文章编号　1004-924X(2020)06-1404-10

循环神经网络多标签航空图像分类

陈科峻[1,2]，张　叶[1*]
(1. 中国科学院大学 长春光学精密机械与物理研究所 应用光学国家重点实验室，吉林 长春 130033；
2. 中国科学院大学，北京 100039)

摘要：由于航空图像背景复杂，包含的物体类别多样，航空图像分类任务仍然面临困难。针对传统航空图像多标签分类算法准确率低、泛化性差的问题，本文提出了一种基于循环神经网络多标签航空图像分类方法。首先，采用超像素分割获取图像的低层特征，通过注意力机制生成注意力特征图；接着，采用交叉验证的方式获取最佳的图像尺度，将多尺度注意力特征图嵌入卷积神经网络中对图像进行特征提取；最后，采用改进的双向长短期记忆网络挖掘标签之间的相关性，改进的双向长短期记忆网络增加了输入门到输出门之间的连接，使输入状态可以更好地控制每一内存单元输出的信息，并且将遗忘门和输入门合并成单一的更新门，使得改进的双向长短期记忆网络可以学到更长时期的历史信息。结果显示，在图像变换尺度为 1，1.3，2 时，模型在 UCM 多标签数据集上的精确率和召回率分别达到了 85.33% 和 87.05%，F1 值达到了 0.862。本文方法相比于原始 VGGNet16 模型，精确率提高了 7.25%，召回率提高了 8.94%。实验表明，该方法可以有效提高航空图像多标签分类任务的准确率。

关　键　词：航空图像分类；多标签；注意力机制；多尺度；卷积神经网络；长短期记忆网络
中图分类号：TP391.7　　**文献标识码**：A　　**doi**：10.3788/OPE.20202806.1404

Recurrent neural network multi-label aerial images classification

CHEN Ke-jun[1,2], ZHANG Ye[1*]

(1. *Changchun Institute of Optics Fine Mechanics and Physics, Chinese Academy of Sciences, State Key Laboratory of Applied Optics, Changchun*, 130033, *China*;
2. *Chinese Academy of Sciences. Beijing* 100039, *China*)
* *Corresponding author, E-mail: yolanda@spirit.ai*

Abstract: Due to the complexity of the background in aerial images and the diversity of object categories, aerial image classification is a challenging task. In order to address the problems of low accuracy and poor generalization in traditional multi-label aerial image classification methods, a method based on recurrent neural networks was proposed. In this method, the super-pixel segmentation algorithm was first used to obtain the low-level features of the image from which an attention map was generated. Subsequently, the best image scale was obtained by cross-validation, and multi-scale attention feature graphs were embedded into aconvolutional neural network in order to extract the features of the image. Finally, tomine the correlation between labels, an improved bidirectional Long Short-Term Memory (LSTM) network was proposed, which increases the connection from the input gate to the

收稿日期：2019-12-02；**修订日期**：2020-01-21.
基金项目：中国科学院青年创新促进协会基金资助项目(No. 2016201)

output gate, so that the input state can efficiently control the output information of each memory unit. The forget gate and the input gate were combined into a single update gate so that the improved bidirectional LSTM network can learn long-term historical information. The results obtained by applying the proposed method to the UCM multi-label dataset indicate that for scale values of 1, 1.3, and 2, the accuracy and recall rates of the model are 85.33% and 87.05% respectively, while the F1 score reached 0.862. The accuracyand recall rates are found to be higher than those of theVGGNet16 model by 7.25% and 8.94% respectively. The experimental results thus indicate that the proposed method can effectively increase the accuracy of multi-label aerial image classification.

Key words: satellite images classification; muilti-label; attention mechanisms; multi-scale; convolutional neural network; Long Short-Term Memory(LSTM) network

1　引　言

随着航空技术的日益成熟，航空图像分辨率日益提高，航空图像在人们日常生活中发挥着越来越重要的作用。自然灾害探测、城市规划、资源勘探及专题地图制作等任务都离不开航空图像分类，因此对航空图像进行准确分类具有重要的意义。在实际的场景中，航空图像通常包含多个不同类别的物体，一张图像往往和多个标签相关联。这在一定程度上给分类模型带来了干扰，即类内呈现较大的多样性，而类间具有较大的相似性。同时，受到视点、旋转、背景等多种变化的影响使得航空图像多标签分类任务依然存在严峻的挑战。

