感影像场景分类算法。通过构建反向传播网络实现了场景图像的复杂度度量；根据图像的复杂度级别，选择 CNN 对图像进行分类，完成了遥感影像的场景分类。使用所提出的算法对 NWPU-RESISC45 公开数据集进行了实验验证，取得了 89.33%(第一类实验)和 92.53%(第二类实验)的分类准确率，平均运行时间为 0.41 s。相比于精调训练的 VGG-16 模型，所提算法的分类准确率分别提升了 2.19%和 2.17%，预测速率提升了 33%，证明了其有效性和实用性。

关键词 遥感；卷积神经网络；图像复杂度；场景分类

中图分类号 TP753 **文献标识码** A **doi**：10.3788/AOS201838.1128001

Scene Classification of Remote Sensing Images Based on Integrated Convolutional Neural Networks

Zhang Xiaonan[1,2**], Zhong Xing[1,3*], Zhu Ruifei[1,3], Gao Fang[3], Zhang Zuoxing[1,2], Bao Songze[1,2], Li Zhuqiang[3]

[1] *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun, Jilin* 130033, *China*;

[2] *University of Chinese Academy of Sciences, Beijing* 100049, *China*;

[3] *Key Laboratory of Satellite Remote Sensing Application Technology of Jilin Province, Chang Guang Satellite Technology Co., Ltd, Changchun, Jilin* 130102, *China*

Abstract A scene classification algorithm of remote sensing images based on the integrated convolutional neural network (CNN) is proposed. A back-propagation network is constructed to measure the complexity of scene images. The classification of these images is conducted with the CNN based on the complexity level of each image, thus, the scene classification of remoting sensing images is achieved. With the proposed algorithm, the experimental verification of the open data of NWPU-RESISC45 is conducted and the classification accuracy of 89.33% for Type I test and that of 92.53% for Type II are obtained, respectively. The average running time is 0.41 s. Compared with the VGG-16 model for fine tuning and training, the classification accuracy by the proposed algorithm is increased by 2.19% and 2.17%, respectively. Simultaneously, the prediction rate is increased by 33%. Thus, the efficiency and practicality of this proposed algorithm are confirmed.

Key words remote sensing; convolutional neural network; image complexity; scene classification

OCIS codes 280.4788; 100.4996; 100.2960

1 引 言

根据遥感图像的内容进行特征提取，使用分类器对特征进行分类，从而实现对遥感场景进行分类与识别。精准的场景分类可以降低地理目标检测、土地利用分析、土地覆盖分析、城市规划等遥感解译任务的难度[1]，并提高解译精度。遥感场景图像不仅包含颜色、纹理等低层信息，还包含很多语义层的信息，这也增大了其准确自动分类的难度。因此，遥感图像场景分类获得了航空和卫星图像分析领域研

收稿日期：2018-04-02；**修回日期**：2018-06-01；**录用日期**：2018-06-13

基金项目：国家自然科学基金青年基金(61505203)

* **E-mail**：ciomper@163.com；** **E-mail**：zhangxiaonan_93@163.com

究者的广泛关注。

早期的场景分类算法多数基于人工特征提取，使用工程性的技巧和专业知识设计针对不同任务的特征描述子，如基于颜色、纹理、空间信息、光谱信息的特征或多特征融合的描述子，这些特征都是场景图像的低层特征[2-4]。此类分类算法的泛化能力较弱，稳健性不强。

随着人工智能算法的不断发展，深度学习已经成为计算机视觉领域最有力的工具之一，其在目标识别、人脸检测、语音识别、语义分割等领域都有了突破性的进展[5-7]。遥感图像场景分类也是深度学习方法的受益者，尤其是卷积神经网络(CNN)在图像领域得到了广泛应用。Krizhevsky 等[8]提出了8层的 AlexNet 模型，大幅提高了图像分类的准确度。Simonyan 等[9]提出了16层的 VGG-16 模型和19层的 VGG-19 模型，分类准确度进一步提升。Szegedy 等[10]在加深网络的同时加宽网络，形成包含子网络的 GoogLeNet 模型。He 等[11]提出的 ResNet 模型解决了网络退化的问题。继承 ResNet 思想，Huang 等[12]提出了稠密连接的的 DenseNet 模型。在遥感场景分类领域，Cheng 等[2]构建了包含45类场景的遥感场景分类数据集 NWPU-RESISC45，并使用 AlexNet、VGG-16 及 GoogLeNet 模型对数据集进行了分类实验，其准确率远高于传统方法的。Yu 等[13]将 CNN 作为特征提取器，混合三种 CNN 提取出来的特征扩充了特征维度，并使用极端学习机(ELM)对扩充后的维度进行分类处理，在 NWPU-RESISC45 数据集上，相比于 VGG-16 模型，其算法的分类准确率提高了3.37%。

精调训练的浅层 CNN，如 AlexNet 模型能快速实现分类，但是分类准确率不够高，在 NWPU-RESISC45 数据集中抽取20%的数据进行训练时，准确率只能达到85.16%[2]。深层 CNN 如 VGG-16 模型能取得较高的分类准确率(90.36%[2])，但是其训练时间长，预测速度低，预测一张场景图像的时长是 AlexNet 模型的9.2倍，其原因主要是增多的卷积层增大了神经网络的计算复杂度。使用多特征混合的方法可以明显提高分类准确率，但是其预测速度仅为 VGG-16 模型的1/3～1/4。在执行遥感图像场景分类任务时，预测速度与分类准确度不能兼得。Li 等[14]的研究表明，遥感图像的复杂度与 CNN 的深度之间有一定的适应性关系，即分别使用浅层和深层的 CNN 对低复杂度和高复杂度的图像进行分类处理时，可以充分发挥多个 CNN 的固有特性，提高分类精度。

本文通过反向传播神经网络(BPNN)构建了图像复杂度的度量模型，训练了多个 CNN，提出了一种集成多个模型的遥感场景分类算法，可将 NWPU-RESISC45 数据集的分类准确率提高2.18%，并将预测速度提升33%。

