第 58 卷 第 4 期/2021 年 2 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于卷积特征和贝叶斯决策的双波段场景分类

邱晓华[1,2*]，李敏[1**]，张丽琼[1]，董琳[2]

[1] 火箭军工程大学作战保障学院，陕西 西安 710025；

[2] 中国人民武装警察部队工程大学信息工程学院，陕西 西安 710086

摘要 针对可见光和近红外双波段场景分类存在图像标注样本少和特征融合质量低的问题，提出了一种基于卷积神经网络(CNN)特征提取和朴素贝叶斯决策融合的双波段场景分类方法。首先，将基于预训练的 CNN 模型作为双波段图像的特征提取器，避免标注样本少导致的过拟合问题；然后，通过主成分分析降维和特征归一化方法，提高支持向量机的计算速度和每个波段的分类性能；最后，以双波段后验概率为朴素贝叶斯先验概率，构建了决策融合模型，实现场景融合分类。在公开数据集上的实验结果表明，相比单一波段分类和双波段特征级联融合分类方法，本方法的识别率有明显提升，可达到 94.3%；比基于传统特征的最优方法高 6.4 个百分点，与基于 CNN 的方法识别率相近，且执行简单高效。

关键词 机器视觉；图像分类；朴素贝叶斯模型；双波段场景；卷积神经网络；决策融合

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.0415006

Dual-Band Scene Classification Based on Convolutional Features and Bayesian Decision

Qiu Xiaohua[1,2*], Li Min[1**], Zhang Liqiong[1], Dong Lin[2]

[1] *College of Operational Support, The Rocket Force University of Engineering, Xi'an, Shaanxi* 710025, *China*;

[2] *College of Information Engineering, Engineering University of PAP, Xi'an, Shaanxi* 710086, *China*

Abstract Aiming at the problems of few labeled samples and low quality of feature fusion in visible and near infrared dual-band scene classification, a dual-band scene classification method based on convolutional neural network (CNN) feature extraction and naive Bayes decision fusion is proposed in this paper. First, the CNN model based on pre training is used as the feature extractor of dual-band image to avoid the over fitting problem caused by few labeled samples. Second, the calculation speed of support vector machine and the classification performance of each band are improved by the dimensionality reduction of principal component analysis and feature normalization method. Finally, using the dual band posterior probability as the naive Bayes prior probability, a decision fusion model is constructed to achieve scene fusion classification. Experimental results on the public dataset show that compared with single-band classification and dual-band feature cascade fusion classification methods, the recognition rate of the method is significantly improved, reaching 94.3%; it is 6.4 percentage points higher than the best method based on traditional features. The recognition rate is similar to the CNN-based method, and the execution is simple and efficient.

Key words machine vision; image classification; naive Bayesian model; dual-band scene; convolutional neural network; decision fusion

OCIS codes 150.0155; 100.4993; 110.4234; 100.3008

收稿日期：2020-06-30；**修回日期**：2020-08-03；**录用日期**：2020-08-12

基金项目：国家自然科学基金(61102170)

* **E-mail**：qxh_1025@163.com；** **E-mail**：proflimin@163.com

1 引 言

随着深度学习技术的广泛研究，凭借海量标注样本的优势，面向可见光波段的图像分类、目标检测等计算机视觉任务取得了突飞猛进的发展，产生了一些经典的卷积神经网络（CNN）模型，如应用于图像分类的 AlexNet[1]、视觉几何组网络（VGGNet）[2]、GoogLeNet[3]、深度残差网络（ResNet）[4]模型。红外图像可为同一场景可见光图像提供补充信息，因此，融合可见光和红外信息的双波段图像比单一波段图像的优势更大。近年来，双波段图像结合深度学习技术，逐渐成为图像分类[5]、目标识别[6]、行人检测[7]、目标检测[8]及目标跟踪[9]等领域的研究热点。

场景分类是计算机视觉领域长期研究的一个主题[10]。目前，基于可见光 RGB（Red，Green，Blue）和近红外（NIR）双波段图像的场景分类方法主要包括基于传统特征的方法和基于 CNN 的方法。基于传统特征的方法主要利用尺度不变特征变换（SIFT）等人工设计特征，如 Brown 等[11]提出了一种基于多光谱 SIFT（MSIFT）的场景分类方法，该方法对双波段图像进行去相关处理，提取每个通道的 SIFT 特征进行级联融合，并用主成分分析（PCA）方法降低融合特征维度。Salamati 等[12]利用 Fisher Vector 方法融合 SIFT 特征和颜色信息，Xiao 等[13]使用直方图统计变换方法对多光谱的梯度信息和颜色信息进行联合编码。张秋实等[14]提取双波段图像的密集 SIFT（DSIFT）特征，采用无字典模型（CLM）进行特征变换，然后基于混合核的支持向量机（SVM）进行分类。

