

分类号 TP75  
U D C           

密 级             
编 号 10486

# 武汉大学

## 硕士专业学位论文

### 基于多尺度空谱特征的高光谱影像分类 研究

研 究 生 姓 名: 王迪

学 号: 2018286190128

指导教师姓名、职称: 史磊 副教授、张良培 教授  
杜博 教授

专 业 类 别 (领 域): 测绘工程

二〇二零年五月

# Hyperspectral Image Classification Based on Multi-Scale Spectral-Spatial Feature

Candidate: WANG DI

Student Number: 2018286190128

Supervisor: A/PROF. SHI LEI、PROF. ZHANG LIANGPEI  
PROF. DU Bo

Major: Surveying and Mapping Engineering

State Key Laboratory of Information Engineering in  
Surveying, Mapping and Remote Sensing  
WUHAN UNIVERSITY

May, 2020

## 论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者（签名）：

2020 年 5 月 22 日



## 摘要

本文提出了三个用于端到端高光谱图像分类的深度网络。

首先从感受野的本质出发，并考虑到信息前向传播时底层细节信息损失的特性，本文利用不同尺寸的输入构建出了一个能够完成信息补偿并通过特征多样化实现利用多尺度特征进行高光谱分类的“多尺度混合补偿网络”。

进一步，为了提取以往常常被人们忽视的空间上下文信息，利用一维序列模型 RNN，本文构建了一个空间上下文信息感知模块，成功解决了过去研究中提取上下文信息过程中存在的结构难以保持的问题，并在多尺度特征提取的基础上综合利用多种光谱及空间信息构建出“混合信息感知网络”，实现了高光谱图像的端到端分类。

接着，针对以往多尺度特征提取的工作往往在空间层面操作而在光谱方面研究进展缓慢的现象，本文又在光谱方面分别基于地物的波长与 CNN 中特征的感受野提出了两种多尺度方案并借助一维序列模型 LSTM 提取多尺度光谱上下文信息，有效的提高了光谱分类的精度。在空间方面，本文将一维交替分组方案进行拓展，开发出了多维交替分组方案，并利用多维序列模型 ConvLSTM 成功的探索出了一种能够在更大尺度下提取空间上下文信息的方法。

为了将提取到的多尺度空间与光谱上下文特征进行结合以产生更有效的空谱多尺度特征，本文从数学方面证明了已有融合方法的不足，并考虑到光谱特征提取和空间特征提取部分不同的重要性，提出了一种自适应的特征融合方式，赋予了模型更强的数据表示能力，最终构建出一种通过端到端提取多尺度空谱上下文信息进行高光谱图像分类的“自适应空谱多尺度网络”。

实验结果表明：与 3-DCNN，SSUN 等现有方法相比，本文提出的方法具有更高的分类精度。由于本文提出的方法能够提取到更有效的空谱特征，因而也得到了边界轮廓更加明显，地物内部更加干净平滑具有区分度的分类图。这对利用高光谱图像进行土地覆盖识别、城市规划、环境监测等应用的相关部门具有重要意义。

**关键词：**多尺度，上下文信息，序列模型，端到端，高光谱图像分类。

## ABSTRACT

In this article, three end-to-end deep networks for hyperspectral image (HSI) classification are proposed.

Firstly, starting from the inherence of receptive field and considering the characteristic of low-level information loss in forward propagation, inputs in different sizes are used to construct a “multiscale hybrid compensation network”, which can achieve the information compensation and HSI classification using multiscale information by diversifying features.-

Furthermore, in order to extract spatial contextual information (SCI) which often ignored by previous literatures. A SCI perception module is proposed with the help of 1-D sequence model RNN and successfully solve the existing problem that difficult to maintain the spatial structure in the past time researches. In addition, the “hybrid information perception network” that utilizes various spectral and spatial information and achieves the end-to-end HSI classification is built on the foundation of multiscale feature extraction.