2　集成神经网络概述

结合图像复杂度与深层 CNN 设计的集成网络结构如图1所示，遥感场景图像分类处理步骤如下。1)计算输入图像的复杂度相关参数，即颜色矩、灰度共生矩阵(GLCM)、信息熵和边缘检测结果；2)将这些参数输入到 BPNN 中计算复杂度，得到标定复杂程度的4种标签；3)根据标签和不同 CNN 的特性，选择对应的网络对场景图像进行特征提取和分类处理。所使用的复杂度相关参数的计算量较小，反向传播(BP)网络层数较少，前两个步骤的运行时间比 CNN 提取特征的运行时间小一个数量级，对集成网络运行效率的影响不大，充分发挥了浅层网络速度快和深层网络准确率高的优势，提高了分类效率。

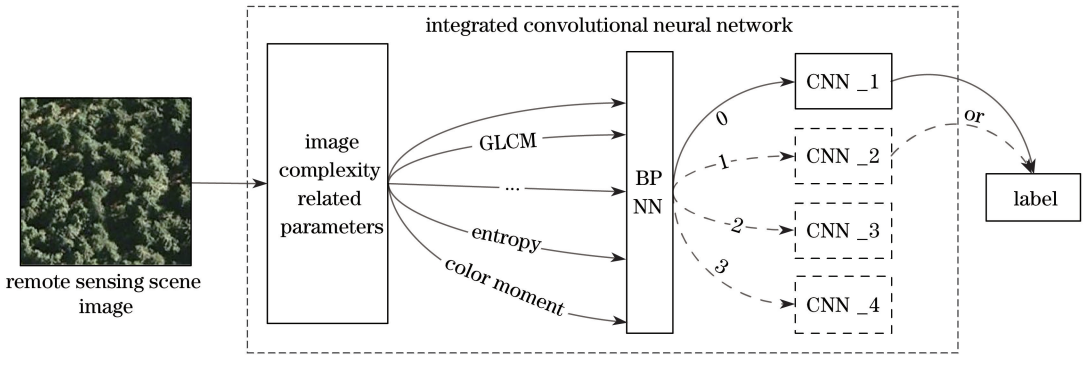

图1　集成神经网络结构

Fig. 1　Architecture of integrated neural network

2.1　使用的 CNN 模型

相比于日常场景图像，遥感影像具有画幅广、观测尺度大、图像特征尺度大等特点。经过对比筛选，使用 AlexNet、ResNet-50、ResNet-152、DenseNet-169 四种网络作为模型集成对象。

1）AlexNet

AlexNet[8]包含 5 个卷积层和 3 个全连接层，使用修正线性单元(RelU)作为激活函数，解决了梯度消失的问题。在第一个卷积层和第二个池化层后添加了正则化层。使用最大池化技术进行池化，在卷积层后使用 Dropout 技术防止过拟合。

2）ResNet

ResNet[11]解决了随着网络深度的加深网络退化的问题，设计了残差模块。假设 x_{l-1} 是网络中第 l 层的输入，类似 AlexNet 的直连式 CNN 在 l 层的输出可以形式化地表示为 $x_l=F(x_{l-1})$，残差块的设计使得 l 层的输出变为 $x_l=F(x_{l-1})+x_{l-1}$，即通过添加恒等映射的方式解决了网络的退化问题。在卷积层和池化层后引入批归一化层，使得网络更容易训练。这里所使用的 ResNet-50、ResNet-152 的分类层维数为适合所使用数据集的 45 维，其余结构参数与文献[11]保持一致。

3）DenseNet[12]

DenseNet 将 ResNet 的思想进行了扩充和发展，使用批正则化技术加速收敛、控制过拟合，将靠前的卷积、池化层的输出分别加在之后若干卷积层的输入上，即将 l 层的输出变为 $x_l=F(x_{l-1})+x_{l-1}+x_{l-2}+\cdots+x_1$，收敛速度比 ResNet 的更快，ImageNet 数据集的测试结果也有了明显的提升。训练的 DenseNet-169 的结构参数与文献[12]基本相同，为了满足所用数据集的需求，将最后的分类层由 1000 维改为 45 维。

2.2　使用 BP 网络度量图像复杂度

2.2.1　图像复杂度

图像复杂度的评估算法有很多，Peters 等[15]总结了 90 年代前基于边缘、灰度、尺寸等特征的复杂度描述方法；Rigau 等[16]提出了较为完整的基于信息论的图像复杂度衡量方法。此外，还有以图像纹理、模糊数学理论作为图像复杂度评判标准的研究[17]。近年来，多特征集成[18]、神经网络计算多特征参数的复杂度评估[19]取得了不错的效果。这些图像复杂度研究的评判标准是人工判别，但这并不适用于分类。所提算法对数据集中的多种场景进行分类测试，以分类准确率和分类效率作为图像复杂度的评判标准；在文献[19]和文献[20]的基础上加入了与颜色相关的描述子及信息熵，形成了包含颜色矩、灰度共生矩阵、信息熵、Canny 边缘检测线占比 4 种参数在内的复杂度评估参数集。

1）颜色矩

Stricker 等[21]提出的颜色矩是一种有效的颜色特征，计算复杂度相对较小，适用于复杂度评估。该方法的核心思想是使用矩表示图像中的颜色分布。所提算法采用颜色的一阶矩、二阶矩和三阶矩表达图像的颜色分布。因此，使用的颜色矩一共只需要 9 个分量(3 个颜色分量，每个分量上 3 个低阶矩)，与其他的颜色特征相比是非常简洁的。颜色矩的计算公式为

$$\mu_i=\frac{1}{N}\sum_{j=1}^{N}p_{i,j},\tag{1}$$

$$\sigma_i=\left[\frac{1}{N}\sum_{j=1}^{N}(p_{i,j}-\mu_i)^2\right]^{\frac{1}{2}},\tag{2}$$

$$s_i=\left[\frac{1}{N}\sum_{j=1}^{N}(p_{i,j}-\mu_i)^3\right]^{\frac{1}{3}},\tag{3}$$

式中 $p_{i,j}$ 为彩色图像第 j 个像素的第 i 个颜色分量；μ_i、σ_i 和 s_i 分别为第 i 个颜色分量的均值、方差和斜度；N 为图像中的像素个数。图像的色调(H)、饱和度(S)、明度(V)分量的前三个颜色矩组成一个 9 维向量，即颜色特征可以表示为

$$\boldsymbol{F}_{\text{color}}=[\mu_H,\sigma_H,s_H,\mu_S,\sigma_S,s_S,\mu_V,\sigma_V,s_V]。\tag{4}$$

