

# 基于注意力机制的多目标优化高光谱波段选择

关世豪<sup>1</sup>, 杨枕<sup>1\*</sup>, 卢珊<sup>2</sup>, 付严宇<sup>1</sup>

<sup>1</sup>中国人民解放军空军航空大学航空作战勤务学院, 吉林 长春 130022;

<sup>2</sup>东北师范大学地理科学学院, 吉林 长春 130024

**摘要** 神经网络的注意力机制可以从数据中提取关键信息, 将这一特性运用在高光谱波段选择上有助于充分学习波段之间的相互依赖和非线性关系, 提取更重要的波段。提出了一种基于注意力机制的多目标优化高光谱波段选择算法。首先, 利用注意力模块和自编码器构建网络; 然后, 将一维光谱数据作为网络输入, 采用两种损失函数并结合多目标优化方法对输入数据进行训练, 使嵌入在网络中的注意力模块充分学习各波段之间的非线性关系, 对信息量大和易于分类的波段赋予更大的权重, 以实现波段选择; 最后, 利用支持向量机分类器和平均光谱散度验证波段子集的性能。实验结果表明: 相比于其他算法, 所提算法在 Botswana 与 Indian Pines 数据集上提取的波段子集的分类精度更高, 信息量更大, 由此证明了所提算法对高光谱波段选择的有效性。

**关键词** 遥感; 高光谱图像; 波段选择; 注意力机制; 多目标优化

中图分类号 TP751.1

文献标志码 A

doi: 10.3788/AOS202040.2128002

## Multi-Objective Optimization of Hyperspectral Band Selection Based on Attention Mechanism

Guan Shihao<sup>1</sup>, Yang Guang<sup>1\*</sup>, Lu Shan<sup>2</sup>, Fu Yanyu<sup>1</sup>

<sup>1</sup>School of Aviation Operations and Services, Aviation University of Air Force, Changchun, Jilin 130022, China;

<sup>2</sup>School of Geographic Science, Northeast Normal University, Changchun, Jilin 130024, China

**Abstract** The attention mechanism of neural networks can extract key information from data, and the application of this feature in the selection of hyperspectral bands can help fully learn the interdependence and nonlinear relations between bands and extract more important bands. This paper presents a multi-objective optimization method for hyperspectral band selection based on the attention mechanism. First, the attention module and autoencoder are used to construct the network. Then, one-dimensional spectral data is provided as input to the network; two loss functions are used and combined with the multi-objective optimization method for training. Therefore, the attention module embedded in the network learns the nonlinear relationship between different bands and assigns more weight to the bands with a large amount of information and easy classification, thereby realizing band selection. Finally, the support vector machine classifier and mean spectral divergence are used to validate the performance of the band subset. The experimental results show that the band subset extracted using this method from the Botswana and Indian Pines datasets is more accurate and informative than the subsets extracted using other algorithms. Thus, it is demonstrated that this algorithm is more effective in selecting hyperspectral bands.

**Key words** remote sensing; hyperspectral image; band selection; attentional mechanism; multi-objective optimization

**OCIS codes** 280.4991; 100.4145; 100.4996

## 1 引 言

高光谱遥感图像是成像光谱仪以数十乃至数百个连续且细分的光谱波段对地面目标区域同时成像

得到的同时具有空间和光谱信息的三维结构图像<sup>[1]</sup>。凭借纳米级的光谱分辨率, 高光谱图像具有很强的地物分辨能力。但其相邻波段的相关性较强, 存在较高的数据冗余<sup>[2]</sup>, 再加上高光谱图像本身

收稿日期: 2020-06-30; 修回日期: 2020-07-09; 录用日期: 2020-07-20

基金项目: 国家自然科学基金(41971290)

\* E-mail: 1026269743@qq.com

存在小样本问题,在数据分析(如分类)时会产生较高的计算成本,并且极易导致“Hughes”现象的产生<sup>[3]</sup>。因此,对高光谱图像进行降维预处理是十分必要的。

高光谱图像的降维方法主要分为特征提取和波段选择两大类<sup>[4]</sup>。特征提取方法就是利用线性或非线性方法将原始特征从高维转换到低维,从而产生新的特征空间。但是原始高光谱数据在经过特征空间转换之后,数据的特征理解性很差,不具备原始的物理特性<sup>[5]</sup>。波段选择方法是从原始波段集合中选择一个波段子集,这样可以在保留原始数据物理意义及光谱特性的同时降低数据的维度。

现有的高光谱波段选择算法可以分为三类:滤波式、封装式和嵌入式。滤波式算法通过计算数据本身的固有属性来评价和选择波段,如最大方差主成分分析(MVPCA)<sup>[6]</sup>、基于几何方法的波段选择(OPBS)<sup>[7]</sup>和近邻传播聚类<sup>[8]</sup>等,此类算法的计算复杂度低,但波段子集的性能不稳定。封装式算法利用后续算法模型检测波段的选择性能,Ma等<sup>[9]</sup>利用随机森林和支持向量机两种分类器,从分类精度、计算时间和稳定性三个方面构建评价标准,进而选择波段。虽然封装式算法所选波段子集的分类精度较高,但是每一次迭代都需要重新训练模型,计算相对比较耗时。嵌入式算法在模型训练过程中通过优化目标函数来完成波段选择任务,避免了模型的重复训练,可以快速获得最优特征子集。Tschannerl等<sup>[10]</sup>利用自编码器重构光谱特征,对不同的输入权重进行稀疏优化,实现了波段选择。

注意力机制(AM)是深度学习领域中的一个研究热点,它通过对深度学习网络中不同的关注部分赋予不同的权重来抽取更加关键的信息<sup>[11]</sup>。Lorenzo等<sup>[12]</sup>构建了基于注意力机制的卷积神经网络(CNN),并用其对高光谱图像进行波段选择,但提取到的波段子集的相关性很强,数据冗余较大。Cai等<sup>[13]</sup>将波段选择任务转换成使用少量特征波段对原始光谱数据进行稀疏重建的任务,但是该算法仅从波段信息量的角度进行波段选择,难以确保所选波段最利于分类。

本文提出了一种基于注意力机制的多目标优化(MOP)波段选择方法(AM\_MOP),该算法将Hu等<sup>[14]</sup>提出的注意力模块 SENet 嵌入到自编码器(AE)构建的编码-解码网络中,结合多目标优化方法将交叉熵损失与均方误差(MSE)损失非负加权之和作为模型的目标损失函数,以充分学习光