近年来，随着基于深度学习的图像分类算法取得了重大的突破[1]，利用计算机视觉技术对航空图像进行分类成为了当前研究的热点[2]，同时也涌现出了大量的相关研究工作：文献[3]采用遗传算法优化 LVQ(Learning Vector Quantization)神经网络的权值与阈值，同时融入相似灰度值创建分类图像特征矢量输入神经网络中进行训练。但基于浅层神经网络的分类模型往往由于特征提取能力有限而导致分类精度不高。文献[4]提出了一种基于多尺度特征融合(Multiscale Features Fusion, MSFF)的航空图像分类方法，对各卷积层和全连接层提取出的不同尺度的特征进行降维和池化操作。将各尺度特征进行编码融合并利用多核支持向量机(Multikernel Support Vector Machine, MKSVM)进行场景分类。但是，基于 SVM(Support Vector Machine)的分类方法往往受限于样本的特征提取方式和核函数参数的选取。文献[5]通过多层卷积神经网络对航空图像进行卷积和池化处理，抽取高层特征构建图像特征库，并根据图像特征和图像特征库中特征向量之间的距离大小，进行图像检索分类。但是通过距离度量图像之间的相似性计算量较大并且容易造成信息的丢失。而对于多标签航空图像分类问题，现有的研究[6-7]大都是假设标签类之间是相互独立的，缺乏对标签之间潜在相关性的挖掘。例如，船舶往往出现在包含有水域或码头的图像中，汽车往往与路面和建筑物共同出现在同一张图像中。一个良好的多标签分类系统不仅需要能够学习整体特征，还应该能够利用标签间的相关性特点进行图像分类。

本文针对传统航空图像分类模型对图像主体特征提取能力不足的问题，提出了采用预训练卷积神经网络结合注意力机制的特征提取方法，将多尺度注意力特征图与卷积神经网络高层特征加权对航空图像进行特征提取。注意力机制的引入，提高了图像主体在模型中的权重，增加了特征的可分性。同时，针对传统研究中缺乏对航空图像标签之间相关性的挖掘的问题，本文考虑到不同标签之间存在着潜在的相互依赖关系，采用改进的双向长短期记忆网络(Bidirectional Long Short-Term Memory, BiLSTM)对图像标签间的高阶依赖关系进行挖掘，以提高网络的多标签分类能力。实验结果表明，本文方法能够在一定程度上提高航空图像多标签分类任务的准确率。

2　多尺度注意力的特征提取网络

传统的卷积神经网络在对图像进行特征提取时，往往难以有效地将图像主体特征与噪声特征区分开来。为了有效提取图像中的主体特征，本

文特征提取网络受到文献[8]的启发，将不同尺度的注意力特征图嵌入卷积神经网络中，以使得网络能够捕获场景中主体对象的显著性信息，提高模型的特征提取能力。同时，将卷积神经网络输出的特征序列化，作为双向长短期记忆网络的输入，利用双向长短期记忆网络的序列学习能力对标签之间的相关性进行挖掘，输出最终的图像类别标签。本文总体的网络结构如图 1 所示。

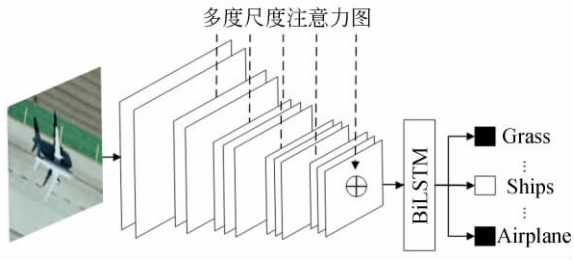


图 1 总体网络框架

Fig. 1 Overall architecture of network

2.1 注意力图生成

航空图像背景复杂，为了提高卷积神经网络对图像主体的特征提取能力，本文将注意力机制与卷积神经网络相结合对图像特征进行提取。基于注意力机制的卷积神经网络首先需要生成注意力特征图。本文通过对图像进行超像素分割，采用计算特征熵的方式提取图像的最优特征，生成注意力特征图。

超像素分割算法（Simple Linear Iterative Clustering，SLIC）[9] 通过将彩色图像转化为 CIELAB 颜色空间和 XY 坐标下的 5 维特征向量，然后对特征向量构造距离度量标准，对图像像素进行局部聚类。SLIC 算法在运算速度、物体轮廓保持及超像素形状方面具有较高的综合性能。在特征提取阶段，可以从超像素图像中获得颜色、纹理、梯度等 12 个低层特征。本文采用计算特征熵的方式提取最优特征，特征熵的计算公式如式(1)所示：

$$entropy = \sum_{I=0}^{255} p_I \ln p_I, \tag{1}$$

式中 p_I 表示特征图中灰度值为 I 的像素在图中所占的比例。选择熵最大的 8 个特征作为最优特征，计算显著性分数，生成注意力特征图。对注意力特征图进行尺度变换之后，嵌入卷积神经网络中对网络进行训练。显著性分数计算公式如式(2)所示：

$$S(s_i) = \left[\sum_{j=1, j\neq i}^{N} \frac{\sqrt{\sum_{m=1}^{8} (F_m(s_i) - F_m(s_j))^2}}{1 + dis(s_i, s_j)} \right] \times c(s_i), \tag{2}$$