2）灰度共生矩阵

灰度共生矩阵[22]是用来提取图像纹理信息的特征描述子，其定义为图像中一个灰度值为 m 的像素和与其相距 $\delta=(\Delta x,\Delta y)$ 的灰度级值为 n 的像素同时出现的联合概率分布。灰度共生矩阵的矩阵元表示为 $p(m,n,d,\theta)(m,n=0,1,2,\cdots,L'-1)$，其中 L' 为图像的灰度级，d 为灰度值分别为 m 和 n 的两个像素点间的距离，θ 为这两个像素点的方位关系。为了将纹理信息抽象成特征向量，从灰度共生矩阵中提取出 5 个常用的特征参数(能量、熵值、对比度、同质性、相关性)进行图像复杂度的计算。

能量用来度量图像灰度的均匀性和纹理粗细度。当图像的纹理较均匀且较粗时，其灰度共生矩阵中的值集中于对角线附近，能量 J 值比较大。J 值的计算公式为

$$J=\sum_{m=1}^{k}\sum_{n=1}^{K}p^2(m,n,d,\theta),\tag{5}$$

式中 k 为矩阵的行数；K 为矩阵的列数。

熵值 H_1 刻画了图像中纹理的信息量。熵值越

小，共生矩阵中的元素值越接近，即图像中的纹理特征越弱，纹理越少。熵值 H_1 的计算公式为

$$H_1=-\sum_{m=1}^{k}\sum_{n=1}^{K}p(m,n,d,\theta)lb\ p(m,n,d,\theta)。(6)$$

对比度 G 是共生矩阵主对角线附近的惯性矩，用来度量图像中纹理的清晰程度。对比度越大，图像中的纹理越明显。对比度的计算公式为

$$G=\sum_{m=1}^{k}\sum_{n=1}^{K}(m-n)^2p(m,n,d,\theta)。\quad(7)$$

同质性反映的是图像纹理的局部信息。若图像纹理在局部范围内的变化较小，则其同质性较高，Q 值较大。同质性 Q 的计算公式为

$$Q=\sum_{m=1}^{k}\sum_{n=1}^{K}\frac{1}{1+(m-n)^2}\cdot p(m,n,d,\theta)。(8)$$

相关性主要用于度量共生矩阵中元素在行(列)方向的相似程度，反映了图像中局部范围内像素灰度值之间的相关性。当行(列)像素的灰度值相似度高时，相关性值较大，图像的复杂度较小，反之复杂度较大。相关性的计算公式为

$$f_{\mathrm{COV}}=\sum_{m=1}^{k}\sum_{n=1}^{K}(m-\mu_1)(n-\mu_2)/(\sigma_1\sigma_2),\ (9)$$

式中 μ_1、μ_2 分别为归一化之后的矩阵中的元素沿行、列方向的均值；σ_1、σ_2 分别表示归一化之后的矩阵中的元素沿行、列方向的方差。

灰度共生矩阵的 5 个纹理特征描述子构成 5 维的向量，即纹理特征的表达式为

$$\boldsymbol{F}_{\mathrm{texture}}=[J,H_1,G,Q,f_{\mathrm{COV}}]。\quad(10)$$

通过计算图像的信息熵 E 和 Canny 边缘检测之后的线占比 L，可以得到描述图像复杂度所需要的 17 维向量。

2.2.2　BPNN

BPNN[20]的基本原理是通过多层全连接神经网络对输入向量 $\boldsymbol{X}$ 进行非线性运算得到输出向量 $\boldsymbol{Y}$，将输出向量 $\boldsymbol{Y}$ 与真实向量 $\boldsymbol{Y}'$ 之间的差异程度作为损失函数，通过 BP 与优化算法减小损失值，达到更新权重的目的。训练 BP 网络的数据集是原始数据集中每幅图像的复杂度参数 F_{complex}，标签是每幅图像的复杂度级别。

3　集成神经网络构建

集成神经网络的构建流程如图 2 所示，主要分为数据预处理、CNN 和 BP 网络训练、使用集成网络进行场景分类三个阶段。数据预处理阶段包括对遥感场景数据集的训练集进行数据增强处理及计算场景图像的复杂度相关系数两部分。训练 CNN 的输入为数据增强之后的训练集，输出为预测的场景类别；训练 BP 网络的输入为图像的复杂度相关系数，输出为图像的复杂度级别。进行分类预测时，先将遥感场景图像对应的复杂度相关系数输入到 BP 网络中，得到输入图像的复杂度级别，再根据复杂度级别选择对应的 CNN 进行预测，得到分类准确率、预测速度、混淆矩阵等评价指标。

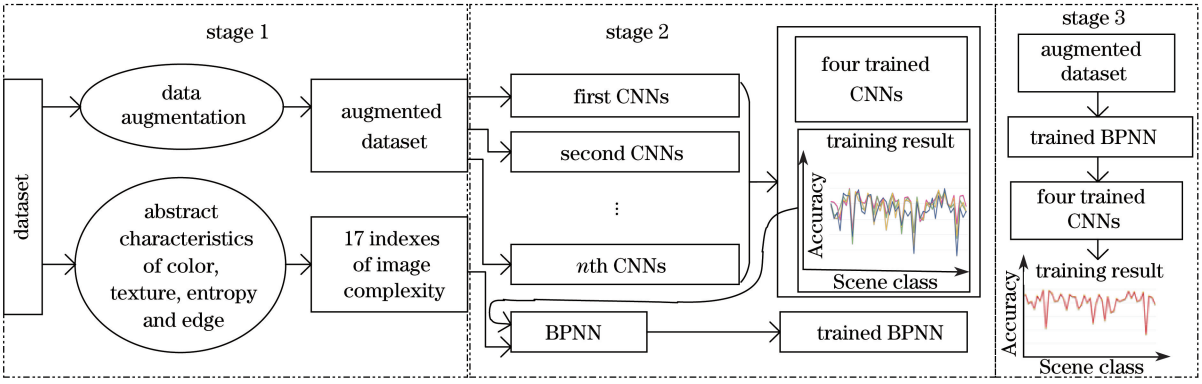

图 2　集成网络构建流程

Fig. 2　Flow chart of integrated network construction

3.1　实验环境及数据集

使用训练机进行模型训练，使用测试机进行模型预测。训练 CNN 使用 TensorFlow 1.4 框架，训练机环境为 Ubuntu 16.04 系统，使用 3.50 GHz Intel Core i7-5930K 中央处理器(CPU)，内存为 64 GB。使用 GTX Titan X GPU 进行加速运算，模型的预测环境为 Windows 10 系统，使用 2.50 GHz Intel Core i5-7300HQ CPU，内存为 8 GB，显卡为 GTX 1050Ti。