Then, in the light of the phenomenon that previous multiscale feature extraction works tend to operate on spatial aspect while progress slowly in spectral dimension. Two schemes that separately based on the wavelength or receptive field are proposed in spectral feature extraction and another 1-D sequence model LSTM is employed to obtain the multiscale spectral contextual information, increasing the spectral classification accuracy effectively. In the spatial feature extraction, we expand the 1-D alternating grouping strategy and develop the multidimensional alternating grouping strategy. A method that extracts SCI in a larger range is successfully exploited using the multidimensional sequence model ConvLSTM.

In addition, aiming to combine the extracted multiscale spatial and spectral contextual features so as to generate the more effective multiscale spectral-spatial joint

features, this article proves the shortage of current fusion methods in mathematics. Considering the different importance of spectral and spatial feature extraction parts, a novel adaptive feature fusion method is proposed, empowering the model stronger data representation capability. Finally, an end-to-end “adaptive spectral-spatial multiscale network” that extracts multiscale spectral-spatial contextual information for HSI classification is developed.

Experiment results show that the proposed methods achieve higher accuracies compared with other state-of-the-art HSI classification methods such as 3-DCNN and SSUN, and obtain a more discriminative classification maps that possess the more clear outlines and more continuous inside objects because of the spectral-spatial feature extracted by the proposed methods. It's of great significance for related departments that apply HSI in land cover recognition, urban planning, environmental monitoring and so on.

**Key words:** multiscale, contextual information, sequence model, end-to-end, HSI classification.



# 目 录

摘要	I
ABSTRACT	II
第 1 章 引言	1
1.1 研究背景和意义 . . . . .	1
1.2 国内外研究现状 . . . . .	2
1.2.1 应对维数灾难的方式 . . . . .	2
1.2.2 光谱冗余和空间变异问题的传统处理方法及局限 . . . . .	2
1.2.3 深度学习在高光谱分类中的应用 . . . . .	3
1.2.4 多尺度特征的提取 . . . . .	4
1.2.5 空间上下文特征信息的获取 . . . . .	4
1.3 研究思路及创新点 . . . . .	5
1.3.1 研究思路 . . . . .	5
1.3.2 研究创新及特色 . . . . .	6
1.3.3 论文章节安排 . . . . .	7
第 2 章 多尺度信息提取和序列模型的基本理论	8
2.1 多尺度信息提取 . . . . .	8
2.1.1 深度学习的层级学习特性 . . . . .	8
2.1.2 感受野 . . . . .	9
2.2 基本的序列模型 . . . . .	9
2.2.1 一维序列模型 . . . . .	10
2.2.1.1 RNN . . . . .	10
2.2.1.2 LSTM . . . . .	10
2.2.2 多维序列模型 . . . . .	12
第 3 章 多尺度混合补偿网络	14
3.1 数据预处理 . . . . .	14
3.2 多尺度混合补偿网络 . . . . .	16

3.2.1 空间特征提取: MSCN . . . . .	16
3.2.2 空谱融合 . . . . .	19
3.3 实验及分析 . . . . .	20
3.4 章末总结 . . . . .	25
<b>第 4 章 混合信息感知网络</b>	<b>26</b>
4.1 准备工作 . . . . .	26
4.2 SCI 感知模块 . . . . .	27
4.3 混合信息感知网络 . . . . .	29
4.4 实验及分析 . . . . .	30
4.5 章末总结 . . . . .	34
<b>第 5 章 自适应空谱多尺度网络</b>	<b>35</b>
5.1 光谱多尺度子网络 . . . . .	36
5.1.1 方案一: 波长多尺度 . . . . .	37
5.1.2 方案二: 感受野多尺度 . . . . .	37
5.2 空间多尺度子网络 . . . . .	38
5.2.1 特征融合模块 . . . . .	41
5.3 实验及分析 . . . . .	43
5.3.1 数据集描述 . . . . .	44
5.3.2 模型和实验设置 . . . . .	45
5.3.2.1 光谱多尺度子网络设置 . . . . .	45
5.3.2.2 空间多尺度子网络设置 . . . . .	45
5.3.2.3 实验设置 . . . . .	46
5.3.3 光谱多尺度子网络分析 . . . . .	46
5.3.3.1 方案一 V.S. 方案二 . . . . .	46
5.3.3.2 SeMN 中方案二的参数分析 . . . . .	47
5.3.4 空间多尺度子网络分析 . . . . .	48
5.3.4.1 参数估计 . . . . .	48
5.3.4.2 分组策略和输入序列设置 . . . . .	49
5.3.4.3 消融实验 . . . . .	50
5.3.5 自适应空谱多尺度网络分析 . . . . .	50
5.3.5.1 特征融合评价 . . . . .	50
5.3.5.2 分类精度评定 . . . . .	51