其中 $F_m(s_i)$ 表示超像素 s_i 对应的第 m 个特征。$c(s_i)$ 的计算公式如式(3)所示：

$$c(s_i) = \exp\left(-\frac{(x_i - x')^2}{2v_x^2} - \frac{(y_i - y')^2}{2v_y^2} \right), \tag{3}$$

其中：$c(s_i)$ 为超像素的坐标 (x_i, y_i) 与图像中心坐标 (x', y') 之间的距离；v_x 和 v_y 是由图像的水平和垂直信息决定的变量；$dis(s_i, s_j)$ 计算公式如式(4)所示：

$$dis(s_i, s_j) = [(L_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2 + \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{Z^2}\beta^2]^{\frac{1}{2}}, \tag{4}$$

其中：[L A B]表示 CIELAB 颜色空间像素的三个颜色分量；(x_i, y_i)，(x_j, y_j) 分别表示超像素 s_i，s_j 的空间坐标；Z 为相邻超像素的空间距离；β 为常数，取值范围为[1，40]；$dis(s_i, s_j)$ 表示超像素之间的颜色-空间加权距离。最终，通过注意力机制生成的注意力特征图可视化结果如图 2 所示。

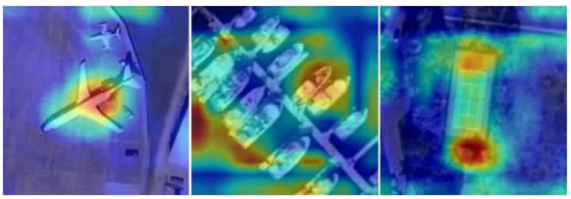


图 2 注意力图

Fig. 2 Attention map

从注意力图可以看出，嵌入注意力机制的特征提取方法，能够突出图像中的主体信息，有利于提高图像特征的可分性。

2.2 特征提取基础网络

在完成注意力特征图的生成之后，需要将注意力特征图嵌入卷积神经网络当中。因此构建基于注意力机制的特征提取网络的第二步则是构建基础卷积神经网络。

近年来，随着深度学习算法的兴起，涌现出了许多经典的卷积神经网络模型。其中，文献[10]对传统的卷积神经网络结构进行改进，提出的 ResNet（Residual Network）解决了层数过深的卷积神经网络模型在训练过程中发生退化而导致

难以收敛的问题，大大提高了卷积神经网络的分类精度。因此，本文以 ResNet50 作为基础网络对输入图像进行特征提取。ResNet50 网络结构参数如表 1 所示。

表 1　ResNet50 网络参数

Tab. 1　Parameters of ResNet50

层名	输出尺寸	ResNet50
卷积层 1	112×112	7×7，64
		最大池化，步长 2
卷积层 2_x	56×56	[1×1　64; 3×3　64; 1×1　256] ×3
卷积层 3_x	28×28	[1×1　128; 3×3　128; 1×1　512] ×4
卷积层 4_x	14×14	[1×1　256; 3×3　256; 1×1　1 024] ×6
卷积层 5_x	7×7	[1×1　512; 3×3　512; 1×1　2 048] ×3
	1×1	平均池化层

ResNet50 网络将多个残差块堆叠在一起，残差块之间存在直连通道，用于将输入信息直接连接到输出端，并且每个残差块都是由卷积核大小为 1×1，3×3，1×1 的三个卷积串联在一起，这样的做法保存了信息的完整性，整个网络只需要学习输入、输出之间的残差，简化了学习目标和难度。通过 ResNet50 提取图像特征所得到的卷积特征图如图 3 所示。

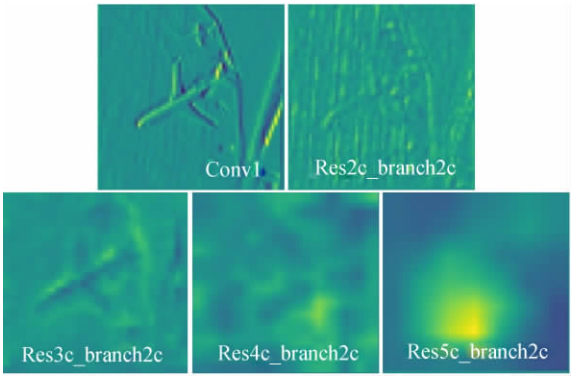


图 3　卷积特征图

Fig. 3　Convolutional feature map

从图 3 可以看出，ResNet50 卷积神经网络的浅层对图像中的轮廓具有很好的特征提取能力。越深的卷积层则越利于提取图像的抽象特征。

最终得到的基于注意力机制的特征提取网络结构如图 4 所示。其中，ResNet50 的卷积层 1 至卷积层 5_x 的结构相同。在每个分支网络的卷积层都嵌入注意力图，使得网络能够从图像中提取到更多层次的对象信息。同时，为了提高模型的收敛速度，本文在网络的激活函数之前，增加批归一化层（Batch Normalization，BN）[11]，提高模型的泛化能力。由于不同尺度大小的输入图像经过卷积和池化之后得到的特征图大小不同，本文在网络结构的最后增加了全局平均池化，以便于对多尺度特征进行融合。