谱数据之间的非线性关系,实现端到端的高光谱波段选择。最后,在 Botswana 与 Indian Pines 高光谱数据集上进行了波段选择实验,以验证所提算法的性能。

## 2 本文算法

高光谱图像的波段选择一般基于以下两个原则<sup>[15]</sup>:1)信息量原则,即少量波段包含全波段数据的大部分信息量;2)利于分类原则,即波段子集有利于地物分类,直接表现在波段子集的分类精度上。Cai等<sup>[13]</sup>依据信息量原则,利用注意力机制筛选少量波段进行光谱稀疏重构,迫使网络选择信息量较大的波段。本文拟借鉴这一方法的思路,并在信息量原则的基础上,考虑了利于分类的原则。

本文算法利用自编码器构建一个编码-解码网络,并将注意力模块嵌入到网络前端。主要思路包括两方面:一方面,通过编码-解码网络进行波段的稀疏重建任务,利用均方误差衡量稀疏重建的误差,选择信息量较大的波段;另一方面,将网络中编码部分提取到的特征向量用于计算交叉熵损失,通过降低交叉熵损失使注意力模块关注更有利于分类的特征(波段)。结合多目标优化理论,通过对比实验设置两类损失的加权系数,使网络中的两种损失同时得到优化。最终通过训练,选择注意力模块中各波段权重大的波段子集,这些波段子集即为波段选择结果。

### 2.1 SENet 结构

SENet 是一种可以嵌入到其他深度学习网络中的模块。其核心思想是通过降低训练损失学习特征通道的权重,使重要性较高的特征通道的权重较大,而重要性较低或无意义的特征通道的权重较小<sup>[14]</sup>。因此,本实验选择 SENet 学习光谱通道的权重。

在 SENet 中,处理过程主要包括 Squeeze、Excitation 和 Reweight 三个步骤<sup>[14]</sup>。Squeeze 部分对维度大小为  $W \times H \times B$  特征图进行全局平均池化(GAP),沿着空间维度进行特征压缩,获取全局感受野,输出大小为  $1 \times 1 \times B$ ;Excitation 部分使输入先经过一个全连接层(FC)实现数据的降维( $1 \times 1 \times (B/k)$ ),其中  $k$  为 SENet 中的参数,然后再经过一个全连接层对数据进行升维( $1 \times 1 \times B$ ),得到各个特征通道的权重;Reweight 部分将权重与原数据相乘,完成各特征通道的加权。SENet 结构如图 1 所示。

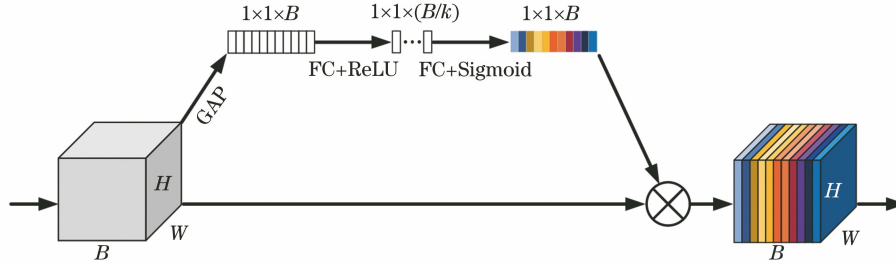


图 1 SENet 结构

Fig. 1 SENet structure

### 2.2 两类损失的多目标优化方法

大部分的机器学习问题本质上是一个多目标优化问题<sup>[16]</sup>。在深度学习领域中,多目标优化问题可以理解为将多个损失函数通过非负加权转换为单目标优化问题。其定义为

$$L = \min \sum_{i=1}^n \omega_i L_i (\omega_i \geq 0), \quad (1)$$

式中: $L$  代表总目标函数; $L_i$  代表第  $i$  个损失函数; $\omega_i$  代表第  $i$  个目标函数的权重; $n$  为损失函数的个数。本文选择均方误差损失 ( $L_{\text{MSE}}$ ) 和交叉熵损失 ( $L_{\text{Cross Entropy}}$ ) 加权作为总损失函数。这两类损失函数公式分别为

$$L_{\text{MSE}} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2, \quad (2)$$

$$L_{\text{Cross Entropy}} = -[x \lg \hat{x} + (1 - x) \lg(1 - \hat{x})]. \quad (3)$$

式中: $y_i$  为单个像素在第  $i$  个波段的真实值; $\hat{y}_i$  为单个像素在第  $i$  个波段的预测值; $m$  为波段数; $x$  代表真实样本标签; $\hat{x}$  代表预测样本标签。

模型的总目标损失为两类损失的加权之和,训练过程中需要设置合理的权重,使两类目标得到优化,但是对两个权重进行遍历组合则计算量较大。本文将均方误差损失的权重设为 1,交叉熵损失的权重系数设为  $\gamma$ ,则总目标损失函数  $L$  可表示为

$$L = L_{\text{MSE}} + \gamma L_{\text{Cross Entropy}}. \quad (4)$$

权重系数  $\gamma$  的大小由后续对比实验确定。

### 2.3 波段选择模型结构

本文提出的波段选择模型结构如图 2 所示,包含注意力模块、光谱重构模块和分类模块。

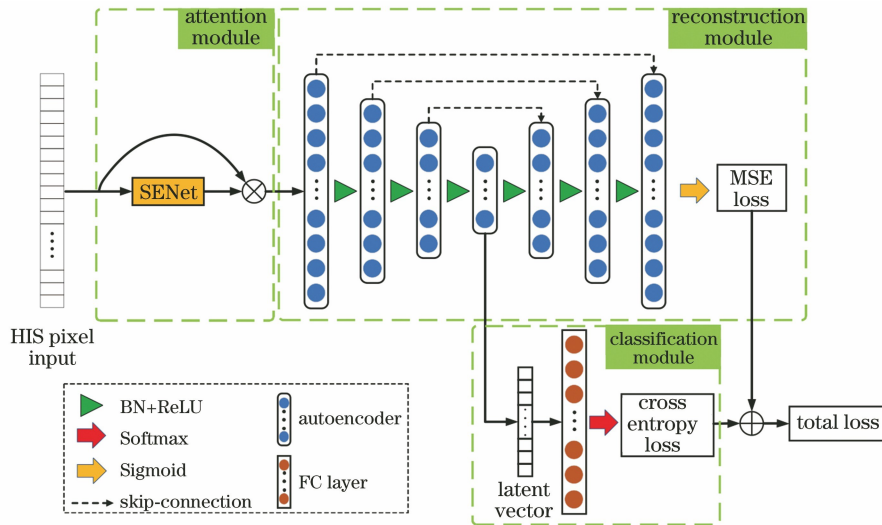


图 2 波段选择模型结构

Fig. 2 Band selection model structure

注意力模块由一个 SENet 组成,该模块将单个像素包含的一维光谱数据作为模型输入。光谱重构模块是由自编码器构成的编码-解码网络,该模块对光谱数据进行编码,提取深层特征(降维),然后将深层特征进行解码重构(升维),计算重构数据与原始