5.3.5.3 分类结果图 . . . . .	55
5.4 章末总结 . . . . .	59
<b>第 6 章 模型对比及分析</b>	<b>60</b>
6.1 模型性能比较 . . . . .	60
6.2 模型复杂度比较 . . . . .	66
6.3 章末总结 . . . . .	66
<b>第 7 章 结论</b>	<b>68</b>
<b>参考文献</b>	<b>69</b>
<b>致谢</b>	<b>76</b>



## 第 1 章 引言

### 1.1 研究背景和意义

随着高光谱传感器技术的发展，可以获得具有数百个通道的高光谱图像，因此在许多领域都有着十分重要的作用。对高光谱图像进行土地覆盖分类一直是近年来的热点问题，它的目标是对高光谱图像中的每一个像元赋予一个唯一的语义标签，进而生成一幅精确完整的分类图。这种分类图可以为农业 [1]，环境监测 [2]，目标探测 [3] 等行业的决策者提供指导。然而，由于高光谱图像本身所存在的复杂的光谱和空间结构，时至今日，这一任务仍然具有一定挑战性。

概括来说，高光谱图像分类任务主要存在以下问题：

- 维数灾难

传统的分类方法诸如 K 近邻 [4]，对数几率回归 (Logistics Regression, LR) [5] 等，直接将原始光谱向量输入到分类器中。这意味着分类器将会在高维特征空间直接处理这上百个波段的特征。然而，实际应用中高光谱图像中的样本数量往往有限，有限的样本在高维空间中分布的极为稀疏，并且这种稀疏程度随特征维数增加而增大，这对分类器判别样本的类别造成了困难，从而造成了分类精度的下降，

- 像素间光谱冗余

高光谱图像不同像素间的许多波段中的光谱曲线表现相似，这不仅对分类没有帮助，还给分类器判断不同类别的地物造成了困难。

- 空间变异

受到周围环境的影响，高光谱图像中的相同地物可能具有截然不同的光谱信息，即同物异谱现象。

如何克服上述问题达到高分类精度并获得高质量的分类结果图目前已成为高光谱分类领域的主要工作目标。

## 1.2 国内外研究现状

### 1.2.1 应对维数灾难的方式

应对高光谱图像中维数灾难最直接的方式就是降维，使得分类器在较低维度下处理特征。然而将高维特征转换到低维势必会损失信息，为了尽可能保留有用的信息，有两种降维手段。第一种称为特征选择，顾名思义，就是选择最有用的特征，舍弃不用那些无用的特征，这个有用与无用则是通过人为设计一些判别准则来确定的 [6]。另一种称为特征提取，就是将原始高维数据投影到低维空间，其目的是保留最有价值的信息，常用的手段包括主成分分析（Principal Component Analysis, PCA）[7] 和独立成分分析（Independent Component Analysis, ICA）[8] 等。

### 1.2.2 光谱冗余和空间变异问题的传统处理方法及局限

对于光谱冗余和空间变异等问题，多特征融合是一种可行的解决手段，其旨在向分类器提供更多样化的特征从而使得更容易的进行判别。因此除了地物本身的光谱向量，通过目标像元周围环境得到的空间信息后来也被人们重视起来，并进一步衍生出了许多联合利用光谱和空间信息进行分类的框架。一种思路是通过特征工程进行，例如基于连续开闭操作的形态学剖面 [9,10]，借助于超像素 [11–13]，稀疏表示 [14,15] 以及多种滤波器工具 [16–19]，设计制造出具有判别力的空间和光谱特征，从而完成分类过程。另一种思路则是先进行预分类得到一个初步分类图，此过程通常采用支持向量机（Support Vector Machine, SVM）分类器，然后通过马尔科夫随机场正则化逐步优化先前的分类结果 [20,21]，这是一个不断提升分类质量的过程。此外，也有一些基于其他机器学习技术例如主动学习，流形学习，循环信念传播 [22] 等进行高光谱分类的方法被提出。