基于 CNN 的方法通常采取预训练或微调的经典网络模型提取卷积特征，通过设计双波段图像的特征级联融合网络，学习训练共同的特征表示。Ševo 等[15]通过组合基于 GoogLeNet 模型的两个子网络和三个分类器，设计了一种双 CNN（Dual CNN）体系架构，两个子网络分别以 RGB 图像和 NIR＋RGB 组合通道图像作为输入。Peng 等[16]通过微调预训练 GoogLeNet 模型提取双波段图像的卷积特征，并采用基于核函数的主成分分析（KPCA）和典型相关分析（CCA）方法进行特征降维和特征融合。江泽涛等[17]使用预训练 ResNet-50 模型提取图像特征，进行级联融合后送入全连接层，然后进行训练和分类。Jiang 等[5]利用全连接层融合基于简单 CNN 的 RGB 和 NIR 两路输出特征，构建并训练双路特征融合模型，取得了较好的分类性能。此外，由于双波段图像缺乏大量的标注样本，数据增强、Dropout 等正则化技术是这类方法训练中常用的技巧，可避免模型出现过拟合问题。

基于传统特征的方法通过词袋模型、Fisher Vector、直方图统计变换等方法对多通道自然场景的结构、纹理、颜色等视觉信息进行联合编码。而 CNN 通过自动学习，将低级特征由底层到高层逐步抽象为高级语义特征[18]，相比传统方法在图像表示上有明显优势。双波段图像的卷积特征不仅包含互补信息，还包括大量冗余信息，导致级联融合的特征质量不高，而学习共同的判别性特征表示仍是目前基于双波段图像计算机视觉任务的一个难点。在双波段图像标注样本匮乏的情况下，基于 CNN 的方法克服了人工设计特征表示能力不足的问题，但模型训练易出现过拟合、级联融合质量低以及共同特征表示学习难等问题。

近年来，单一波段多分类器融合[19-20]的决策级融合在双波段图像目标识别[21]、目标检测[22]领域中得到广泛的应用。为解决双波段图像场景分类中存在的问题，本文利用决策级融合简单快速的优势，提出了一种基于预训练网络模型卷积特征和朴素贝叶斯模型（Naive Bayes model）决策融合的双波段图像场景分类方法，为双波段图像场景分类提供了一种新的融合思路。

2 融合模型

本方法首先利用基于 ImageNet 数据集预训练的 CNN 模型（以 VGG-16 为例）分别提取双波段图像的卷积特征，避免因标注数据集匮乏引起的深度网络模型训练过拟合问题。其次，采用 PCA 方法进行特征选择，并降低卷积特征维度，避免高维度卷积特征占用存储和计算资源多的问题。然后，基于 SVM 分类器计算每个波段图像分类的后验概率。最后，通过朴素贝叶斯模型融合后验概率，输出双波段图像的分类识别标签。算法的总体框架如图 1 所示，其中，VGG-16 模型包含 5 个卷积块（C1～C5）和 3 个全连接层（F6～F8），每个卷积块包含卷积层（conv）和池化层（pool）。