光谱数据的均方误差损失。每个自编码器的输出均经过批归一化(BN)层与 ReLU 激活函数。为了避免梯度消失导致注意力模块得不到有效学习的情况,网络中的编码器和与之相对应的解码器之间使用跳跃连接(skip-connection)<sup>[17]</sup>。分类模块将光谱

重构模块中编码部分提取到的深层特征向量(latent vector)输入到全连接层中,经过 Softmax 激活函数后计算交叉熵损失。最后将两类损失加权求和得到

总目标损失,再通过训练使总目标损失收敛。模型中各层的输入、输出尺寸以及激活函数如表 1 所示( $b$  代表高光谱图像的波段数)。

表 1 模型中的数据尺寸及激活函数变化

Table 1 Data size and activation function change in the model

Module	Layer	Input size	Output size	Activation
	Input		$1 \times 1 \times b$	
Attention module	FC-1(fully connected layer)	$1 \times 1 \times b$	$1 \times 1 \times (b/16)$	ReLU
	FC-2(fully connected layer)	$1 \times 1 \times (b/16)$	$1 \times 1 \times b$	Sigmoid
Reconstruction module	Encoder-1(autoencoder)	$1 \times 1 \times b$	$1 \times 1 \times 256$	
	BN-1(batch normalization)	$1 \times 1 \times 256$	$1 \times 1 \times 256$	ReLU
	Encoder-2(autoencoder)	$1 \times 1 \times 256$	$1 \times 1 \times 128$	
	BN-2(batch normalization)	$1 \times 1 \times 128$	$1 \times 1 \times 128$	ReLU
	Encoder-3(autoencoder)	$1 \times 1 \times 128$	$1 \times 1 \times 64$	
	BN-3(batch normalization)	$1 \times 1 \times 64$	$1 \times 1 \times 64$	ReLU
	Encoder-4(autoencoder)	$1 \times 1 \times 64$	$1 \times 1 \times 64$	
	BN-4(batch normalization)	$1 \times 1 \times 64$	$1 \times 1 \times 64$	ReLU
	Decoder-1(autoencoder)	$1 \times 1 \times 64$	$1 \times 1 \times 128$	
	BN-5(batch normalization)	$1 \times 1 \times 128$	$1 \times 1 \times 128$	ReLU
	Decoder-2(autoencoder)	$1 \times 1 \times 128$	$1 \times 1 \times 256$	
	BN-6(batch normalization)	$1 \times 1 \times 256$	$1 \times 1 \times 256$	ReLU
Decoder-3(autoencoder)	$1 \times 1 \times 256$	$1 \times 1 \times b$	Sigmoid	
Classification module	Latent vector	$1 \times 1 \times 64$	$1 \times 1 \times 64$	
	FC-3(fully connected layer)	$1 \times 1 \times 64$	Number of class	Softmax

本文算法的具体实现步骤如下:

1) 数据预处理。去除高光谱数据中的水吸收及噪声波段,对原数据进行 min-max 标准化,将所有像元值缩放到 $[0, 1]$ 范围内。

2) 模型训练。将一维光谱数据( $1 \times 1 \times b$ )作为模型输入进行训练,直到损失收敛。

3) 选择波段。提取注意力模块中各波段的权重,比较每个波段权重的大小,选择权重较大的波段作为最终的波段。

### 3 数据集与评价标准

#### 3.1 实验数据集

本文选择 Botswana 与 Indian Pines 高光谱图像数据集对模型波段的选择效果进行验证。Botswana 原始数据集包含 242 个光谱波段,去除噪声和受水汽影响的波段(去除波段序号为 1~9, 56~81, 98~10, 120~133, 165~186)后剩余 145

个波段;Indian Pines 原始数据集包含 220 个光谱波段,去除噪声和水汽影响的波段(去除波段序号为 104~108, 150~163, 220)后剩余 200 个波段。Botswana 数据集和 Indian Pines 数据集的拍摄地点、成像传感器、分辨率、地物类别数及样本数等信息如表 2 所示,图 3 和图 4 分别展示了 Botswana 和 Indian Pines 数据集的真彩色图像和地物真值图。

#### 3.2 评价标准

本实验从分类精度和信息量两个方面对波段子集进行评价。采用总体分类精度(OA)、平均分类精度(AA)和 Kappa 系数这三种常用指标来衡量波段子集的分类精度。其中:总体分类精度是指在测试集上分类正确的样本占所有样本的比例;平均分类精度是指在每一地物类别样本中分类正确的样本占总样本的比例;Kappa 系数也是一种衡量分类精度的指标,主要用于衡量模型预测效果和实际分类效果的一致性。利用混淆矩阵计算 Kappa 系数,计算公式为

表 2 高光谱图像数据集  
Table 2 Hyperspectral image data set

Item	Botswana	Indian Pines
Shooting area	Okavango Delta, Botswana	Indiana, USA
Imaging spectrometer	Hyperion	AVIRIS
Spectral range /nm	400–2500	400–2500
Number of wavelengths (remove strong noise and water vapor band)	145	200
Image size /(pixel×pixel)	1476×256	145×145
Spatial resolution /m	30	20
Sample size	3248	10249
Object types	14	16