NWPU-RESISC45 数据集[2]包含 45 类遥感场景：飞机、机场、棒球场、篮球场、沙滩、桥梁、稀疏的植被、宫殿、圆形农田、云、商业区、密集住宅区、沙

漠、森林、公路、高尔夫球场、田径场、港口、工业区、十字路口、岛屿、湖泊、公园、住宅区、移动房屋停放场、山脉、立交桥、教堂、停车场、铁路、火车站、矩形农田、河流、环形路口、机场跑道、海冰、船只、湿地、稀疏住宅区、体育场、油罐、网球场、梯田、热电站、草地，每类包括 700 张 256 pixel×256 pixel 的红绿蓝三通道彩色(RGB)图像，空间分辨率从 0.2 m 到 30 m不等。数据集中的图像来自 Google 地图，覆盖的区域包括全球 100 多个国家和地区，总计 31500 幅图像。该数据集中的天气、季节、光照、视角等因素变化比较丰富，对于场景分类算法是一个考验。图 3 所示为 45 类场景图像中的代表性图像。

图 3　NWPU-RESISC45 数据集的场景图

Fig. 3　Scene images of NWPU-RESISC45 dataset

3.2　训练 CNN

使用 NWPU-RESISC45 数据集训练了 AlexNet、ResNet-50、ResNet-101、ResNet-152、DenseNet-121、DenseNet-161、DenseNet-169、Inception-v3、Inception-Res-v2、Xception 共 10 种 CNN 网络。为了与其他论文成果进行横向比较，在数据集划分上采取与文献[2]及[13]相同的策略，即设置两类实验，第一类实验从 NWPU-RESISC45 数据集的每类场景中提取 10%即 70×45 幅图像作为训练集，其余 90%作为测试集；第二类实验从每类场景中提取 20%即 140×45 幅图像作为训练集，其余 80%作为测试集。

在训练前对训练集进行了数据增强[23]处理，对训练集进行随机旋转、翻转、平移、剪切、放缩和对比度拉伸等变换操作，最终的训练集扩充成原来的 10 倍。训练阶段，打乱扩充后数据集的顺序，将其以批为单位输入到网络中，批数为 b_s，网络的输入为 b_s×256×256×3 的四维张量。将所有的训练集按批次输入到网络中进行训练称为一个循环，在训练 AlexNet 和 ResNet-50 时设定每批包含 256 幅图像，进行 300 次循环训练；在训练 ResNet-50 和 DenseNet-169 时每批包含 128 幅图像，进行 600 次循环训练。训练策略使用精调策略[3]。在训练过程中为了加速训练，使用 Adam 优化方法[24]，并设置阶梯式下降的学习率，如图 4 所示。

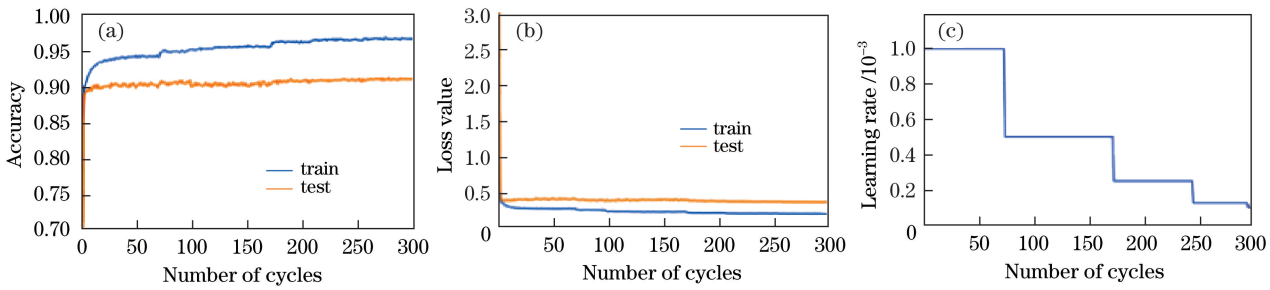

图 4　ResNet-50 训练过程中准确率、损失值和学习率随循环次数的变化。(a)准确率；(b)损失值；(c)学习率

Fig. 4　Accuracy, loss value and learning rate versus number of cycles in training process of ResNet-50. (a) Accuracy; (b) loss value; (c) learning rate

如图 4(c)所示，训练阶段设置的初始学习率为 0.001，之后通过监控测试损失对学习率进行调整。当测试损失持续 30 个循环不再减小时，将学习率降低为之前的 0.5。在训练 ResNet-50 时，学习率经过了 3 次自动衰减，从 0.001 衰减为 0.000125。由图 4(a)可知，随着学习率的减小，训练准确率在一定范围内有小幅度跃升的趋势，这也从侧面证明了渐变学习率对训练网络具有积极作用。在图 4(a)、(b)中，蓝色曲线和橘色曲线之间的差异是由于训练集在数据集中占比过低，存在过拟合现象，这样的低占比设置是为了与其他研究成果进行横向对比。

在测试阶段使用未打乱顺序的测试集对网络的分类性能进行测试，将同一种类别的测试集数据作为一个批次输入到网络，网络输出预测结果后可以方便地统计每一类的分类准确率，对所有类别的分类准确率取平均可以得到网络整体的分类准确率。统计了 10 种网络在两类实验中的分类准确率，在筛选网络的过程中综合考虑模型的运行时间、模型之间的互补性质及模型的分类准确率，其中 Inception-v3、Inception-Res-v2、Xception 网络在遥感数据集上的表现没有 ResNet 和 DenseNet 的出色，准确率不高，预测速度较慢，将其排除。在取舍 ResNet-101、DenseNet-121 及 DenseNet-169 时，最大程度保留了网络的多样性，包括网络深度、网络结构、卷积核参数的多样性等。为了保证 BP 神经网络的复杂度度量的准确率，只选择了 10 个模型中的 4 个，这是因为设计的 BP 神经网络的输入维度只有 17 维，输出维度越小，预测准确率越高。当输出维度增大到 5 维时，BP 网络复杂度预测的准确率会减小到 50%以下，这对于整个集成网络是不利的。筛选出来的 4 种网络在两类实验中的参数和测试结果见表 1。