2.1 卷积特征提取

深度学习区别于传统机器学习方法的最主要特点在于表示学习部分具有分布式表示特性[23]。此

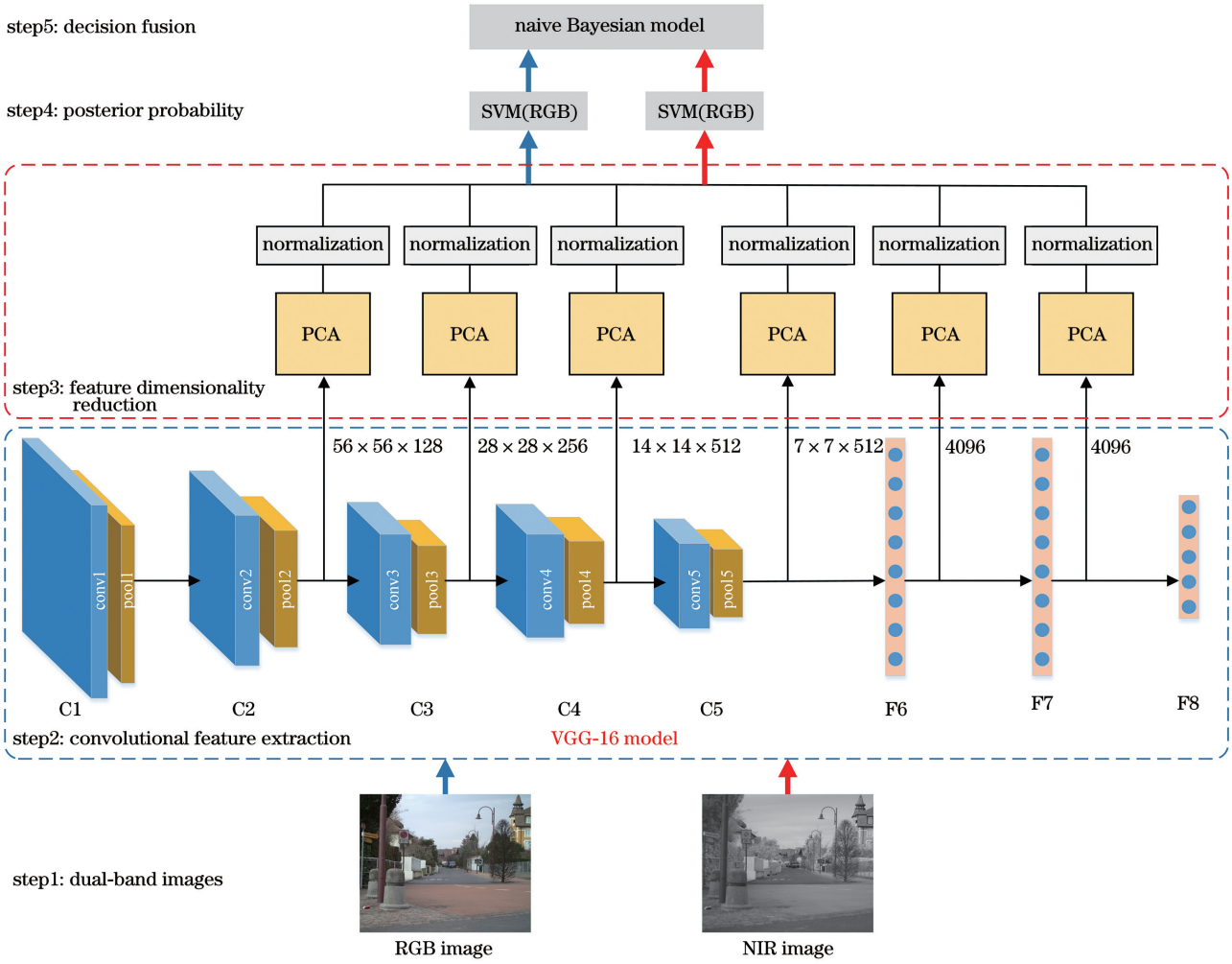

图 1　本方法的框架图

Fig. 1　Framework diagram of our method

外，AlexNet、VGGNet 等深度 CNN 通过增加网络层数，将底层的视觉特征逐步抽象为高层的语义特征。因此，整个网络体系架构由下自上呈现出特征表示的层次性[24]。采用基于 ImageNet 数据集预训练的 VGGNet 和 ResNet 模型，并根据网络结构特征表示的层次性，依次提取 RGB 和 NIR 双波段图像的低级、中级和高级卷积特征，如图 1 中卷积块 C1 和 C2 提取的特征为低级特征，卷积块 C3～C5 提取的特征为中级特征，全连接层 F6～F8 提取的特征为高级特征。由于卷积块 C2 能代表低级特征，且最后一个全连接层的特征表示为特定数据集的语义模式，因此，不提取预训练网络模型的卷积块 C1 和最后一个全连接层 F8 的特征。表 1 为 VGGNet 和 ResNet 模型特征提取对应的网络层及特征大小，其中，ResNet 的高级层 G6 为全局平均池化层(Global average pooling layer)。可以发现，对于每个输入样本，可得到大小为 $r\times r\times K$ 的卷积特征。其中，$r\times r$ 为卷积核大小，K 为卷积核的数量。全连接层和全局平均池化层的特征可看成大小为 $1\times1\times K$ 的卷积特征。同时，将卷积特征转化为一维特征向量 $\boldsymbol{f}_i\in\boldsymbol{R}^{r^2K}$ $(i=1,2,\cdots,n)$，其中，n 为样本数。

表 1　VGGNet 和 ResNet 的网络层及特征维度

Table 1　Layers and feature dimension of the VGGNet and ResNet

Hierarchical feature	Low level	Middle level			High level	
	C2	C3	C4	C5	F6(G6)	F7
VGGNet	56×56×128	28×28×256	14×14×512	7×7×512	4096	4096
ResNet-50	56×56×256	28×28×512	14×14×1024	7×7×2048	2048	–

2.2 特征降维与归一化设计

提取的高维度卷积特征消耗的计算资源较多，且含有一定的噪声和较多的冗余信息，因此，通过 PCA 将卷积特征从高维空间投影到低维空间。PCA 是机器学习中最常用的降维方法，主要思想是通过对高维特征的协方差矩阵进行特征值分解，保留前几个最大特征值对应的特征向量，从而构成低维特征空间。PCA 可以通过设置固定维度和设置重构阈值[25]两种方法获取低维特征空间的维度，其中，设置重构阈值的数学理论是方差最大化理论。在信号处理中，通常认为信号的方差较大，噪声的方差较小，信噪比就是信号与噪声的方差比，即信噪比越大越好。PCA 也可看成逐一选取方差最大方向，即通过 PCA 投影后，保留方差较大的前几个特征向量，该做法与保留前几个最大特征值得到的特征向量等价。重构阈值的实质是低维空间方差和与高维空间方差和的百分比。在协方差矩阵中，方差可经奇异值或特征值解释，因此，重构阈值可表示为

$$\frac{\sum_{d=1}^{d_{\text{low}}}\lambda_d}{\sum_{d=1}^{d_{\text{high}}}\lambda_d}\geqslant t, \tag{1}$$

式中，λ_d 为高维特征协方差矩阵的第 d 个特征值，$d_{\text{high}}=r^2K$ 为高维空间的维数，d_{low} 为低维空间的维数，t 为重构阈值。可以发现，t 越大，d_{low} 也越大。

实验主要研究双波段场景图像数据集，不同波段具有不同的成像特性，且提取的特征为预训练深度 CNN 模型不同层的卷积特征。此外，数据集的样本数远小于卷积特征的维数，因此，从特征重构的角度出发，通过设置重构阈值计算低维特征空间的维度。重构阈值的大小直接影响了特征重构的质量，进而影响后续 SVM 的分类精度。