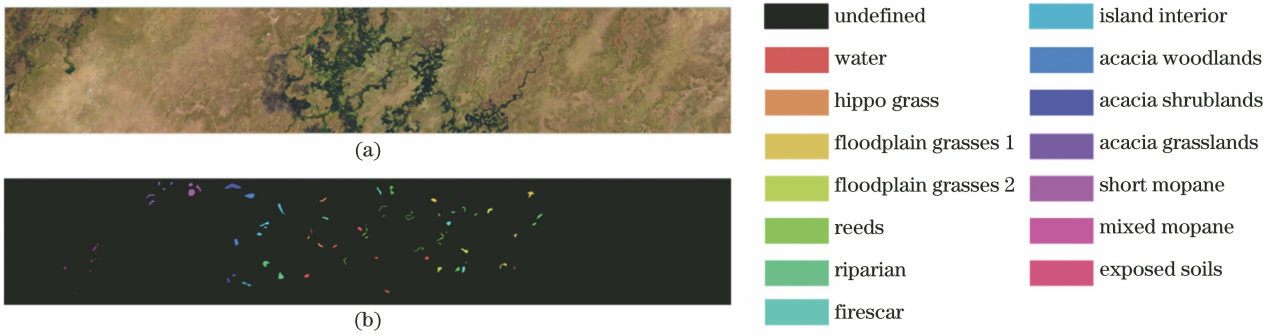


图 3 Botswana 数据集的真彩色图像和地物真值图。(a)真彩色图像;(b)地物真值图

Fig. 3 True color image and ground truth map of Botswana data set. (a) True color image; (b) ground truth map

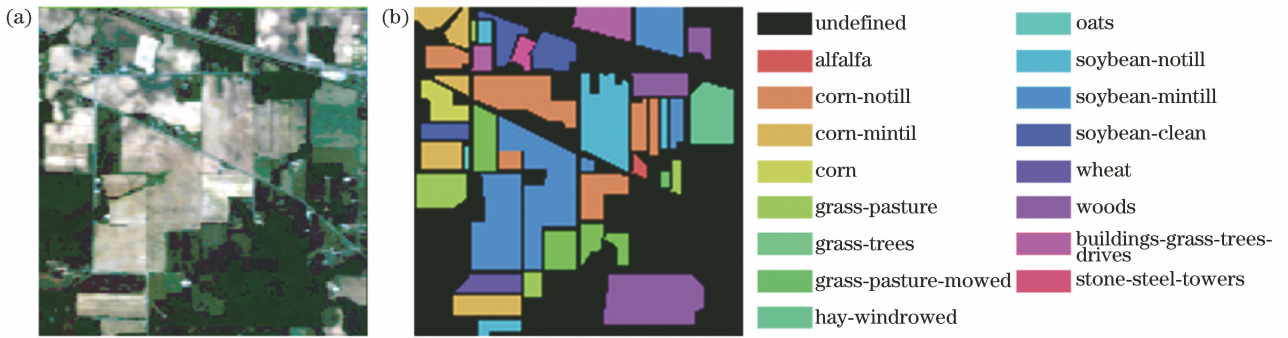


图 4 Indian Pines 数据集的真彩色图像和地物真值图。(a)真彩色图像;(b)地物真值图

Fig. 4 True color image and ground truth map of Indian Pines data set. (a) True color image; (b) ground truth map

$$C_{\text{Kappa}} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})}, \quad (5)$$

式中： $N$  表示总样本的个数； $r$  表示混淆矩阵的行数； $x_{i+}$  表示混淆矩阵第  $i$  行元素之和； $x_{+i}$  表示混淆矩阵第  $i$  列元素之和； $x_{ii}$  代表混淆矩阵第  $i$  行、第  $i$  列的元素值。

利用支持向量机(SVM)计算总体分类精度、平均分类精度和 Kappa 系数。计算策略如下：从波段

子集中随机选取 10% 的样本作为训练集，对 SVM 分类器进行训练，将其余数据作为验证集。对每种算法提取到的波段子集进行 10 次重复实验。

平均光谱散度(MSD)<sup>[18]</sup>用来衡量波段之间的相关性,MSD 值越大,所选波段之间的相关性就越小,信息量也就越大。本文采用 MSD 定量分析波段子集的信息量,计算公式为

$$\alpha_{\text{MSD}} = \frac{2}{k(k-1)} \sum_{i=1}^k \sum_{j=1}^k D_{\text{SKL}}(\mathbf{X}_i \parallel \mathbf{X}_j), \quad (6)$$

式中： $D_{\text{SKL}}$  表示对称 KL 散度 (symmetrical

Kullback-Leibler divergence)<sup>[19]</sup>;  $\mathbf{X}_i$  和  $\mathbf{X}_j$  分别表示第  $i$  和第  $j$  波段的图像;  $k$  表示波段子集中的波段数。  $D_{SKL}$  定义如下

$$D_{SKL}(\mathbf{X}_i \parallel \mathbf{X}_j) = D_{KL}(\mathbf{X}_i \parallel \mathbf{X}_j) + D_{KL}(\mathbf{X}_j \parallel \mathbf{X}_i), \quad (7)$$

其中,

$$D_{KL}(\mathbf{X}_i \parallel \mathbf{X}_j) = \sum_{y \in \Omega} p_i(y) \lg \frac{p_i(y)}{p_j(y)} \quad (8)$$

式中:  $D_{KL}$  表示 KL 散度;  $y$  和  $\Omega$  分别表示图像中的灰度级和灰度级空间;  $p_i(y)$  表示第  $i$  波段图像各灰度级的概率分布;  $p_j(y)$  表示第  $j$  波段图像各灰度级的概率分布。

## 4 实验结果及分析

### 4.1 实验方案

所有实验均在同一台计算机上完成, 计算机配

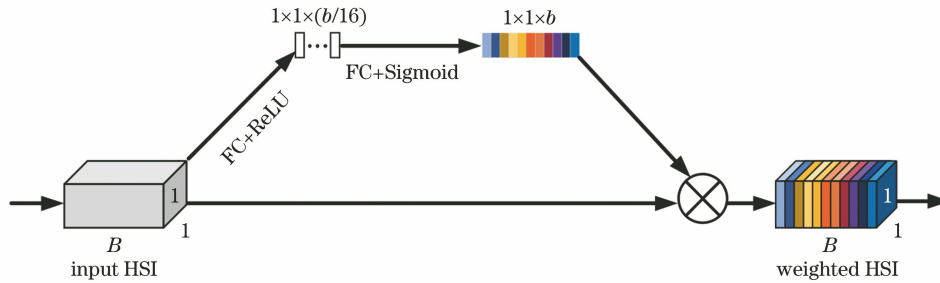


图 5 实验中使用的 SENet 结构

Fig. 5 SENet structure in the experiment

### 4.2 实验结果分析

#### 4.2.1 损失函数权重的选择

将权重系数  $\gamma$  分别设为 0.1、0.3、0.5、0.7、0.9, 在两个数据集上进行对比实验, 选择 10 个最具代表性的波段, 使用 SVM 分类器验证波段选择的效果, 结果如表 3 所示。选择使分类精度和 Kappa 系数最高的权重  $\gamma$  作为损失函数权重。因此,  $\gamma$  在 Botswana 数据集上设为 0.3, 在 Indian Pines 数据集上设为 0.5。