表 1　各个网络的训练参数及结果

Table 1　Training parameters and results of each network

Model	Input size /(pixel×pixel)	Batch size /frame	Number of cycles	Training accuracy /%	
				Experiment I	Experiment II
AlexNet	224×224	256	300	81.22	85.46
ResNet-50	224×224	256	300	86.52	90.52
ResNet-152	224×224	128	600	85.11	90.11
DenseNet-169	224×224	128	600	82.44	87.44
VGG-16[2]	-	-	-	87.15	90.36
Proposed model	-	-	-	88.47	92.53

由表 1 可知，第二类实验的准确率均高于第一类实验的，即所有网络的准确率都会随着训练集占比的增大而增大。单个模型中表现最突出的是 ResNet-50，其平均分类准确率最高；AlexNet 的平均分类准确率最低，但是其层数最少，速度最快。DenseNet-169 的层数最多，整体准确率反而有所减小，这说明层数增多不一定会带来整体准确率的提升，选择 DenseNet-169 是因其在树丛、海冰、岛等场景的识别准确率上高于其他网络的。在两类实验中，集成模型的分类准确率均高于其余单个网络的。

图 5 所示为第二类实验中训练的 4 个 CNN 的各类识别准确率，可以看出，AlexNet 对沙滩、雪山、森林、桥梁、云、公路这些比较宏观的场景进行分类时的识别率较高；ResNet-50 对飞机、棒球场、篮球场、高尔夫球场、田径场这些具有固定形状的场景进行分类时的识别率较高；DenseNet-169 对树丛、海冰、岛、公园、港口这些含有更多语义特征的抽象场景进行分类时的识别率较高；ResNet-152 对热电站、湖泊、活动房屋停放场、梯田等场景进行分类时的识别率较高。

文献[4]的研究显示，遥感图像的复杂度与 CNN 的深度之间有一定适应性关系，即分别使用浅层和深层的 CNN 对低复杂度和高复杂度的图像进行分类处理时，可以充分发挥 CNN 的固有特性，提高分类精度。从图像复杂度角度出发，沙滩、雪山、云所包含的信息量及纹理颜色信息比飞机、球场、梯田等的少很多，即宏观场景的图像复杂度较小。因此，将能够通过浅层 AlexNet 准确识别的场景图像认定为简单场景，复杂度级别为 1；将能通过中层 ResNet-50 准确识别的场景图像的复杂度设置为 2；将能在 ResNet-152 和 DenseNet-169 网络中准确识别的场景的复杂度级别分别设置为 3 和 4。

3.3　训练复杂度度量 BP 网络

通过统计图 5 的结果可以制作训练 BP 网络所需要的输出标签，将 BP 网络的输入设置为与图像复杂度相关的 17 个参数，即 9 个颜色矩参数、6 个

图 5　CNN 的分类结果

Fig. 5　Classification results based on CNN

纹理参数、线占比和信息熵。

通过计算图像的信息熵 E 和 Canny 边缘检测之后的线占比 L，可以得到描述图像复杂度所需要的 17 维向量。经过归一化处理之后，将 17 维向量作为 BP 神经网络的输入向量。

设计的 BP 网络采用一维卷积对输入的 17 维数据进行组合，可以达到更好的分类效果。该层的卷积核大小设置为 11，步长设置为 2，输出向量维度为(None，4，32)，其中 None 代表训练过程中一个批次包含的数据数量，其不会随网络结构的变化而发生变化。激活层以 ReLU 作为激活单元，之后将激活后向量通过最大池化操作进行下采样处理。经过维度拉伸及两个全连接层可以得到 4 维输出向量，该输出即为图像的复杂度标签。

随机在数据集中选取 80% 即 45×560 幅图像作为训练集，剩下的 20% 作为测试集，BPNN 的训练集和测试集不需要与 CNN 的训练集、测试集相同。训练 BP 网络的输入为场景图像的 17 维复杂度描述子 F_{complex}，即先打乱训练集的顺序，再计算每幅图像的 F_{complex}，以批为单位将一批 F_{complex} 输入到网络中进行训练。输出为每幅图像的复杂度级别。训练使用随机梯度下降(SGD)方法[25]，学习率设置为 0.001，将 b_s 设置为 4500，进行 10000 次循环训练。训练结果如图 6 所示，可以看出，训练的 BP 网络的准确率为 66.73%，即输入一幅图像时，BP 网络有 66.73% 的概率为该图像匹配到最合适的 CNN 模型进行场景分类处理。BP 网络的准确率越高，整个集成模型的性能提升越高。

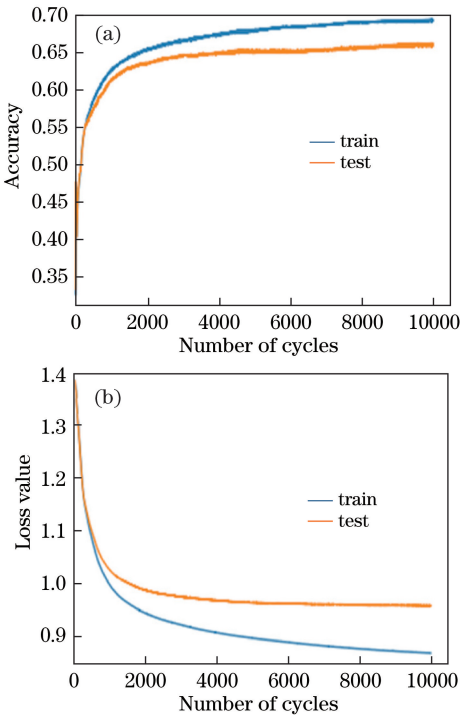

图 6　BP 网络的训练过程中准确率和损失值随循环次数的变化曲线。(a)准确率；(b)损失值

Fig. 6　Accuracy and loss value versus number of cycles in training process of BP network. (a) Accuracy; (b) loss value