利用 SVM 计算后验概率前，采用 L_2 范数归一化处理低维特征空间的每个样本，有利于计算两个样本之间的距离相似度。L_2 范数归一化方法首先计算每个样本的 L_2 范数，然后将该样本中的元素除以该范数，归一化处理的目的是使每个样本的 L_2 范数为 1。令 $\boldsymbol{f}_i'=(f_1',f_2',\cdots,f_{d_{\text{low}}}')^{\mathrm{T}}$ 为低维空间第 i 个样本的特征向量，$\boldsymbol{f}_i''=(f_1'',f_2'',\cdots,f_{d_{\text{low}}}'')^{\mathrm{T}}$ 为第 i 个样本经 L_2 范数归一化处理后的特征向量，则 $\boldsymbol{f}_i''$ 中第 j 个元素 f_j'' 可表示为

$$f_j''=\frac{f_j'}{(|f_1'|^2+|f_2'|^2+\cdots+|f_{d_{\text{low}}}'|^2)^{\frac{1}{2}}},\quad j=1,2,\cdots,d_{\text{low}}。 \tag{2}$$

2.3 贝叶斯决策融合模型构建

SVM 分类器通常产生模式识别中的类别标签，通过拟合 Sigmoid 模型的方法，可将 SVM 的无阈值输出转换为后验概率输出[26]。本方法采用基于线性核的 SVM 分别计算双波段图像中每个样本的后验概率，并通过朴素贝叶斯模型进行融合分类，得到双波段图像共同的分类标签。朴素贝叶斯模型是一种基于贝叶斯理论和条件独立性假设的分类方法。本方法将 SVM 分类器输出的后验概率作为先验概率，通过计算条件概率获得融合分类的后验概率，从而构建朴素贝叶斯决策融合模型，其后验概率的计算与分类过程如下。

假设 $\omega_k(k=1,2,\cdots,c)$ 为双波段图像数据集的样本类别，$\boldsymbol{S}=\{s_1,s_2\}$，$s_1$ 和 s_2 为双波段图像对应的两个相互独立的 SVM 分类器，$P(s_m)$ 为第 m 个 SVM 分类器将样本 $\boldsymbol{x}$ 标记为所有类别的后验概率。依据条件独立性假设，得到条件概率 $P(\boldsymbol{S}|\omega_k)$ 为

$$P(\boldsymbol{S}|\omega_k)=P(s_1,s_2|\omega_k)=P(s_1|\omega_k)P(s_2|\omega_k)。 \tag{3}$$

通过先验概率 $P(\omega_k)$ 和条件概率 $P(\boldsymbol{S}|\omega_k)$ 计算用决策融合模型标记样本 $\boldsymbol{x}$ 的后验概率 $P(\omega_k|\boldsymbol{S})$，可表示为

$$P(\omega_k|\boldsymbol{S})=\frac{P(\omega_k)P(\boldsymbol{S}|\omega_k)}{P(\boldsymbol{S})}=\frac{P(\omega_k)P(s_1|\omega_k)P(s_2|\omega_k)}{P(\boldsymbol{S})}, \tag{4}$$

式中，$P(\boldsymbol{S})$ 为 s_1 和 s_2 的联合概率，与 ω_k 无关，可以忽略。则样本 $\boldsymbol{x}$ 对类别 ω_k 的支持 $\mu_k(\boldsymbol{x})$ 可表示为

$$\mu_k(\boldsymbol{x})\propto P(\omega_k)P(s_1|\omega_k)P(s_2|\omega_k), \tag{5}$$

式中，$\propto$ 为正比符号，样本 $\boldsymbol{x}$ 的最终类别为 $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_c)$ 中最大值对应的类别。贝叶斯决策融合方法在样本数为 N 的数据集上的具体实现：两个波段的分类器 s_1 和 s_2 通过测试样本集计算，分别获得一个 $c\times c$ 的混淆矩阵 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$。每个 $\boldsymbol{C}$ 的第 (k,s) 个元素 $c_{k,s}^m$ 为数据集中真实类别标签 ω_k 被 SVM 分类器判别为类别标签 ω_s 的样本个数。假设 N_k 为数据集中类别为 ω_k 的总样本数，$c_{k,s}^m/N_k$ 为后验概率估计，N_k/N 为先验概率估计，则样本 $\boldsymbol{x}$ 对类别 ω_k 的支持 $\mu_k(\boldsymbol{x})$ 可表示为

$$\mu_k(\boldsymbol{x})\propto\frac{1}{N_k}(c_{k,s_1}^1\times c_{k,s_2}^2)。 \tag{6}$$

根据 $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_c)$ 的最大值规则，将样本 $\boldsymbol{x}$ 标记为类别 ω_k。

3　实验与分析

3.1　数据集与实验平台

实验采用的验证数据集是唯一公开的 RGB-NIR 双波段自然场景基准数据集[11]，该数据集包含 477 对 RGB-NIR 图像，包括乡村(52)、田野(51)、森林(53)、室内(56)、山峰(55)、古老建筑(51)、街道(50)、城市(58)、水域(51)9 类，括号中为每类样本的数量。虽然数据集的图像数量有限，但包含相互干扰、具有挑战性的类别，如乡村与田野、街道与城市。图 2 为 RGB-NIR 数据集中的部分图像对，按文献[10]的训练和测试样本设置方法，随机选取 99 对图像(每类 11 对)作为测试集，其余图像用于训练。同时，随机选择 20 组训练/测试组进行实验，并以分类精度的平均值(M)和均方差(S)评估本方法法的分类性能。