同时采用光谱和空间特征相当于扩大了特征的广度，而提取特征的有效程度同样值得注意。上述提到的传统分类方法，尤其是特征工程思路中产生的特征是人工设计的产物，其有效性很大程度上依赖于设计者的经验和专业程度。另外，这些特征事实上属于浅层特征因为产生这些特征的模型是浅层模型，例如线性 SVM 和 LR 是单层分类器，而径向基核 SVM 和随机森林则可以被看作是双层模型 [23]。这些通过浅层模型获取的浅层特征往往难以反映出目标的本质属性，此外这些浅层特征在复杂条件

下也不够稳定 [24]，这意味着仅凭这些特征很难在不同环境下识别出相同的地物。

### 1.2.3 深度学习在高光谱分类中的应用

近年来，利用深度神经网络 (Deep Neural Network, DNN) 自动获取出目标深层特征的“深度学习”方法逐渐显示出其强大的威力，其不仅在计算机视觉领域崭露头角并逐渐占据主流，也被应用于高光谱分类领域并且取得了非常好的效果。2014 年，一种典型的 DNN，栈式自编码机 (Stacked Auto Encoder, SAE)，被首先用于高光谱分类 [25]。该模型是通过不断的应用编码-解码过程，重构输入数据，然后再去掉解码层以表示已经得到了合适的编码层来提取特征，从而逐渐拓展网络深度，其最终得到的网络具有提取深层特征的能力。不过，不管是 SAE，还是后来也被应用于高光谱分类的深度信念网络 [26]，它们都是通过堆叠全连接层而成的，这意味着很大的计算量，此外，由于这些网络的输入是一维向量，这意味着提取空间特征时需要将二维乃至三维的矩阵块展平成一维向量，这无疑损失了空间结构信息，并且是一个无法弥补的损失，因此 DNN 后来逐渐淡出了人们的视野。与此同时，另一种深度网络，卷积神经网络 (Convolutional Neural Network, CNN) 则由于其具有局部感知和权重共享的良好特性，在提取特征时不破坏原有的空间结构，因此得到了人们的广泛关注，并且已经有大量基于 CNN 的工作问世 [27–32]。除了将高光谱图像的像元作为光谱向量处理以外，因为高光谱图像中包含数百个波段，这些波段又分布在狭窄的波长范围内，这意味着相近波段之间可能会存在关联，这一思想催生了将高光谱图像作为序列数据进行处理的一系列模型。此类模型旨在提取上下文信息来丰富提取到的特征，通常这些模型以循环神经网络 (Recurrent Neural Network, RNN) 及其变体结构例如长短期记忆网络 (Long Short-Term Memory, LSTM) 以及门控循环单元 (Gated Recurrent Unit, GRU) 为基础因为这些结构的序列化建模能力已在自然语言处理领域得到证实因为它们可以准确捕捉长句子的时序变化 [33–35]。借鉴于这一思想，如果将高光谱图像的像素看作反映了地物反射率随波长不断变化的序列数据，就可以利用这些模型来对像素的光谱变化建模，代表性的工作有 [36–38] 等。例如，[37] 和 [36] 分别采用具有  $T_b$  个时间步的 LSTM 和 RNN， $T_b$  是高光谱图像的波段数。[39] 将光谱向量分成若干组，然后按顺序将它们输入 LSTM，减轻了 LSTM 提取信息的困难。[40] 则采用级联 RNN 进一步获取 RNN 层之间的关系。在高光谱图像分类领域，光谱信息的提取仅需利用单个像素