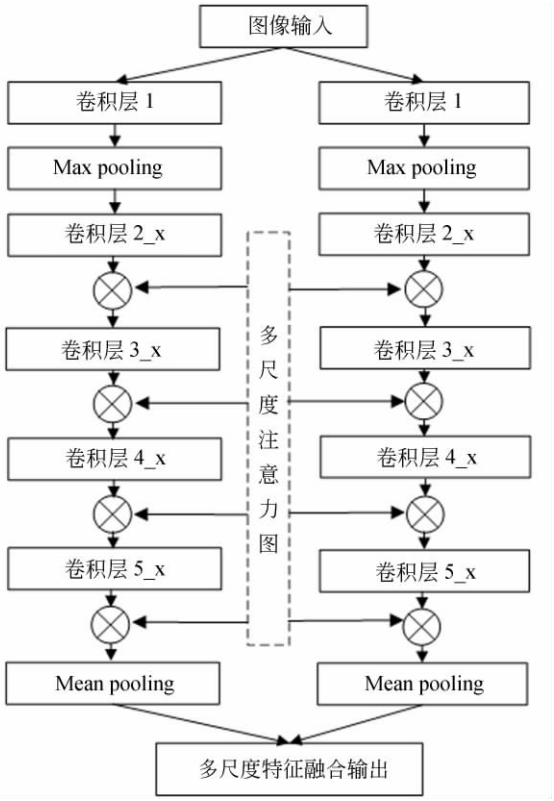


图 4　特征提取网络框架

Fig. 4　Framework of feature extraction network

3　循环神经网络多标签分类

3.1　多标签相关性分析

为了对航空图像多分类任务中各标签之间相关性的挖掘，本文利用 UCM 多标签数据集[12]进

行标签之间相关性的研究。UCM 数据集[13]由美国国家地质调查局(USGS)地图中心提供,数据集中包含 2 100 张航拍图片,一共 21 个类别。UCM 多标签数据集是在 UCM 数据集的基础之上,根据原始对象为每个图像样本分配一个或多个标签来重新标记组合而成。新定义的标签类总数为 17 个,包括:飞机、沙地、路面、建筑物、汽车、灌木丛、球场、树林、码头、储水罐、水域、草地、活动房屋、轮船、裸土、海洋和田野。表 2 为每个类别的图像数量分布。

表 2　UCM 多标签数据集每一类别图像数量

Tab. 2　Number of images per category in UCM multi-label dataset

类别名称	总数/张	类别名称	总数/张
Airplane	100	Mobile-home	102
Bare-soil	718	Pavement	1 300
Buildings	691	Sand	294
Cars	886	Sea	100
Chaparral	115	Ship	102
Court	105	Tanks	100
Dock	100	Trees	1 009
Field	104	Water	203
Grass	975		

UCM 多标签数据集样例如图 5 所示,其中场景(a)包含了飞机,裸地,汽车,草地,道路的标签。场景(b)包含了沙地、海洋的标签。场景(c)包含了裸地、网球场、草地、树木的标签。

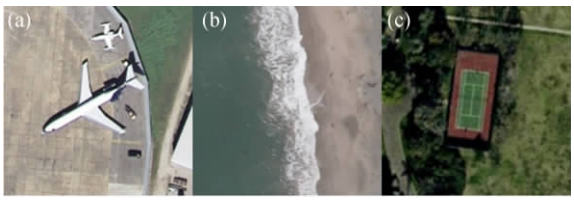


图 5　UCM 多标签数据集样例

Fig. 5　Sample of UCM multi-label dataset

在传统的研究当中,常见的假设是图像中的多个标签是相互独立的,模型分别对每个类别进行预测。通过计算类别之间的条件概率,可以发现实际的场景中,多个标签之间往往是相互关联的。条件概率的计算公式如式(5)所示:

$$P(C_p \mid C_r) = \frac{P(C_p, C_r)}{P(C_r)}, \tag{5}$$

其中:C_r 表示标签中的一个类别,C_p 表示潜在的共现类别。$P(C_r)$表示类别 C_r 出现的先验概率,$P(C_p, C_r)$表示类别 C_p,C_r 的联合概率,$P(C_p \mid C_r)$表示类别 C_p,C_r 的条件概率。UCM 多标签数据集条件概率矩阵如图 6 所示。