表 3 不同权重系数下两个数据集的实验结果

Table 3 Experimental results of two data sets with different weight coefficients

$\gamma$	Botswana			Indian Pines		
	OA / %	AA / %	Kappa	OA / %	AA / %	Kappa
0.1	88.9	89.5	0.873	73.1	71.4	0.708
0.3	89.3	89.8	0.886	73.6	71.5	0.706
0.5	88.6	87.1	0.869	74.3	70.4	0.712
0.7	87.2	86.8	0.853	72.1	69.5	0.698
0.9	85.3	86.7	0.839	69.7	66.1	0.664

置为 Intel (R) Core (TM) i7-8750H CPU @ 2.20 GHz 2.21 GHz, 8 GB 运行内存, NVIDIA GeForce GTX 1060 显卡。软件平台采用 Windows 10 系统下的 Python 3.6.0 和 PyTorch 0.4.0 深度学习框架。

本文采用光谱数据进行波段选择实验, 数据输入维度为  $1 \times 1 \times b$ , 利用 SENet 完成注意力学习时无需进行空间维度的压缩。除此之外, 在 Excitation 步骤中, 参考文献[14], 令  $k=16$ 。本文实验中使用的 SENet 结构如图 5 所示。

本实验使用 SGD 优化器进行网络训练, 初始学习率设为 0.01, 动量设为 0.8, 权重衰减率设为 0.0005。模型的 batch size 设为 32, 网络训练 epoch 数设为 100。为了充分学习样本数据的特征空间, 本实验将所有标注和未标注的数据都输入到模型中进行训练。

#### 4.2.2 AM\_MOP 算法实现过程分析

为了进一步分析 AM\_MOP 算法的波段选择过程, 使用 AM\_MOP 算法分别在 Botswana 和 Indian Pines 数据集上提取 10 个最具代表性的波段, 分析训练过程中两个数据集的总体分类精度、训练损失和各波段权重的变化。将所有权重缩放到  $[0, 1]$  范围内, 用热力图显示不同训练阶段的波段权重变化, 其中横轴代表高光谱数据的波段序号, 纵轴代表模型训练 epoch 数, 如图 6 和图 7 所示。

对于 Botswana 数据集, 观察图 6(a) 可以发现: 在训练过程中, 波段子集的总体分类精度在 0~30 个 epoch 范围内增加得较快, 从 74.4% 增加到 88.3%; 在 30~100 个 epoch 范围内, 分类精度从 88.3% 增加到 89.3%, 分类精度增加得较慢。观察图 6(b) 所示的训练损失和图 6(c) 所示的波段权重变化可知: 在 0~30 个 epoch 范围内, 训练损失下降得较快, 从 0.97 下降到 0.14, 波段权重热力图也发生了较大变化; 在 30~100 个 epoch 范围内, 训练损失下降的幅度减缓并逐渐趋近于 0, 波段权重热力图的变化也较小。

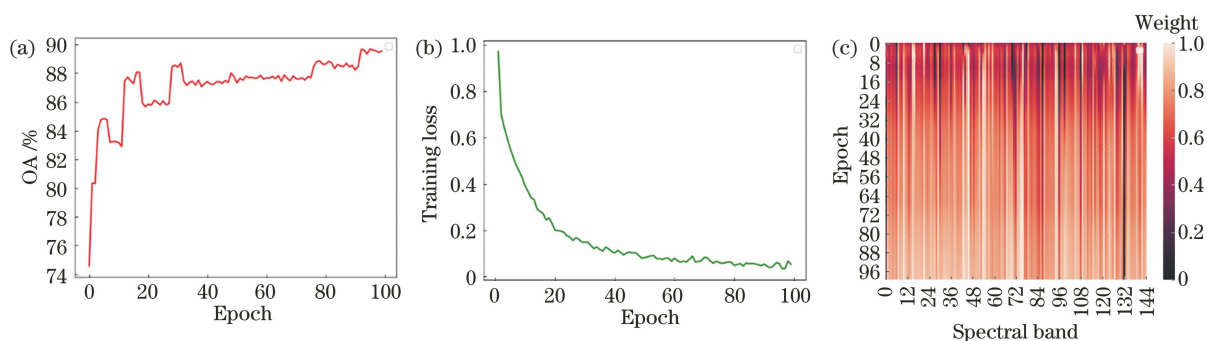


图 6 Botswana 数据集上的总体分类精度、训练损失和波段权重变化。(a) 总体分类精度; (b) 训练损失; (c) 波段权重热力图

Fig. 6 Overall classification accuracy, training loss, and band weight changes in the Botswana data set.

(a) Overall classification accuracy; (b) training loss; (c) band weight thermal map

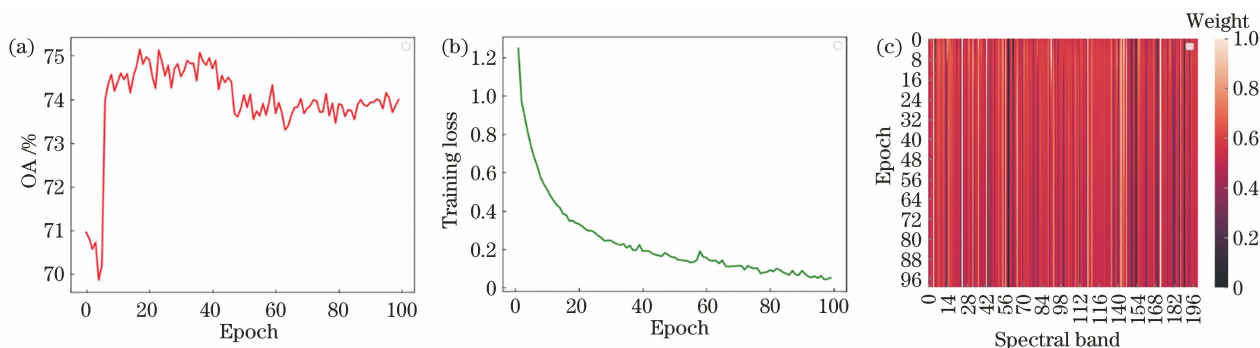


图 7 Indian Pines 数据集上的总体分类精度、训练损失和波段权重变化。(a) 总体分类精度; (b) 训练损失; (c) 波段权重热力图

Fig. 7 Overall classification accuracy, training loss and band weight changes on the Indian Pines data set.