在不同波长的反射率值，而空间信息的获得则需要同时采用不同的像素，通常相应模型的输入是以目标像元为中心进行开窗得到的特定尺寸的图像块，这一图像块叫做补丁。然而，如果想要挖掘补丁中像素之间的联系，却不能像单个像素那样将具有多个维度的补丁不能直接输入这些序列模型中，因为这些模型往往需要一维的输入值。为了解决这一问题，[37] 在行方向上切分了空间补丁的第一主成分，然后把这些行送入 LSTM。[40] 和 [38] 则分别使用了 RNN 和 GRU 从经过 CNN 或全连接层得到的特征间获取空间信息。此外还有一些工作混合利用了 CNN, RNN 及一些其他的结构等等，如 [29, 39] 等。

#### 1.2.4 多尺度特征的提取

在早期的研究工作中 [25, 31]，最终提取到的特征是单一尺度的，这意味着输入分类器的特征图中每一个像素的感受野大小均相同，这限制了分类精度的提高，而地物不论怎样缩放其类别应该是保持不变的，因此不同尺度特征应该被考虑在内，这有助于提高模型的鲁棒性，并且类似的思想已经被广泛的应用于计算机视觉领域的训练集增强中。有许多工作提出了获取多尺度特征的方法，最自然的想法分别采用不同尺寸和步长的卷积和池化操作并行处理，如 Inception 结构 [41–43], ASPP 结构 [44–46]。2018 年，一种新颖的多尺度 CNN 被用于高光谱分类中 [39]。考虑到深度学习的层级学习特性，CNN 中的串联的特征图的感受野应该是不断递增的，作者认为这自然而然带来了不同尺度的特征，于是其将这些特征图合并，从而获得了多尺度特征，并最后进行分类，取得了良好的分类效果。事实上，特征的尺度，也就是感受野，对应于原始图像上的覆盖范围。因此，多尺度可以通过不同尺寸的输入来控制。

#### 1.2.5 空间上下文特征信息的获取

在深度学习发展起来以后，已经有大量工作利用空谱融合网络进行高光谱分类，然而，大部分方法在空间部分单纯依赖 CNN 局部感知的能力。在 CNN 的卷积操作中，相邻像素值是利用点乘并行处理的，这意味着 CNN 难以感知到相邻像元间存在的关系信息，这些信息也就是空间上下文信息。在  $3 \times 3$  大小的补丁上提取空间上下文信息是最简单的一种方式，论文 [47] 试图利用 RNN 提取  $3 \times 3$  补丁的空间上下文信息，其输入序列通过周围像素与中心像素在特征空间的欧氏距离的来确定，然而这样却忽视

了像元间存在的空间结构。

此外，前述工作在利用序列模型处理空间数据时常常需要把多维数据转成 1 维，此时高光谱图像中的原始空间结构信息将不可避免的损失掉。如果使得每个时间步的输入都是一维，那么待操作数据的尺寸也不会特别大（否则需要过多的时间步及大量的计算）。换句话说，输入尺寸的限制使得序列模型仅能够提取像素级或者小区域的空间上下文信息。事实上，之所以输入的尺寸有所限制是因为这些序列模型中的隐藏状态往往是全连接层。为了在更大的区域提取空间序列信息同时保持输入数据的空间结构。[48] 在 LSTM 的基础上通过将控制门中的全连接层替换成卷积层，从而提出了卷积 LSTM(Convolutional LSTM, ConvLSTM)。已经有一些工作将 ConvLSTM 应用于高光谱图像分类中 [49–51]。例如，[49] 将 2 维或 3 维卷积与 ConvLSTM 结合去捕捉 CNN 特征图不同通道间的关系。双向 ConvLSTM 则被 [50] 所使用去提取高光谱图像的空谱特征从而对高光谱图像进行分类。