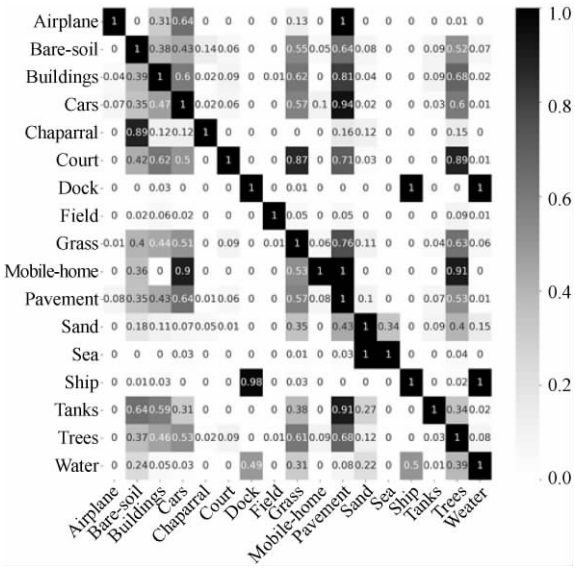


图 6　UCM 多标签数据集条件概率矩阵

Fig. 6　UCM multi-label dataset conditional probabilities matrix

从条件概率矩阵可以看出,$P(ship \mid dock)$的值为 0.98,表示出现船这一类别的图像中,同时出现码头的概率很高。$P(dock \mid ship)$的值为 1,表示出现码头这一类别的图像中出现船的概率也很高,也即这两个类别具有很高的相关度。因而可以采用卷积神经网络与循环神经网络相结合的方式,将特征提取网络输出的特征图向量化为W^2维的向量作为循环神经网络的输入,其中 W 为特征图的尺寸。

3.2　改进型长短期记忆(LSTM)网络

结合 UCM 多标签数据集条件概率矩阵可以看出,同一图像中不同的标签序列之间存在相关性。因此本文受文献[14]的启发,采用长短期记忆网络(Long Short-Term Memory,LSTM)[15]对标签之间的潜在相关性进行挖掘。其中,LSTM 网络通过门控单元将短期记忆与长期记

忆结合起来，被广泛应用于序列数据问题的处理。

在传统 LSTM 网络中，输入门输入的信息状态不能影响输出门的输出信息，并且遗忘门和输入门之间是相互独立的。本文在传统 LSTM 网络的基础上，增加输入门到输出门的连接，并且将遗忘门和输入门合并成一个单一的更新门，由原本的遗忘门和输入门分别决定信息的剔除和保留，变为遗忘门和输入门共同进行决策，以使输入状态更好地控制每一内存单元输出的信息，改进的 LSTM 网络结构如图 7 所示。

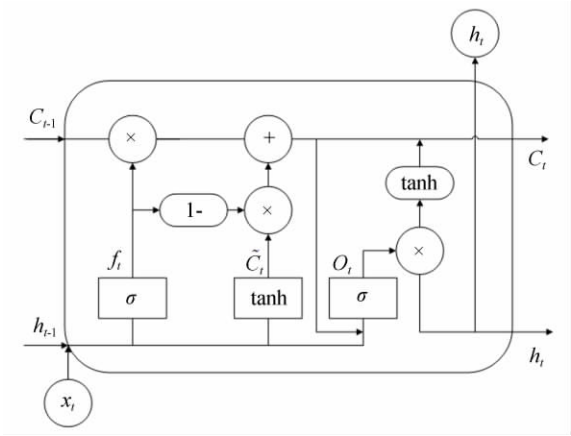


图 7　改进型 LSTM 单元

Fig. 7　Improvement of LSTM cell

3.2.1　改进型 LSTM 网络门限更新过程

改进的 LSTM 网络首先进行细胞状态中冗余信息的剔除，这个过程通过遗忘门的控制完成。公式(6)表示的是 LSTM 网络的遗忘门信息过滤过程：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{6}$$

其中：b 表示偏置，W 表示权重，σ 表示 sigmoid 函数变换，f_t 表示遗忘门选择输入的信息。遗忘门通过读取输入数据 x_t 和前一个状态的输出 h_{t-1}，输出一个在 0 到 1 之间的数值给每个细胞状态 C_{t-1}。其中 1 表示对信息完全保留，0 表示对信息完全舍弃。下一步需要确定存放在细胞状态中的信息。输入门由 sigmoid 函数决定需要更新的值，tanh 层创建新的候选值向量 $\tilde{C}_t$ 加入到状态中。式(7)表示信息经过 sigmoid 函数变换得到的输入信息，式(8)表示信息经过 tanh 变换得到的输入信息，式(9)表示最终输入门选择存入细胞状态中的信息：