(a) Overall classification accuracy; (b) training loss; (c) band weight thermal map

对于 Indian Pines 数据集,从图 7(a)中可以看出:所选波段子集的总体分类精度在 0~20 个 epoch 范围内增加得较快,从 70.9% 增加到 74.7%;在 20~100 个 epoch 范围内,分类精度在 74% 左右波动。对比观察图 7(b) 所示的训练损失和图 7(c) 所示的波段权重变化可知:在 0~20 个 epoch 范围内,训练损失下降得较快,从 1.25 下降到 0.28,波段权重热力图也发生了较大变化;在 20~100 个 epoch 范围内,训练损失下降的幅度减缓并逐渐趋近于 0,波段权重热力图的变化也较小。

综合 Botswana 和 Indian Pines 数据集上的实验结果可以发现,本文提出的波段选择算法可以在深度学习网络训练过程中,通过收敛训练损失来优化注意力模块上各波段的权重。训练损失下降得越快,波段权重更新得就越快;当训练损失趋于 0 时,各波段权重几乎不再变化。在这一过程中,注意力模块上的信息量较大,有利于分类的波段被赋予更大的权重,进而达到提取特征波段子集的目的。

#### 4.2.3 波段选择算法的性能对比

为了验证本文所提算法的性能,将本文所提

AM\_MOP 算法与 MVPCA<sup>[6]</sup>、OPBS<sup>[7]</sup>、ISSC<sup>[20]</sup> 这三种常用的波段选择算法,以及其他三种基于注意力机制的波段选择算法 BS\_Net\_FC<sup>[13]</sup>、BS\_Net\_Conv<sup>[13]</sup>、BS\_CNN<sup>[12]</sup> 进行对比。所有算法的参数及训练策略均采用原文中性能最优的参数和策略。

在 Botswana 数据集上进行波段选择实验,分别采用不同的波段选择算法提取 5~35 之间不同大小的波段子集,使用 SVM 分类器验证波段子集的分类精度,结果如图 8 所示。从图中可以看出:当波段数小于 15 时,AM\_MOP 算法提取到的波段子集的三个指标明显高于其他算法;当波段数大于等于 15 时,除了 MVPCA、OPBS 和 ISSC 三种算法,其他算法提取的波段子集的分类精度和 Kappa 系数十分接近,但 AM\_MOP 算法仍略高于其他几种算法。实验结果证明,AM\_MOP 算法在 Botswana 数据集上的波段提取性能优于其他几类算法,尤其是当波段数较小时,AM\_MOP 算法能够提取更具代表性的波段。此外,从整体情况来看,基于注意力机制的波段选择算法普遍比常规算法提取的波段子集的分类精度更高。

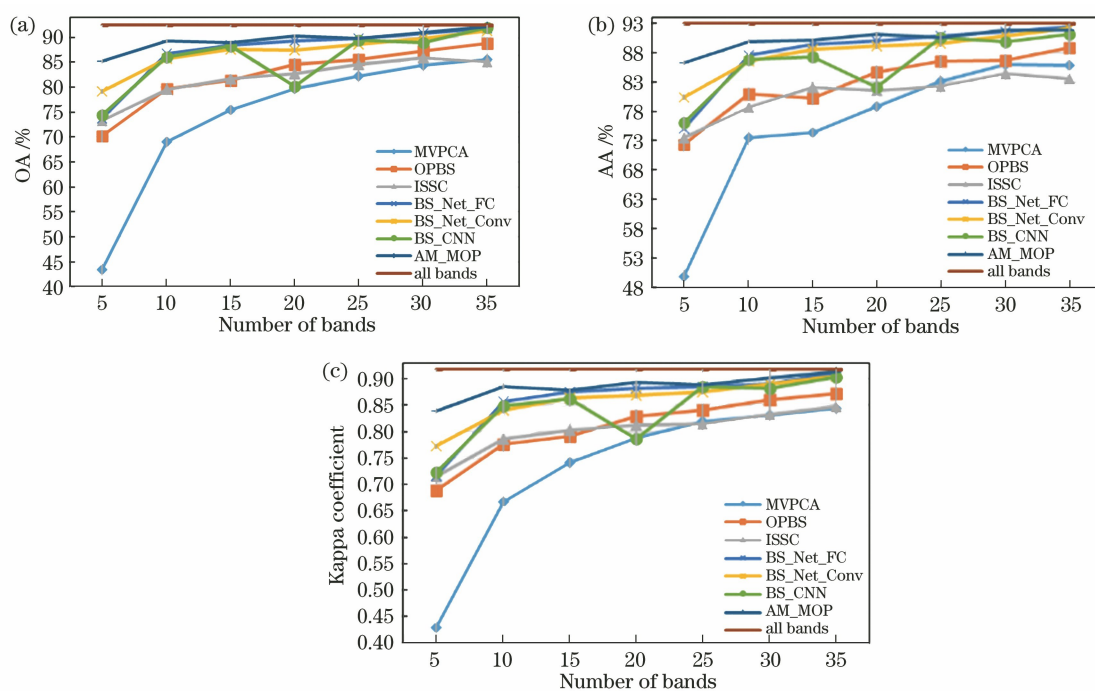


图 8 各算法在 Botswana 数据集上的总体分类精度、平均分类精度和 Kappa 系数。(a)总体分类精度；  
(b)平均分类精度；(c) Kappa 系数

Fig. 8 Overall classification accuracy, average classification accuracy and Kappa coefficient of each algorithm in the Botswana data set. (a) Overall classification accuracy; (b) average classification accuracy; (c) Kappa coefficient