图 2　RGB-NIR 数据集的示例图像

Fig. 2　Example image of the RGB-NIR dataset

仿真验证的硬件平台：处理器为 2.8 GHz 英特尔 Core i7-7700HQ，内存为 16 GB，显卡为 NVIDIA GeForce 940MX。软件环境：系统为 Windows10，集成开发环境为 PyCharm 2.4，PCA 和线性 SVM 算法的实现采用基于 Python 语言 sklearn 库集成 PCA 模块和 SVM 模块中核函数为“linear”的 SVC 模块，PCA 和 SVM 的关键参数分别为维度因子 n_{com} 和惩罚因子 C，其中，n_{com} 的取值为(0，1)，$C=1$ 为 SVC 模块的默认参数。贝叶斯决策融合模型中，$\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 通过基于 Python 语言 sklearn 库集成 confusion_matrix 模块计算获得。深度学习框架采用前端 Keras 和后端 TensorFlow 平台，深度 CNN 模型为基于 ImageNet 预训练的 VGG-16、VGG-19 和 ResNet-50 模型。

3.2　实验结果与分析

1) PCA 降维算法的性能评估

线性核 SVM 算法的时间复杂度与输入样本的特征维度成线性关系，因此降低输入样本的特征维度可加快分类速度。以重构阈值 t 为 0.99 的 PCA 方法对 CNN 特征进行降维，通过对比 CNN 原始特征和 PCA 降维特征的大小与分类精度，评估 PCA 降维方法的性能。表 2 为 VGG-16 模型各层 CNN 特征和 PCA 特征的维度，图 3 为基于 VGG-16 模型各层的 CNN 特征和 PCA 特征的分类精度，可以发现，由于数据集样本少，即使设置最大的重构阈值，PCA 特征也远远小于 CNN 特征的维度。且 PCA 算法虽然大幅降低了 CNN 特征的维度，但对分类精度的降低并不明显，在高层网络中的分类精度还有所提高，这表明采用 PCA 方法对 CNN 特征进行降维处理是有效的。

表 2　VGG-16 模型不同特征的维度

Table 2　Dimensions of different features of the VGG-16 model

Layer	C2	C3	C4	C5	F6	F7
CNN feature	401408	200704	100352	25088	4096	4096
PCA feature of RGB	359	360	361	344	328	312
PCA feature of NIR	361	362	361	350	338	322

2) 不同重构阈值的分类性能

为了分析不同重构阈值 t 对算法分类性能的影响，将基于 1 组重构阈值的 PCA 方法作用于 VGG-16 模型的各层卷积特征，分析基于单波段和双波段的场景分类精度。图 4 为 1 组阈值 t(0.1，0.2，0.3，0.4，0.5，0.6，0.7，0.8，0.9，0.99)在基于中高级 C5、F6 和 F7 层卷积特征的分类精度。可以发

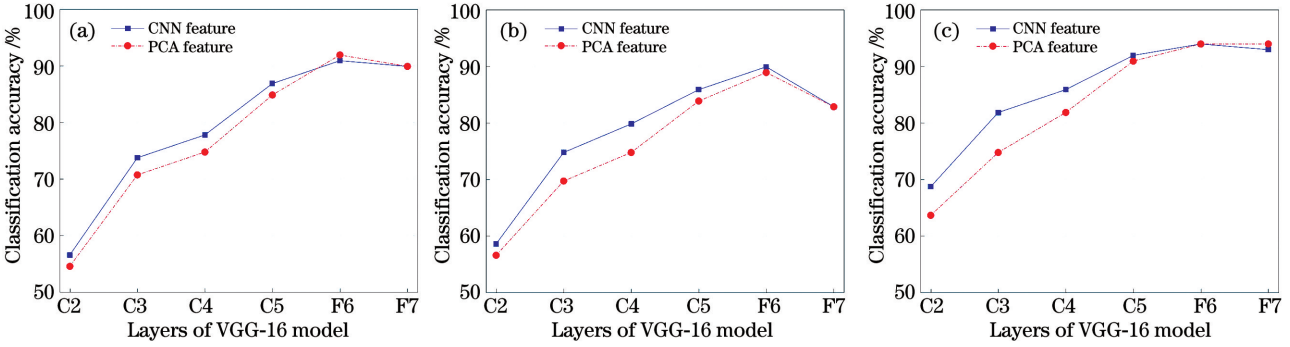

图 3　两种特征的分类精度。(a)RGB 图像;(b)NIR 图像;(c)RGB-NIR 图像

Fig. 3　Classification accuracies of the two features. (a) RGB image; (b) NIR image; (c) RGB-NIR image