4　实验结果与分析

在使用集成后的模型进行预测时，先计算待预测图像的复杂度描述子 F_{complex}，然后将 F_{complex} 输入到 BP 网络中，BP 网络输出复杂度级别，根据复杂度级别选择其对应的 CNN 网络，最后对场景图像进行预测。对未参加训练的测试集进行分类预测后统计其分类性能，得到混淆矩阵，如图 7 所示，混淆矩阵中每一个元素 $f_{\text{Con}}(i,j)$ 代表标签为 i 的图像被识别为 j 的概率。其对角线元素代表每一类的识别准确率，可以看出，设计的集成网络的分类性能较好，45 类场景中有 33 类可达到 90% 以上的识别率，

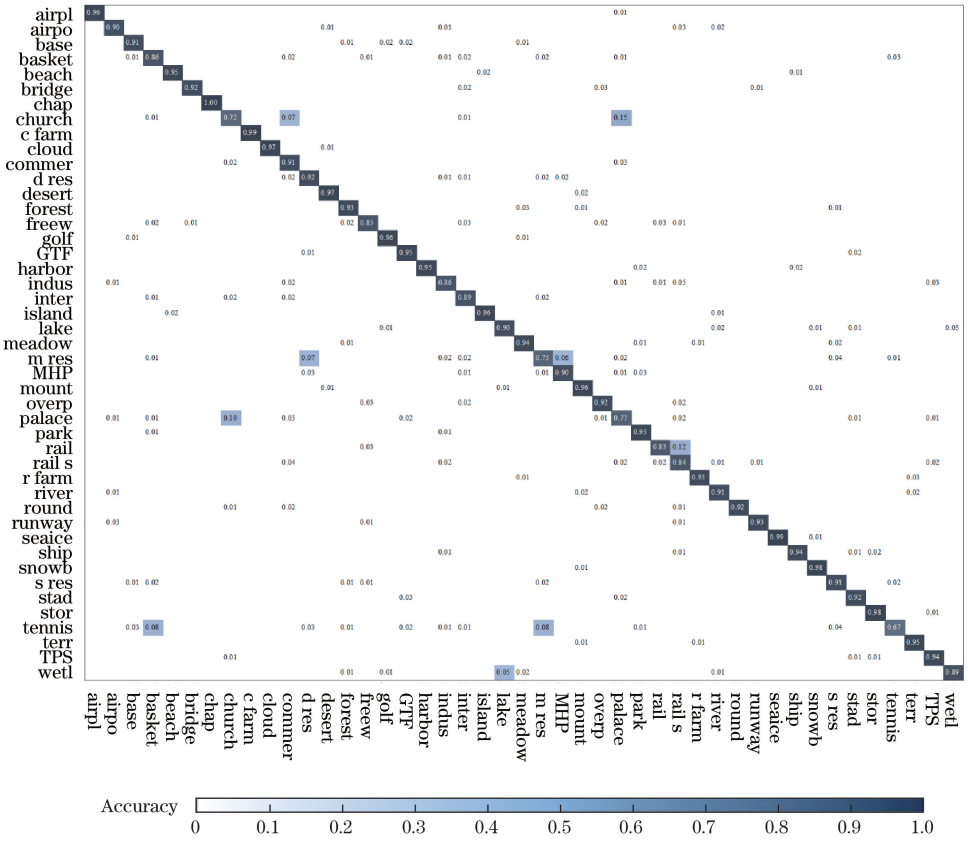

图 7　集成模型对数据集进行分类预测后得到的混淆矩阵

Fig. 7　Confusion matrix obtained after classification prediction of dataset by integrated model

对树丛的识别准确率为 100%，识别率最低的是网球场，只有 67%的识别准确率。分类网络容易误判的场景还有中度住宅区和密集住宅区、火车站和铁路、宫殿和教堂等。其原因一方面是这些场景的相似度较高，另一方面是数据集的数量较少。在实际遥感解译任务中通过增大数据量可以提高网络的分辨能力。

为了对比所提算法与其他算法的性能，根据分类算法的混淆矩阵，绘制了文献[2]中几种代表性算法、所提算法及通过多数投票方式集成的网络结构在 NWPU-RESISC45 数据集上的分类准确率，如图 8 所示。从文献[2]中基于低层特征的算法中选取基于颜色直方图的分类算法，从基于非监督学习的方法中选取视觉词袋(BoVW)模型，从基于深度学习的方法中选取 VGG-16 模型，所选取的三种算法均是其同类算法中准确率较高的算法。同时，为

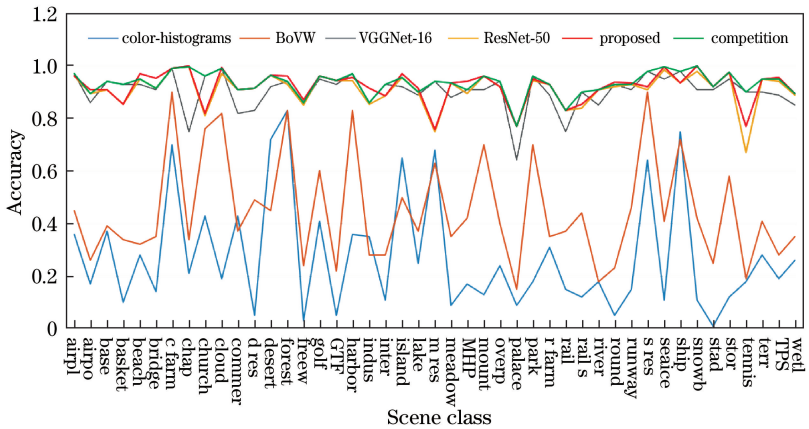

图 8　与其他算法单类的准确率对比

Fig. 8　Single accuracy comparison with those of other algorithms

了检验所提集成方法的效果，添加了没有采用集成策略的单一 ResNet-50 模型和采用多数投票方式集成的竞争模型。对比实验的训练集与测试集的选取与第三节所述的第二类实验的相同，即随机选取20%作为训练集，80%作为测试集。训练与测试环境采用第三节所述的训练机和测试机，结果如图 8 所示。可以看出，采用深度学习的算法在每个类别上的分类准确率均高于提取低层特征的算法及非监督学习方法的。