图 9 是各算法在 Botswana 数据集上的平均光谱散度。尽管 AM\_MOP 算法提取的波段子集在分类性能上最优,但是当波段子集大小为 5、10、20 时, AM\_MOP 算法的平均光谱散度并没有达到最高;当波段子集大小为 5 时, MVPCA、OPBS、BS\_Net\_FC、BS\_Net\_Conv 算法的分类精度和 Kappa 系数低于 AM\_MOP 算法,但平均光谱散度高于 AM\_MOP 算法。根据文献[21]中的结论,当存在噪声波段时,波段子集的平均光谱散度也会增加。为了进一步验证这一结论,本文根据(6)式计算 Botswana 原始数据集中 1~9、89~97、10~19、98~106 波段的平均光谱散度,计算结果分别为 121.3、118.6、21.2、33.6。显然,在 4 个波段子集大小相同的情况下,前两个波段子集的平均光谱散度远高于后两个,这是因为前两个波段子集是 Botswana 数据集在预处理过程中被剔除的噪声波段。尽管前两个波段子集的平均光谱散度较高,但噪声波段对于分类的意义不大,甚至会起到负面作用。

在 Indian Pines 数据集上进行波段选择实验,分别采用不同的波段选择算法提取 5~35 之间不同大小的波段子集,使用 SVM 分类器验证波段子集的分类精度,结果如图 10 所示。从图中可以看出: AM\_MOP 算法提取到的波段子集的分类精度和

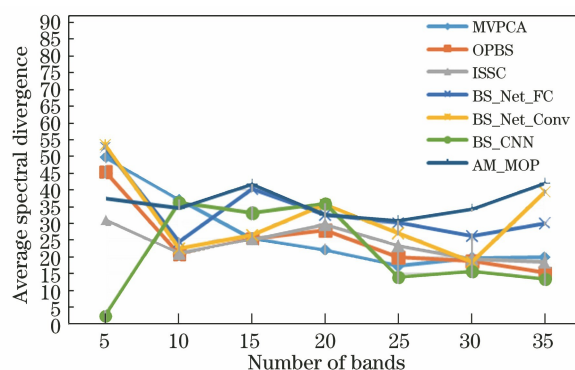


图 9 各算法在 Botswana 数据集上平均光谱散度  
Fig. 9 Average spectral divergence of each algorithm on the Botswana data set

Kappa 系数在 5~35 波段范围内优于其他算法; BS\_CNN 算法提取的波段子集在三个指标上远低于 ISSC 算法;另外,当波段子集的大小大于 20 时, AM\_MOP 算法提取的波段子集的分类精度与全波段高光谱数据的分类精度几乎一致。实验结果表明, AM\_MOP 算法在 Indian Pines 数据集上的波段选择性能优于其他几类算法。

图 11 是各算法在 Indian Pines 数据集上的平均光谱散度。从图 11 中可以看出,虽然 MVPCA 与 BS\_CNN 算法的平均光谱散度较高,但其分类精度较低。与 Botswana 数据集上的情况类似,尽管



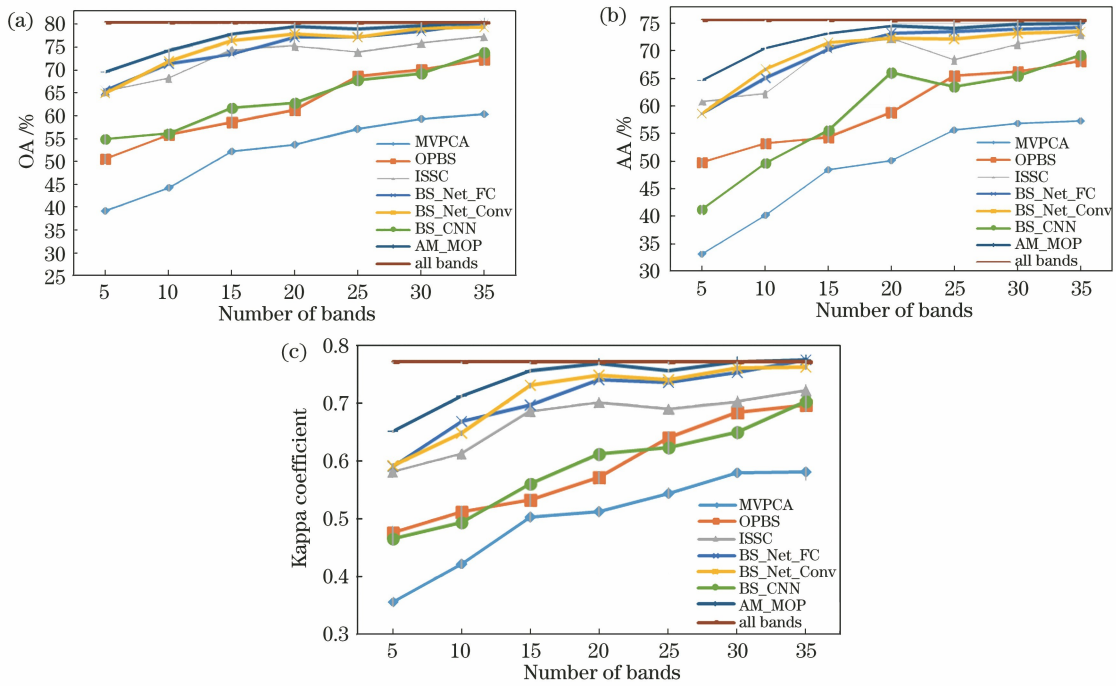


图 10 各算法在 Indian Pines 数据集上的总体分类精度、平均分类精度和 Kappa 系数。(a)总体分类精度；(b)平均分类精度；(c) Kappa 系数

Fig. 10 Overall classification accuracy, average classification accuracy and Kappa coefficient of each algorithm in the Indian Pines data set. (a) Overall classification accuracy; (b) average classification accuracy; (c) Kappa coefficient

MVPCA 与 BS\_CNN 算法提取的波段子集的平均光谱散度较高,但分类精度较低,这可能是由于波段子集中存在噪声波段。相比之下,虽然 AM\_MOP 算法的平均光谱散度较低,但其具有更高的分类精度。