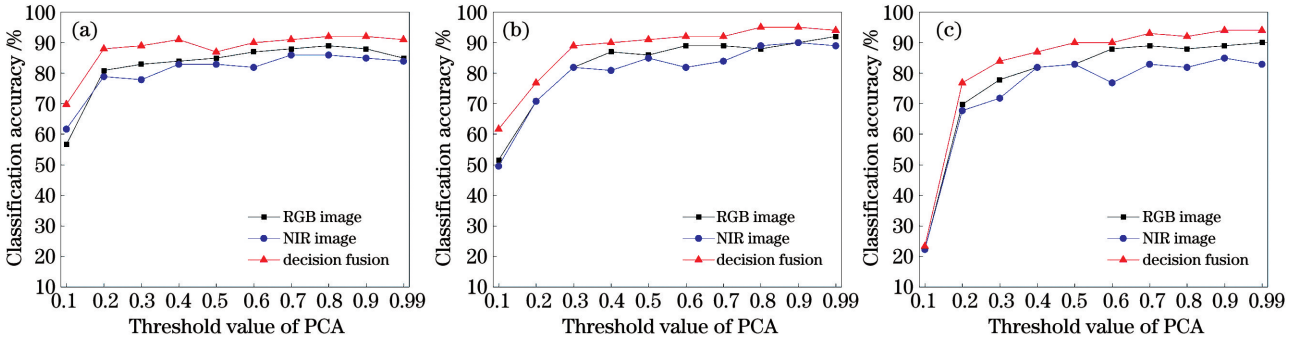

图 4　不同阈值对模型分类精度的影响。(a)C5 层;(b)F6 层;(c)F7 层

Fig. 4　Influence of the different threshold value on model classification accuracy. (a) C5 layer; (b) F6 layer; (c) F7 layer

现,随着重构阈值的不断增加,融合分类精度逐步提高,当阈值 t 的取值大于 0.5 时,融合分类精度趋于平稳。鉴于数据集样本少,重构阈值 t 的取值为 0.9～0.99。此外,无论阈值如何变化,双波段的决策融合分类精度均比 RGB 波段的分类精度高 4%～8%,比 NIR 波段的分类精度高 6%～10%。

3) 基于不同 CNN 模型的分类性能

图 5 为基于不同 CNN 模型的融合分类精度(t 取 0.9),其中,feature fusion 为将 RGB 和 NIR 图像的特征进行级联融合。可以发现,本方法在不同 CNN 模型和不同卷积层的融合分类精度均高于单波段的分类精度,而 feature fusion 并没有明显提升单波段的分类精度,甚至低于 RGB 图像的分类精度。此外,基于相同 CNN 模型的融合分类精度由底层到高层逐渐提高,原因是每个波段的分类性能与同一 CNN 模型不同卷积层的特征表示能力相关,预训练 CNN 模型的特征表示能力由底层到高层逐渐增强。在同一级卷积层中,每个波段的分类精度与 CNN 模型自身的性能相关,导致不同 CNN 模型的融合分类精度存在差异。不同重构阈值得到不同 CNN 模型的最佳融合分类精度如表 3 所示,可以发现,VGG-16 模型在 F7 层和 t 取 0.95 时的分类精度最高,为(93.3±2.0)%;VGG-19 模型在 F6 层和 t 为 0.99 时的分

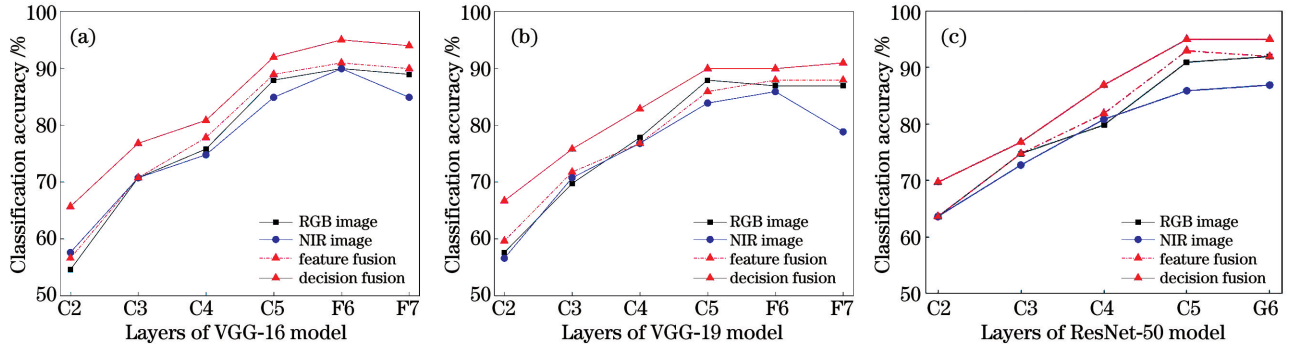

图 5　不同 CNN 模型的分类精度。(a)VGG-16 模型;(b)VGG-19 模型;(c)ResNet-50 模型

Fig. 5　Classification accuracies of different CNN models. (a) VGG-16 model; (b) VGG-19 model; (c) ResNet-50 model

类精度最高，为(92.0±2.5)%；而 ResNet-50 模型在 G6 层和 t 为 0.99 时的分类精度最高，为(94.3±2.1)%。对比不同 CNN 模型和不同层，可以发现，CNN 模型的高层均取得最佳值，且 ResNet-50 模型的分类性能优于 VGG-16 模型和 VGG-19 模型。

表 3 不同 CNN 模型在不同 t 时的分类精度

Table 3 Classification accuracies of different CNN models at different t unit: %

Model	C5			F6(G6)			F7		
	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
VGG-16	90.6±2.5	90.3±2.4	90.5±2.4	91.9±2.3	92.0±2.5	91.9±2.1	92.4±2.7	**93.3±2.0**	92.9±2.5
VGG-19	90.1±2.3	89.8±2.3	89.9±2.3	91.1±2.6	91.3±2.5	**92.0±2.5**	91.5±3.3	91.3±3.4	90.7±3.0
ResNet-50	91.8±1.9	92.1±2.1	92.2±2.0	94.0±2.1	94.0±2.2	**94.3±2.1**	–	–	–