为了从数值上对各种算法进行性能比较，测试并统计了图 8 中算法的平均准确率、准确率标准差、单幅场景影像的平均预测时间，结果见表 2。预测时间是指在模型训练完毕后，从输入一幅测试集遥感影像到得到该影像的场景类别所用的时间。在所提算法中，预测时间包含对测试集影像提取复杂度描述子、复杂度度量、CNN 预测类别三部分的时间。

表 2　几种算法的性能对比

Table 2　Performance comparison among several algorithms

Method	Color-histogram	BoVW	VGG-16	ResNet-50	Proposed	Competition
Accuracy /%	27.52	44.97	90.36	90.59	92.53	93.41
Standard deviation	0.2184	0.2051	0.0673	0.0657	0.0593	0.0451
Prediction time /s	-	-	0.62	0.47	0.41	2.26

由表 2 可知，所提集成模型的准确率和预测时间均优于 VGG-16 及 ResNet-50 的，而基于多数投票策略的竞争网络在分类实验中表现出了非常高的准确率，但该方法在预测阶段需要使用 4 个网络分别进行预测，预测时间大于 2 s，效率较低。随着平均准确率的增大，准确率的标准差减小，说明分类系统的稳健性提高，而且标准差的减小有利于在遥感自动解译任务中对错分类行为进行风险评估。

在第二类实验中，集成网络与其他算法的性能比较见表 3，其中，SPM 表示空间金字塔匹配。可以看出，基于全局特征信息(GIST)、局部二进制模式(LBP)、局部约束线性编码(LLC)、视觉词袋(BoVW)等提取低层特征的传统方法在该数据集上的分类准确率低于 50%，采用深度 CNN 可将准确率提升到 80%以上；采用深度更大、参数更多的 GoogLeNet 对数据集进行分类测试时，其结果逊于 VGG 的，这是因为这些网络的初始化权重都是在自然场景数据集上训练得到的，并不适合于遥感场景的分类[2]，故只挑选了 4 个有优势的网络作为集成网络的子网络，这 4 种网络中性能最高的是 ResNet-50。文献[13]采用的 Two-stream 深度融合框架(DFF)算法的训练阶段未选择精调策略，其准确率比其他没有采用精调策略的同类算法的高，但并不及采用精调策略训练的网络。训练的 ResNet-50 模型及混合模型的准确率远远大于 GIST、BoVW 等传统分类方法的，采用精调训练策略时单一模型的准确率比文献[13]所采取的训练策略的高 7%。集成网络在分类任务中可以提高整体的准确率，相比于 ResNet-50，其分类准确率分别高 1.65%(第一类实验)、1.94%(第二类实验)；相比于文献[2]中性能最高的 VGG-16，其分类准确率分别高 2.19%(第一类实验)、2.17%(第二类实验)，故集成网络在遥感场景分类任务中具有一定的优势。

表 3　与其他方法的平均准确率对比

Table 3　Average accuracy comparison with those of other algorithms

Method	Accuracy /% (experiment I)	Accuracy /% (experiment II)
GIST[2]	15.90	17.88
LBP[2]	19.20	21.74
Color histograms[2]	24.84	27.52
BoVW+SPM[2]	27.83	32.96
LLC[2]	38.81	40.03
BoVW[2]	41.72	44.97
GoogLeNet[2]	82.57	86.02
VGG-16[2]	87.15	90.36
AlexNet[2]	81.22	85.16
Two-streamDFF[13]	80.22	83.16
ResNet-50	87.69	90.59
Proposed model	89.34	92.53

所提集成网络除了在准确率上有所提升之外，分类所消耗的预测时间也比单个模型的短，如图 9 所示。AlexNet 在测试机上的预测时间为 0.07 s，整体分类准确率为 85.16%(第二类实验)；文献[2]训练的 VGG-16 模型在测试机上的预测时间为 0.62 s，分类准确率为 90.36%；训练的 ResNet-50 模型在测试机上的预测时间为 0.47 s，预测速度比 VGG-16 模型的提升了 12%，分类准确率为 90.59%；

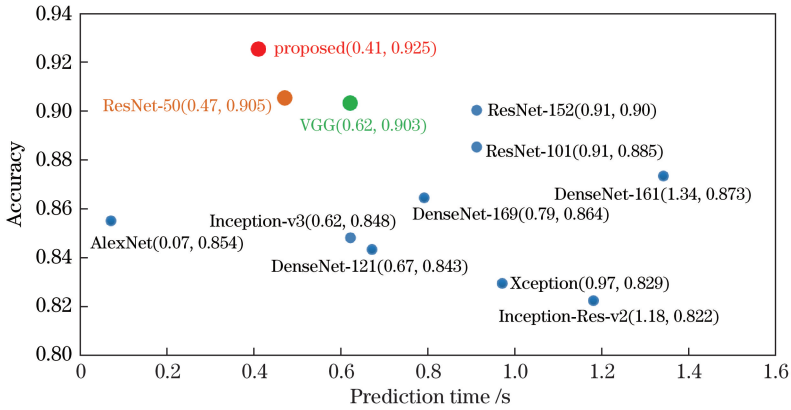

图 9　多种模型的分类准确率和预测时间

Fig. 9　Classification accuracies and prediction time of various models

集成网络在测试机上计算复杂度及使用 BP 网络进行复杂度度量的整体运行时间为 0.06 s，整个集成网络的分类平均预测时间为 0.41 s，分类准确率为 92.53%，相比于 VGG-16 模型，预测速度提升了 33%，准确率提升了 2.17%。

由图 5 及图 9 可知，集成神经网络能够提高分类准确率和分类预测速度，主要原因是其利用了 AlexNet 速度快和 ResNet-50、ResNet-152、DenseNet-169 分类准的优势，而二者的比例关系是影响集成网络性能的关键。通过重新设置场景图像的复杂度级别标签，控制使用 AlexNet 进行分类的简单场景的种类数量，统计了集成网络的预测速度和分类准确率，如图 10 所示。可以看出，使用 AlexNet 进行分类的简单场景的种类数量越多，分类速度越快；分类准确率先是小幅度增大后出现急剧减小，其原因在于单个 AlexNet 本身不能为所有场景提供很高的分类准确率。在实际遥感的自动解译任务中，可以根据任务所需要的准确率和预测速度对简单场景的数量进行设定。