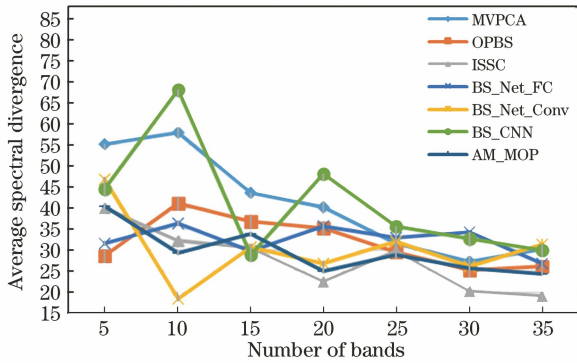


图 11 各算法在 Indian Pines 数据集上的平均光谱散度  
Fig. 11 Average spectral divergence of each algorithm on the Indian Pines data set

## 5 结 论

本文结合注意力机制和多目标优化理论提出了一种端到端的高光谱图像波段选择算法。该算法的主要思想是将波段选择任务转化为一个注意力模块的优化任务:利用自编码器构建编码-解码网络,选择两种目标损失函数对网络进行训练,使嵌入在深

度学习网络中的注意力模块学习各波段之间的非线性关系;当训练损失收敛时,注意力模块对信息量大、有利于分类的波段赋予更大的权重,最终实现波段选择。两个数据集上的实验结果表明:基于注意力机制的波段选择算法的性能优于常规算法;相对于其他基于注意力机制的波段选择算法,本文算法提取到的波段子集的分类精度更高,信息量更大。

但是本文仅针对 Botswana 与 Indian Pines 两个数据集进行了实验,在其他高光谱数据集上的性能仍需验证。另外,在将特征提取能力更强的深度学习网络作为骨干网络、进一步优化训练策略以及降低计算复杂度等方面仍需进行进一步研究。

## 参 考 文 献

[1] Bioucas-Dias J M, Plaza A, Camps-Valls G, et al. Hyperspectral remote sensing data analysis and future challenges[J]. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(2): 6-36.  
 [2] Zhang B. Advancement of hyperspectral image processing and information extraction[J]. Journal of Remote Sensing, 2016, 20(5): 1062-1090.  
 张兵. 高光谱图像处理与信息提取前沿[J]. 遥感学报, 2016, 20(5): 1062-1090.  
 [3] Hughes G. On the mean accuracy of statistical

- pattern recognizers [J]. *IEEE Transactions on Information Theory*, 1968, 14(1): 55-63.
- [4] Su H J, Du P J, Sheng Y H. Study on band selection algorithms of hyperspectral image data [J]. *Application Research of Computers*, 2008, 25(4): 1093-1096.  
苏红军, 杜培军, 盛业华. 高光谱影像波段选择算法研究[J]. *计算机应用研究*, 2008, 25(4): 1093-1096.
- [5] Yuan Y F. Band selection and classification for hyperspectral image based on particle swarm optimization and mutual information [D]. Xi'an: Xidian University, 2014: 9-10.  
袁永福. 基于粒子群和互信息的高光谱图像波段选择和分类[D]. 西安: 西安电子科技大学, 2014: 9-10.
- [6] Chang C I, Du Q, Sun T L, et al. A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 1999, 37(6): 2631-2641.
- [7] Zhang W Q, Li X R, Dou Y X, et al. A geometry-based band selection approach for hyperspectral image analysis [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(8): 4318-4333.
- [8] Ren Z W, Wu L D. Hyperspectral band selection based on affinity propagation clustering [J]. *Laser & Optoelectronics Progress*, 2018, 55(10): 103002.  
任智伟, 吴玲达. 基于 AP 聚类的高光谱波段选择 [J]. *激光与光电子学进展*, 2018, 55(10): 103002.
- [9] Ma L, Fu T Y, Blaschke T, et al. Evaluation of feature selection methods for object-based land cover mapping of unmanned aerial vehicle imagery using random forest and support vector machine classifiers [J]. *ISPRS International Journal of Geo-Information*, 2017, 6(2): 51.
- [10] Tschannerl J, Ren J C, Zabalza J, et al. Segmented autoencoders for unsupervised embedded hyperspectral band selection [C]//2018 7th European Workshop on Visual Information Processing (EUVIP). November 26-28, 2018, Tampere, Finland. New York: IEEE Press, 2018: 1-6.
- [11] Zhu Z L, Rao Y, Wu Y, et al. Research progress of attention mechanism in deep learning [J]. *Journal of Chinese Information Processing*, 2019, 33(6): 1-11.  
朱张莉, 饶元, 吴渊, 等. 注意力机制在深度学习中的研究进展 [J]. *中文信息学报*, 2019, 33(6): 1-11.
- [12] Lorenzo P R, Tulczyjew L, Marcinkiewicz M, et al. Band selection from hyperspectral images using attention-based convolutional neural networks [J/OL]. 2018: arxiv:1811.02667 [2020-06-30]. <https://arxiv.org/abs/1811.02667?context=cs.LG>.
- [13] Cai Y M, Liu X B, Cai Z H. BS-nets: an end-to-end framework for band selection of hyperspectral image [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(3): 1969-1984.
- [14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [15] Liu J P, Zhao Y S. Methods on optimal bands selection in hyperspectral remote sensing data interpretation [J]. *Journal of the Graduate School of the Chinese Academy of Sciences*, 1999(2): 153-161.  
刘建平, 赵英时. 高光谱遥感数据解译的最佳波段选择方法研究 [J]. *中国科学院研究生院学报*, 1999(2): 153-161.
- [16] Jin Y C. *Multi-objective machine learning* [M]. Berlin: Springer-Verlag, 2006.
- [17] Kuga R, Kanezaki A, Samejima M, et al. Multi-task learning using multi-modal encoder-decoder networks with shared skip connections [C]//2017 IEEE International Conference on Computer Vision Workshops (ICCVW). October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 403-411.
- [18] Geng X R, Sun K, Ji L Y, et al. A fast volume-gradient-based band selection method for hyperspectral image [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(11): 7111-7119.
- [19] Martínez-Usómartínez-uso A, Pla F, Sotoca J M, et al. Clustering-based hyperspectral band selection using information measures [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, 45(12): 4158-4171.
- [20] Sun W W, Zhang L P, Du B, et al. Band selection using improved sparse subspace clustering for hyperspectral imagery classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2784-2797.
- [21] Gong M G, Zhang M Y, Yuan Y. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(1): 544-557.