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \tag{7}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \tag{8}$$

$$C_t = f_t \cdot C_{t-1} + (1 - f_t) \cdot \tilde{C}_t, \tag{9}$$

其中：tanh 表示 tanh 函数变换，i_t 表示输入信息经过 sigmoid 函数计算之后的输出值，$\tilde{C}_t$ 表示输入信息经过 tanh 函数变换之后的输出值，C_t 表示输入门最终输入的细胞状态。此时输入细胞的状态由 C_{t-1} 更新为 C_t，tanh 通过对细胞状态的处理得到介于 −1 和 1 之间的值，并将其和 sigmoid 函数的输出相乘，经过改进的 LSTM 网络的输出融入了细胞输入状态，最终获得确定的输出门信息如公式(10)所示，LSTM 单元的最终输出如公式(11)所示：

$$o_t = \sigma(W_o[C_t, h_{t-1}, x_t] + b_o), \tag{10}$$

$$h_t = o_t \cdot \tanh(C_t), \tag{11}$$

其中：o_t 表示输出门输出的信息，h_t 表示该 LSTM 单元的最终输出。

考虑到多标签图像中，类与类之间的依赖关系是双向的，而基于单向的 LSTM 网络不足以全面描述标签类别间的关系。因此本文在模型中采用 BiLSTM 网络对多标签序列进行学习。BiLSTM 网络结构如图 8 所示。

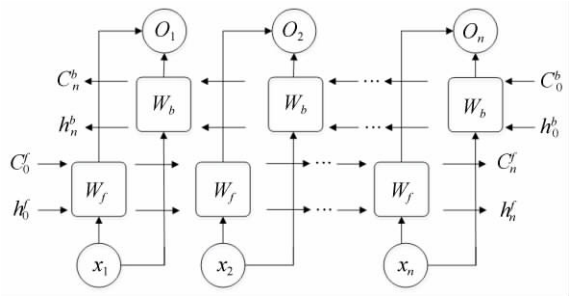


图 8　BiLSTM 网络结构

Fig. 8　Structure of bidirectional LSTM network

4　实验与结果分析

4.1　模型训练

为了验证本文方法的有效性，本文设计了实验进行验证。首先，对特征提取网络中多尺度注意力特征图尺度的选择上，大尺度的图像包含的

信息丰富，细节清晰，但是会增加模型的计算量。而小尺度的图像虽然计算量较小，但是缺乏细节。本文以224×224的图像作为基准，选定尺度0.8，1，1.3，2，分别得到179×179，256×256，291×291，448×448不同大小尺度的图像，通过不同组合交叉验证，得到的最优尺度组合为1、1.3、2。同时，为了使模型能够充分收敛，本文在原始图像数据集上进行了数据增强，对原始图像进行缩放、旋转、模糊等数据增强方式，对原始数据集进行了10倍地有效扩充。最终训练图片的总数量达到了20 000张。实验中将数据集按照8∶2的比例随机划分为训练集和测试集。

本文的实验在Linux环境下进行，采用的GPU为英伟达K80，在基于Tensorflow的深度学习框架进行模型训练。模型训练过程中采用动量梯度下降法(Momentum)，动量momentum=0.9，初始学习率lr=0.01，并且当测试集精度饱和时进行0.9倍的权值衰减，进行80 000次迭代的训练。训练的损失函数定义为均方误差损失函数。训练过程的损失曲线如图9所示。

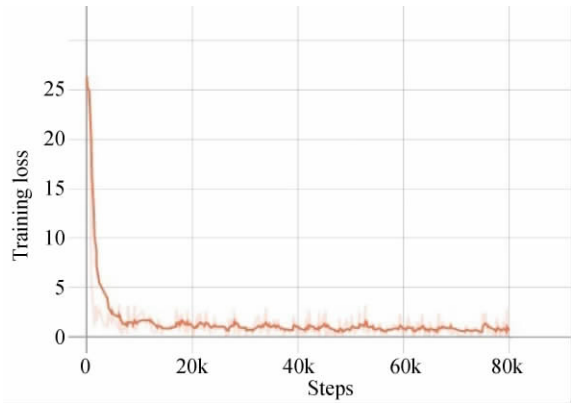


图9 网络训练损失曲线图

Fig. 9 Network training loss curve

4.2 实验结果与分析

为了对实验结果进行分析，全面评估模型的性能，本文通过计算精确率，召回率和$F1$值来评判模型的性能。精确率表示的是在所有被预测为正的样本中实际为正样本的概率，代表了正样本结果中预测的准确程度，精确率的公式如式(12)：

$$precision = \frac{TP}{TP + FP}, \tag{12}$$

其中：TP表示模型判断为正样本，实际为正样本。FP表示模型判断为正样本，实际为负样本。召回率表示的是实际为正的样本中被预测为正样本的概率。召回率公式如式(13)：