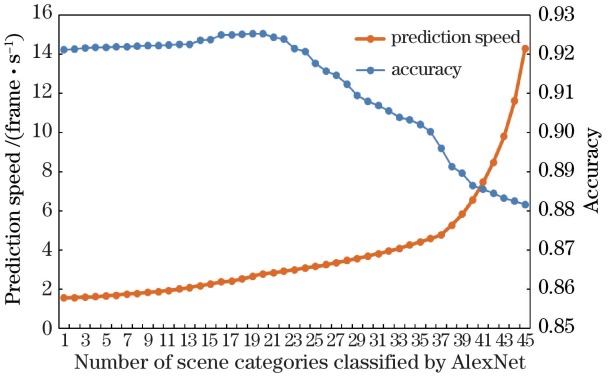

图 10　使用 AlexNet 进行分类的场景类别数量对集成模型的性能影响

Fig. 10　Impact of number of scene categories classified by AlexNet on performance of integrated model

5　结　　论

集成了多个 CNN 对遥感场景进行分类识别，在使用 CNN 进行预测前判别图像复杂度，进而找到与待预测场景最匹配的 CNN。利用浅层网络识别复杂度较低的场景图像，达到快速识别的目的；通过深层的网络识别复杂度较高的场景图像，达到精准识别的目的。研究结果表明，ResNet-50 在 NWPU-RESISC45 数据集上的表现优于 VGG-16 的，设计的集成模型在预测准确率和预测速度上均优于 VGG-16 和 ResNet-50 等单一模型，分类的整体效率得到提高。实验结果证明了集成模型的高效性和可行性。

集成模型需要训练并存储多个模型，相比于单个卷积神经网络，需要更长的训练时间和模型存储空间，可以通过适当的模型压缩减少模型参数的数量，还可采取多显卡并行优化的加速训练方法来缩减训练时长。

参　考　文　献

[1] Li E Z, Xia J S, Du P J, *et al*. Integrating multilayer features of convolutional neural networks for remote sensing scene classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(10): 5653-5665.

[2] Cheng G, Han J W, Lu X Q. Remote sensing image scene classification: Benchmark and state of the art[J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.

[3] Zou Q, Ni L H, Zhang T, *et al*. Deep learning based feature selection for remote sensing scene classification[J]. IEEE Geoscience and Remote

Sensing Letters, 2015, 12(11): 2321-2325.

[4] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(8): 1778-1790.

[5] Hu F, Xia G S, Hu J W, *et al*. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery[J]. Remote Sensing, 2015, 7(11): 14680-14707.

[6] Stumpf A, Kerle N. Object-oriented mapping of landslides using random forests[J]. Remote Sensing of Environment, 2011, 115(10): 2564-2577.

[7] Wang Y B, Zhang L Q, Tong X H, *et al*. A three-layered graph-based learning approach for remote sensing image retrieval[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10): 6020-6034.

[8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]. International Conference on Advances in Neural Information Processing Systems, 2012: 1097-1105.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10)[2018-03-15]. https://arxiv.org/pdf/1409.1556v6.pdf.

[10] Szegedy C, Liu W, Jia Y Q, *et al*. Going deeper with convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1-9.

[11] 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, 2016: 770-778.

[12] Huang G, Liu Z, van der Maaten L, *et al*. Densely connected convolutional networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[13] Yu Y L, Liu F X. A two-stream deep fusion framework for high-resolution aerial scene classification[J]. Computational Intelligence and Neuroscience, 2018, 2018: 8639367.

[14] Li H F, Peng J, Tao C, *et al*. What do we learn by semantic scene understanding for remote sensing imagery in CNN framework?[EB/OL]. (2017-05-19)[2018-03-16]. https://arxiv.org/ftp/arxiv/papers/1705/1705.07077.pdf

[15] Peters R A, Strickland R N. Image complexity metrics for automatic target recognizers[C]. Automatic Target Recognizer System and Technology Conference, 1990: 1-17.

[16] Rigau J, Feixas M, Sbert M. An information-theoretic framework for image complexity[C]. Computational Aesthetics 2005: Eurographics Workshop on Computational Aesthetics in Graphics, Visualization and Imaging, 2005: 177-184.

[17] Cardaci M, Di Gesù V, Petrou M, *et al*. On the evaluation of images complexity: A fuzzy approach[C]. International Workshop on Fuzzy Logic and Applications, 2005: 305-311.

[18] Song Q H, Chen Z B, Sun S H, *et al*. A scene recognition method based on image complexity[J]. Proceedings of SPIE, 2014, 9282: 928221.

[19] Chen Y Q, Duan J, Zhu Y, *et al*. Research on the image complexity based on neural network[C]. Machine Learning and Cybernetics, 2015, 1: 295-300.

[20] Chen Y Q, Duan J, Zhu Y, *et al*. Research on the image complexity based on texture features[J]. Chinese Optics, 2015, 8(3): 407-414.
陈燕芹, 段锦, 祝勇, 等. 基于纹理特征的图像复杂度研究[J]. 中国光学, 2015, 8(3): 407-414.

[21] Stricker M A, Orengo M. Similarity of color images[J]. Proceedings of SPIE, 1995, 2420: 381-393.

[22] Gao C C, Hui X W. GLCM-based texture feature extraction[J]. Computer Systems and Applications, 2010, 19(6): 195-198.
高程程, 惠晓威. 基于灰度共生矩阵的纹理特征提取[J]. 计算机系统应用, 2010, 19(6): 195-198.

[23] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning[EB/OL]. (2017-12-13)[2018-03-17]. https://arxiv.org/pdf/1712.04621.pdf.

[24] Kingma D P, Ba J L. Adam: A method for stochastic optimization[EB/OL]. (2017-01-30)[2018-3-17]. https://arxiv.org/pdf/1412.6980.pdf.

[25] Wang B B, Wang Y X. Some properties relating to stochastic gradient descent methods[J]. Journal of Mathematics, 2011, 31(6): 1041-1044.
汪宝彬, 汪玉霞. 随机梯度下降法的一些性质[J]. 数学杂志, 2011, 31(6): 1041-1044.