## 1.3 研究思路及创新点

### 1.3.1 研究思路

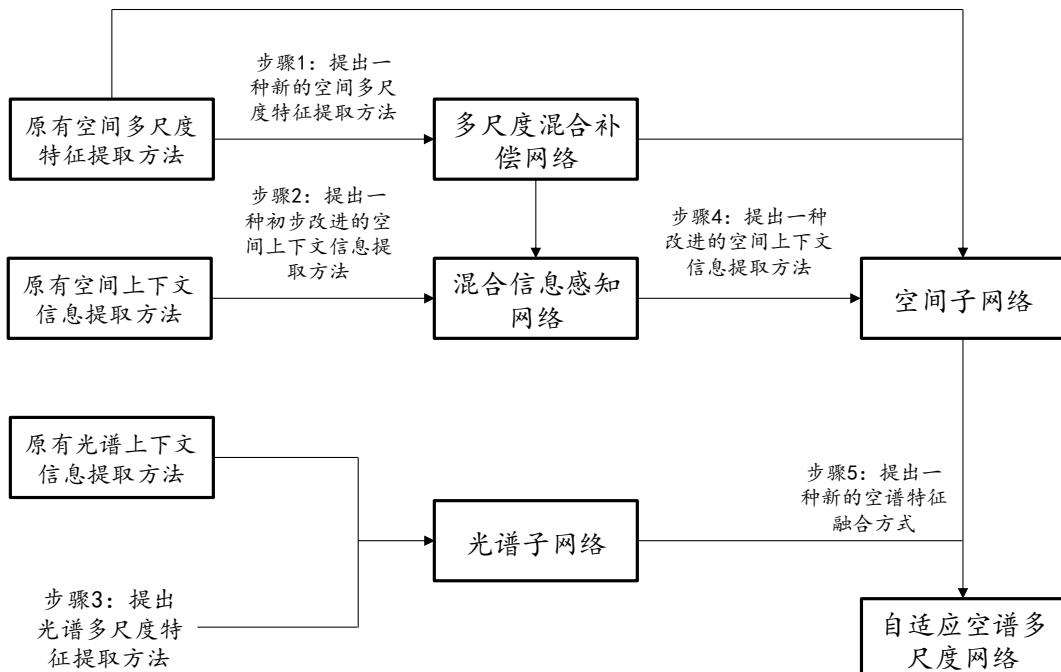


图 1.1: 本文研究思路

图1.1展示了本文的研究思路：首先，受上述工作启发，本文提出了一个不同的获取多尺度特征的思路：利用以目标像素为中心得到的不同尺寸的空间补丁，通过一个跳跃连接模块来多样化深度网络提取到的特征，以增强网络的表示能力。考虑到这个模块具有信息补偿的特性，这个网络也被叫做“多尺度混合信息补偿网络”。

紧接着，在通过多尺寸空间补丁块提取多样化特征以保证网络基本能力的基础上，为了同时兼顾空间上下文信息提取与原始结构保持，本文设计了一个新颖的空间上下文提取模块，进而构建了一个能够综合提取多种信息的“混合信息感知网络”。

上述工作分别在多尺度特征提取和空间上下文信息提取两方面进行了初步尝试，其仍然存在许多方面需要改进，例如：“混合信息感知网络”中的空间上下文信息提取模块依然是在 $3\times 3$ 大小的空间补丁上提取得到的，可以想见，其对分类精度虽然有所提升，但是提升仍然较为有限。此外，在该模块中空间结构仍有所损失，因为其中区域级上下文信息的提取需要将每个时间步的输入矩阵展平成一维向量，这也损失了一部分结构信息。

为了获得更强力的上下文特征同时保证所获取特征的尺度多样性。在通用空谱融合网络框架结构的基础上，本文在光谱和空间两方面分别构思了两个子网络：在光谱方面，我们提出了新的光谱多尺度特征提取方案，并结合先前的光谱上下文信息提取方法，构建了光谱子网络；在空间方面，我们不仅利用多尺度特征保证分类精度，还提出了新的空间上下文信息提取方法，这种方法可以作用于更大尺寸的空间块，并进而开发出了空间子网络。这两个子网络均具有获取有效多尺度上下文特征的能力。最后，我们提出了自适应特征融合模块将空谱两子网络得到的特征更有效的结合起来，进一步增强网络的表征能力。由于整个过程完全是端到端的，其也可以被视为一个更大的分类网络，在后续内容中将其称为“自适应空谱多尺度网络”。

### 1.3.2 研究创新及特色

关于空间多尺度特征和光谱上下文信息的提取方法已经较为成熟。然而，由于光谱信息仅由一个一维向量表示，光谱多尺度特征的提取仍有一定困难，且现有相关研究较少。此外，空谱融合框架中的空间补丁作为一个多维矩阵，一般的一维序列模型诸如RNN, LSTM及GRU等，仅能提取像素级上下文信息，进行区域级上下文信息的提取很容易损失空间结构，并丢失空间信息，从而影响分类精度。如何在保持原有