$$recall = \frac{TP}{TP + FN}, \tag{13}$$

其中，FN表示模型判断为负样本，实际为负样本。引入$F1$-Score作为综合评价模型的指标。$F1$值是精确率和召回率的调和平均，$F1$值的公式如式(14)：

$$F1 = \frac{2}{\frac{1}{recall} + \frac{1}{precision}}. \tag{14}$$

为了验证本文算法的有效性，本文在UCM多标签数据集上进行了对比试验。分别使用在ImageNet图像上训练的VGGNet16[16]和ResNet50作为预训练模型。其中方法1表示以ResNet50结合多尺度注意力机制，采用未改进的BiLSTM网络对多标签数据进行分类实验。本文方法采用预训练ResNet50与多尺度注意力机制相结合的方式对图像特征进行提取，结合改进型BiLSTM网络挖掘UCM多标签数据集标签之间相关性信息。对比试验结果如表3所示。

表3 UCM多标签数据集实验结果

Tab. 3 UCM multi-label dataset experimental result

模型	精确率/%	召回率/%	F1值
VGGNet16	78.08	78.11	0.781
ResNet50	80.69	82.29	0.815
文献[14]	78.08	86.35	0.82
方法1	82.22	83.17	0.827
本文方法	85.33	87.05	0.862

从实验结果可以看出，本文方法的分类性能相比于标准的卷积神经网络均有所提升。同时，本文对双向长短期记忆网络进行了改进，对比于方法1采用的传统长短期记忆网络的方法，模型的精确率提高了3.11%，模型的召回率提高了3.88%，验证了对传统BiLSTM网络改进的有效性。本文实验的部分分类预测结果如表4所示。

表 4　UCM 多标签数据集测试样例结果

Tab. 4　Example predictions on UCM multi-label dataset

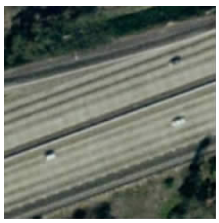
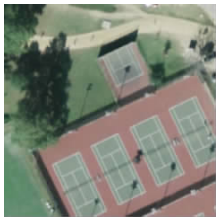
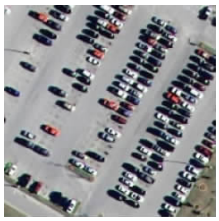
UCM 多标签数据集	真实值	VGGNet16	ResNet50	方法 1	本文方法
	Dock Ship Water	[Dock] Ship Water	[Dock] Ship Water	Dock Ship Water (Sea)	Dock Ship Water
	Cars Grass Trees Pavement	Cars [Grass] (Sand) (Chaparral) Trees Pavement	Cars Grass (Bare-soil) (Chaparral) Trees Pavement	Cars Grass (Chaparral) Trees Pavement	Cars Grass (Chaparral) Trees Pavement
	Court Pavement Grass Trees	Court Pavement Grass (Sand) Trees	Court Pavement Grass (Sand) Trees	Court Pavement Grass Trees	Court Pavement Grass Trees
	Cars Grass Pavement	Cars (Bare-soil) [Grass] Pavement	Cars (Bare-soil) (Trees) [Grass] Pavement	Cars [Grass] (Trees) Pavement	Cars [Grass] Pavement

表 4 中，小括号中的值代表假正例，表示原图中不存在而模型误测出的类别，即误识。中括号中的值代表假负例，表示图像中存在而模型没有预测出的类别，即漏识。

从表 4 可以看出，本文模型可以检测出多标签航空图像中绝大部分的标签类别，但在个别情形下发生了漏识或误识的情况。究其原因，在类别特征区分度不高，多个相似类别的特征混杂在一起的情况下，则可能会导致误识的发生。例如表 4 中第 2 张测试图像，“Grass”，“Chaparral”，“Trees”的图像特征具有一定相似性，并且“Chaparral”这一类别的物体在图像中特征区分度较弱，因此所有模型都对“Chaparral”这一类别产生了误识。同时，对于图像中目标较小的物体，由于物体特征不明显，因此导致漏识情况的发生。例如表 4 中最后一张图像，“Grass”这一类别的物体在图像中占比较小，所有模型都对“Grass”这一类别产生了漏识。但同时也可以看出，本文模型的多标签分类方法相比于传统方法，误识和漏识的情况得到了明显地改善，识别准确率有了较大提升，体现出了本文模型良好的多标签分类性能。

5　结　论

针对航空图像多标签分类问题，本文提出了一种基于循环神经网络的航空图像多标签分类方法。通过将多尺度注意力图嵌入预训练的 ResNet50 卷积神经网络中对图像进行特征提取。在模型层中加入批归一化层，提高模型的泛化能力。针对标签序列，采用改进的 BiLSTM 网络进行序