4) 与其他方法的对比

实验选用 PCA 重构阈值 t 为 0.99、ResNet-50 模型 G6 层的融合方法与现有八种算法进行对比，包括四种基于传统特征的方法和四种基于 CNN 的方法。基于传统特征的方法包括基于多光谱 SIFT 的方法(MSIFT)[11]、基于 Fisher Vector 的方法(Fisher Vector)[12]、基于直方图统计变换的方法(mCENTRIST)[13]、基于密集 SIFT 和无字典模型的方法(DSIFT_CLM)[14]；基于 CNN 的方法包括基于双 CNN 的方法(Dual CNN)[15]、基于核函数主成分分析和典型相关分析的方法(CNN_KPCA_CCA)[16]、基于多路 CNN 的方法(MCNN)[17]和基于双通道 CNN 的方法(DC_ CNN)[5]，不同方法的双波段场景分类精度如表 4 所示，表中前三种方法的分类精度均采用随机选择 10 组训练和测试样本得到的分类精度平均值和均方差，而本方法采用 20 组。相比其他采用 1 组训练和测试样本方法的分类精度，本方法更加科学客观。可以发现，基于传统特征方法的分类精度普遍低于基于 CNN 的方法，原因是传统特征均为低级特征，不包含语义信息；而基于 CNN 的方法由底层到高层，低级特征逐层抽象为语义信息。与基于传统特征的最优方法 Fisher Vector 相比，本方法的分类精度提高了 6.4 个百分点。其他基于 CNN 的方法在微调模型或训练网络的耗时较长，而本方法仅 SVM 分类需要进行训练，且采用 PCA 方法降低了高维卷积特征的维度，有效减少了 SVM 分类器的训练时间。此外，PCA 方法和 SVM 分类器的时间复杂度均为线性复杂度。因此，本方法不仅分类精度与基于 CNN 的方法相当，且具有更高的执行效率。图 6 为本方法在 20 组训练/测试组中最好和最差融合分类精度的混淆矩阵，正对角线为正确分类精度，其余位置为误分类精度。

表 4 不同方法在 RGB-NIR 数据集上的分类精度比较

Table 4 Classification accuracy comparison of different methods

Method	Train/test group	Year	Classification accuracy /%		
			RGB	NIR	RGB+NIR
MSIFT	10	2011	62.9±3.1	–	73.1±3.3
Fisher Vector	10	2011	84.5±2.3	–	87.9±2.2
mCENTRIST	10	2014	78.9±5.1	–	84.5±2.1
DSIFT_CLM	1	2018	–	–	86.9
Dual CNN (GoogLeNet)	1	2017	–	–	92.5
CNN_KPCA_CCA (GoogLeNet)	1	2018	–	–	90.8
MCNN (ResNet-50)	1	2019	–	–	93.5
DC_CNN	1	2019	–	–	95.0
Our method (worst)	1(20)	2020	87.9	80.8	88.9
Our method (best)	1(20)	2020	96.0	93.9	98.0
Our method (ResNet-50)	20	2020	**92.3±1.9**	**88.7±3.2**	**94.3±2.1**

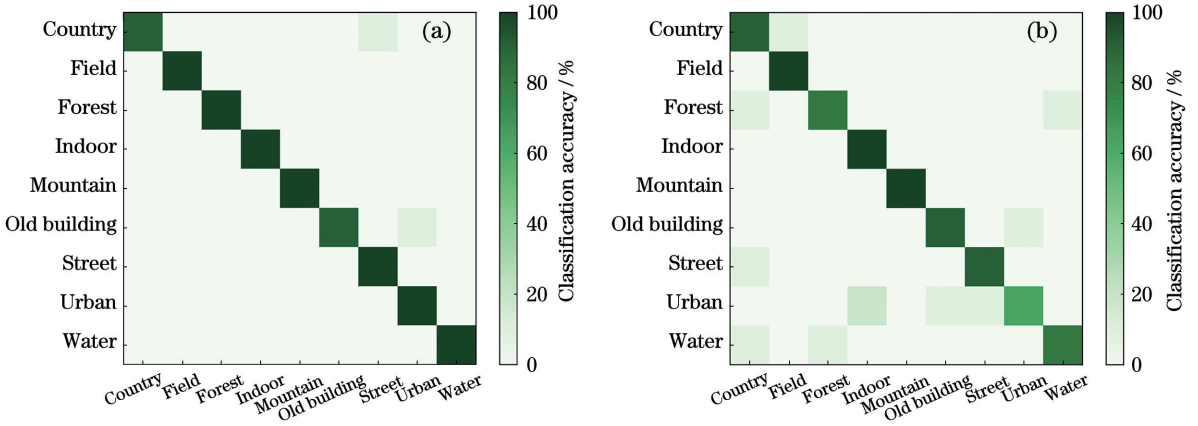

图 6 本方法的分类精度混淆矩阵。(a)20 组中最好的分类精度(98.0%);(b)20 组中最差的分类精度(88.9%)

Fig. 6 Classification accuracy confusion matrix of our method. (a) Best classification accuracy in the 20 groups (98.0%); (b) worst classification accuracy in the 20 groups (88.9%)

4 结 论

在双波段图像缺少标注样本和特征级联融合分类精度不高的情况下,利用预训练的经典 CNN 模型提取卷积特征,通过 PCA 方法降维和 SVM 计算后验概率;采用朴素贝叶斯决策融合分类,避免了网络模型训练出现过拟合以及特征融合共同表示学习难的问题。实验结果表明,在场景分类中,双波段图像决策级融合分类性能明显优于单一波段分类和级联特征融合分类的性能。本方法在 VGG-16 模型第 2 个全连接层、ResNet-50 模型全局平均池化层取得了最佳融合效果,与其他方法相比,具有分类精度高、处理速度快的优势,可适用于特殊领域数据样本少的计算机视觉任务。

参 考 文 献

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2020-06-15]. http://arxiv. org/abs/1409. 1556.