多维数据结构的同时提取有效的空间上下文信息进行高光谱图像分类是一个亟待解决的难题。

致力于改善上述问题，本研究作出如下创新：

- 1) 本文提出一种基于多尺寸空间补丁的高光谱图像分类网络，该模型能够以更新颖的方式获取多尺度特征。
- 2) 本文提出了一种新的在小尺寸补丁上提取空间上下文信息的方法。不同于前述忽视乃至破坏原有空间结构的方法，本文的方法不仅提取到了多级上下文信息，同时保持了原有的空间结构。
- 3) 以往多尺度特征的提取往往针对空间方面，光谱特征的有效性往往不受重视，这使得分类精度往往很高，而分类结果图却由于过于重视空间特征产生过平滑现象从而并不十分理想。本研究针对光谱向量通过专门设计的光谱子网络提取多尺度特征，并通过实验展现了提取到的光谱特征对分类质量的带来的影响。另外，不同于 1) 2)，本研究同时还在借助多尺度特征提取模式保证分类精度的基础上进一步探索，试图在大尺寸补丁上获取更有效的空间上下文信息，并由此开发出了空间子网络。最后，本研究提出一种新的特征融合模块用来融合上述空间和光谱子网络提取到的特征。该模块能够自适应学习同时赋予网络更强的表示能力。理论证明和实验结果均展现了所提出模块的相比于以往特征融合方法的优越性。

### 1.3.3 论文章节安排

第一章是引言，给出了本文研究背景、研究现状、研究创新并梳理一下研究脉络；第二章介绍了多尺度信息提取和序列模型的基本理论框架；第三章主要介绍提出的“多尺度混合补偿网络”及信息补偿模块，第四章则对应于“混合信息感知网络”和空间上下文信息提取模块。包含“光谱子网络”、“空间子网络”以及提出的特征融合模块的“自适应空谱多尺度网络”将在第五章被详细讨论。三种方法各自的实验部分均已包含到对应的章节中。紧接着，第六章对三四五章中各自最有力的模型进行了进一步的综合性对比。最后，第七章总结全文。

## 第 2 章 多尺度信息提取和序列模型的基本理论

### 2.1 多尺度信息提取

#### 2.1.1 深度学习的层级学习特性

机器学习即计算机能够自己获取知识，自动从原始数据中提取模式的能力。简单的机器学习算法很大程度上依赖于输入数据的表示，这种表示就是提供给机器的特征。在深度学习发展起来之前，有大量的工作从特征工程方面入手，即如何手工设计出更有效的特征，这种方法的确解决了许多人工智能任务。然而，这种方法不仅费时费力，还依赖于设计者本身的经验与专业水平。另外，仍有一些任务无法显式的构造出特征。例如根据车轮识别汽车，此时很难通过像素值来描述车轮。此外，因场景而异的图像也加剧了识别的难度。

解决这种问题的另一种思路是不仅仅要把表示映射到输出，更要利用机器学习算法挖掘表示本身。为此，表示学习应运而生。其目的在于自动获取有效的特征，这种特征往往比手动设计表现的要好。表示学习几乎不需人工干预，就能使计算机快速的适应新的人工智能任务。

然而，有许多任务，其被机器所能理解到的表示需要接近人类水平的理解，而简单的机器学习算法仅能自动提取到输入数据的底层特征，例如仅有一个隐层的自编码机。这意味高层语义信息与提取到的底层特征间存在着语义鸿沟。

深度学习让计算机通过较简单的概念来构建复杂的概念，通过简单表示表达复杂表示，解决了表示学习中的核心问题 [52,53]。假设有一个系统  $F$ ，其有  $n$  层，输入是  $I$ ，输出是  $O$ ，可以表示为

$$I \Rightarrow F_1 \Rightarrow F_2 \cdots \Rightarrow F_n \Rightarrow O \quad (2.1)$$