列特征提取，增加了BiLSTM网络输入门到输出门之间的连接，使得输入状态更好地控制每一内存单元输出的信息。通过这种设计，模型输出是结构化的序列，而不是离散的值。最后，本文在UCM多标签数据集上进行了实验。实验结果表明，本文方法相比于VGGNet16模型，精确率提高了7.25%，召回率提高了8.94%。相比于ResNet50模型，精确率提高了4.64%，召回率提高了4.76%。相比于传统的卷积神经网络模型，本文方法表现出了良好的多标签分类性能。但对于图像中类别特征不明显或特征之间区分度不高的情况下，本文模型也会存在漏识或误识的情况，如何针对性地解决这一问题将是本文下一步的研究重点。

参考文献：

[1] 郑远攀，李广阳，李晔. 深度学习在图像识别中的应用研究综述[J]. 计算机工程与应用，2019，55(12):20-36.
ZHENG Y P, LI G Y, LI Y. Survey of application of deep learning in image recognition[J]. *Computer Engineering and Applications*, 2019, 55(12):20-36. (in Chinese)

[2] 李晓斌，江碧涛，王生进. 光学遥感图像场景分类技术综述和比较[J]. 无线电工程，2019，49(4)：265-271.
LI X B, JIAN B T, WANG SH J. A review and comparison of optical remote sensing scene classification[J]. *Radio Engineering*, 2019, 49(4): 265-271. (in Chinese)

[3] 邓凌云. 遥感图像分类中的遗传算法LVQ神经网络运用[J]. 现代电子技术，2020，43(1)：40-43.
DENG L Y. Application of LVQ neural network in remote sensing image classification[J]. *Modern Electronics Technique*, 2020, 43(1):40-43. (in Chinese)

[4] 杨州，慕晓冬，王舒洋，等. 基于多尺度特征融合的遥感图像场景分类[J]. 光学 精密工程，2018，26(12)：232-240.
YANG ZH, MU X D, WANG SH Y, *et al*.. Scene classification of remote sensing images based on multi-scale features fusion [J]. *Opt. Precision Eng.*, 2018, 26(12): 232-240. (in Chinese)

[5] 李宇，刘雪莹，张洪群，等. 基于卷积神经网络的光学遥感图像检索[J]. 光学 精密工程，2018，26(1)：200-207.
LI Y, LIU X Y, ZHANG H Q, *et al*.. Optical remote sensing image retrieval based on convolution neural network[J]. *Opt. Precision Eng.*, 2018, 26(1): 200-207. (in Chinese)

[6] ZEGGADA A, BENBRAIKA S, MELGANI F, *et al*.. Multi-label conditional random field classification for UAV images[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(3): 399-403.

[7] KODA S, ZEGGADA A, MELGANI F, *et al*.. Spatial and structured SVM for multi-label image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(10): 5948-5960.

[8] 边小勇，费雄君，穆楠. 基于尺度注意力网络的遥感图像场景分类[J]. 计算机应用，2020，40(3)：872-877.
BIAN X Y, FEI X J, MU N. Remote sensing image scene classification based scale-attention network [J]. *Journal of Computer Applications*, 2020, 40(3):872-877. (in Chinese)

[9] ACHANTA R, SHAJI A, SMITH K, *et al*.. SLIC superpixel compared to state-of-the-art superpixel methods[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 2274-2282.

[10] 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*, 2016: 770-778.

[11] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [J]. *Arxiv Preprint* Arxiv: 1502.03167, 2015.

[12] CHAUDHURI B, DEMIR B, CHAUDHURI S, *et al*.. Multi-label remote sensing image retrieval using a semi-supervised graph-theoretic method [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(2):1144-1158.

[13] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]. *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. *ACM*, 2010: 270-279.

[14] HUA Y, MOU L, ZHU X X. Recurrently exploring class-wise attention in a hybrid convolutional and bidirectional LSTM network for multi-label aerial image

classification[J]. *ISPRS journal of photogrammetry and remote sensing*, 2019, 149: 188-199.

[15] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. *Neural computation*, 1997, 9(8): 1735-1780.

[16] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. *Arxiv Preprint Arxiv*: 1409.1556, 2014.

作者简介：

陈科峻(1993－)，男，广西南宁人，硕士研究生，2016 年于哈尔滨工业大学获得学士学位，主要从事计算机视觉，模式识别，机器学习方面的研究。E-mail：ckj409399@sina.com

通讯作者：

张　叶(1982－)，女，吉林长春人，研究员，博士生导师，吉林大学学士，中国科学院长春光学精密机械与物理研究所博士，主要从事计算机视觉，模式识别，机器学习等方面的研究。E-mail：yolanda@spirit.ai