[3] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.

[4] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[5] Jiang J H, Feng X A, Liu F, et al. Multi-spectral RGB-NIR image classification using double-channel CNN[J]. IEEE Access, 2019, 7: 20607-20613.

[6] Liu F, Shen T S, Ma X X. Convolutional neural network based multi-band ship target recognition with feature fusion[J]. Acta Optica Sinica, 2017, 37(10): 1015002.
刘峰, 沈同圣, 马新星. 特征融合的卷积神经网络多波段舰船目标识别[J]. 光学学报, 2017, 37(10): 1015002.

[7] Ding L, Wang Y, Laganière R, et al. Convolutional neural networks for multispectral pedestrian detection [J]. Signal Processing: Image Communication, 2020, 82: 115764.

[8] Zhang Q, Huang N C, Yao L, et al. RGB-T salient object detection via fusing multi-level CNN features [J]. IEEE Transactions on Image Processing, 2020, 29: 3321-3335.

[9] Zhang X C, Ye P, Peng S Y, et al. DSiamMFT: an RGB-T fusion tracking method via dynamic Siamese networks using multi-layer feature fusion[J]. Signal Processing: Image Communication, 2020, 84: 115756.

[10] Xie L, Lee F, Liu L, et al. Scene recognition: a comprehensive survey [J]. Pattern Recognition, 2020, 102: 107205.

[11] Brown M, Süsstrunk S. Multi-spectral SIFT for scene category recognition[C]//CVPR 2011, June 20-25, 2011, Providence, RI, USA. New York: IEEE Press, 2011: 177-184.

[12] Salamati N, Larlus D, Csurka G. Combining visible and near-infrared cues for image categorisation[C]// 22nd British Machine Vision Conference (BMVC 2011), August 30-September 1, 2011, Dundee,

Scotland. UK: BMVA Press, 2011: 1-11.

[13] Xiao Y, Wu J X, Yuan J S. mCENTRIST: a multi-channel feature generation mechanism for scene categorization[J]. IEEE Transactions on Image Processing, 2014, 23(2): 823-836.

[14] Zhang Q S, Li W, Li L, et al. Infrared and visible image fusion classification based on a codebookless model(CLM)[J]. Journal of Beijing University of Chemical Technology (Natural Science Edition), 2018, 45(2): 71-76.
张秋实, 李伟, 李禄, 等. 基于无字典模型的红外与可见光图像融合分类[J]. 北京化工大学学报(自然科学版), 2018, 45(2): 71-76.

[15] Ševo I, Avramović A. Multispectral scene recognition based on dual convolutional neural networks[C]//Proceedings of the 10th International Symposium on Image and Signal Processing and Analysis, September 18-20, 2017, Ljubljana, Slovenia. New York: IEEE Press, 2017: 126-130.

[16] Peng X S, Li Y X, Wei X, et al. RGB-NIR image categorization with prior knowledge transfer[J]. EURASIP Journal on Image and Video Processing, 2018, 2018(1): 1-11.

[17] Jiang Z T, Qin J Q, Hu S. Multi-spectral scene recognition method based on multi-way convolution neural network[J]. Computer Science, 2019, 46(9): 265-270.
江泽涛, 秦嘉奇, 胡硕. 基于多路卷积神经网络的多光谱场景识别方法[J]. 计算机科学, 2019, 46(9): 265-270.

[18] Yosinski J, Clune J, Bengio Y, et al. How transferable are features in deep neural networks? [EB/OL]. [2020-06-13]. https://arxiv.org/abs/1411.1792v1.

[19] Zhao H H, Liu H. Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition[J]. Granular Computing, 2020, 5(3): 411-418.

[20] Woźniak M, Graña M, Corchado E. A survey of multiple classifier systems as hybrid systems[J]. Information Fusion, 2014, 16: 3-17.

[21] Zeng H, Yang B, Wang X Q, et al. RGB-D object recognition using multi-modal deep neural network and DS evidence theory[J]. Sensors, 2019, 19(3): 529.

[22] Tang C, Ling Y S, Yang H, et al. Decision-level fusion detection for infrared and visible spectra based on deep learning[J]. Infrared and Laser Engineering, 2019, 48(6): 456-470.
唐聪, 凌永顺, 杨华, 等. 基于深度学习的红外与可见光决策级融合检测[J]. 红外与激光工程, 2019, 48(6): 456-470.

[23] 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.

[24] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[EB/OL]. [2020-06-15]. https://arxiv.org/abs/1311.2901.

[25] Zhou Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016: 229-232.
周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 229-232.

[26] Lin H T, Lin C J, Weng R C. A note on Platt′s probabilistic outputs for support vector machines[J]. Machine Learning, 2007, 68(3): 267-276.