倘若  $O = I$ ，即输入  $I$  经过该系统无任何信息损失（实际上不可能发生，因为  $I(F_i, F_{i+2}) \leq I(F_i, F_{i+1})$ ， $I(\cdot)$  是互信息），即输入  $I$  在每层  $F_i$  均没有信息丢失，即在任意一层  $F_i$ ，其均为输入  $I$  的另一种表示。现在令系统  $F$  通过学习得到的输出

$O$  仍为输入  $I$ , 则表明通过该系统自动化的提取到了输入  $I$  的一些列层次化的表示  $F_1, F_2, \dots, F_n$ 。

深度学习正是通过堆叠多个层, 将前一层的输出作为下一层的输入, 使得信息得到了分层表达。从而完成了层级化学习。

### 2.1.2 感受野

感受野是 CNN 中每一层特征图上的像素在输入图像上映射的区域大小, 即特征图上一个点对应于输入图的区域。感受野的计算十分简单, 一种是从前往后顺推, 另一种则是倒推, 其中顺推公式为

$$r_k = r_{k-1} + (f_k - 1) \cdot \prod_{i=1}^{k-1} s_i \quad (2.2)$$

上式中,  $r_k$  是第  $k$  层的感受野,  $f_k$  则是第  $k$  层的卷积核大小或池化尺寸大小 ( $k = 1, 2, \dots$ ),  $s$  是步长。第 0 层是输入层, 初始化  $r_0 = 1, r_1 = f_1$ 。

逆推公式则从后向前进行: 假设一个  $n$  层的网络,  $r_k$  是第  $n$  层像素在第  $k$  层的感受野,  $f_k, s_k$  分别是第  $k$  层到第  $k+1$  层的滤波器尺寸和步长, 则有

$$r_k = (r_{k+1} - 1) \cdot s_k + f_k \quad (2.3)$$

$r_1$  即为所求第  $n$  层特征图上像素的感受野。

借助于深度学习的层级学习特性, 信息在 CNN 中分层表达的过程中, 能够产生出许多具有不同感受野的特征, 而感受野的大小表明了特征的感知范围, 即特征的尺度。感受野这一概念将贯穿本文始终, 特别的, 第三章介绍了一种能够构造出同时具有不同感受野的特征的方法, 实现了多尺度特征提取。

## 2.2 基本的序列模型

基于 RNN 构建的模型的序列建模能力已经在 NLP 领域被证实, 并对高光谱分类领域, 尤其上下文信息提取方面具有深远影响。本节将对 RNN 及其变种 LSTM 进行简要介绍, 因为这些模型只能处理一维数据, 因此它们都是一维序列模型。

## 2.2.1 一维序列模型

### 2.2.1.1 RNN

RNN 通常被用来处理序列数据，尤其是在时序数据处理方面，因为它可以完成从过去时刻到当前时刻的信息转换并且不断地更新神经元，因此 RNN 具有捕捉上下文信息的能力。假设有一组序列数据  $x = \{x_1, x_2, \dots, x_i, \dots, x_t\}$ ,  $x_i$  是  $i$  时刻的特征向量，则 RNN 的计算过程为

$$\begin{aligned} h_t &= g(W_{hh}h_{t-1} + b_{hh} + W_{hi}x_t + b_{hi}) \\ o_t &= W_{oh}h_t + b_{oh} \end{aligned} \quad (2.4)$$

$h_t, x_t$  和  $o_t$  分别是时刻  $t$  的隐藏状态，输入和输出值。 $(W_{hh}, b_{hh})$ ,  $(W_{hi}, b_{hi})$  和  $(W_{oh}, b_{oh})$  则是隐层到隐层、输入到隐层以及隐层到输出的权值矩阵和偏置。 $g$  是激活函数，在 RNN 中通常采用线性激活单元 (Rectified Linear Unit, ReLU)  $g_r(x) = \max(0, x)$  或双曲正切函数  $g_t(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ 。本文中， $h_0$  的初始化隐藏状态均为 0。

### 2.2.1.2 LSTM

虽然 RNN 已经应用于许多序列建模方面。然而，当处理长序列数据时，它在反向传播的过程中仍然有梯度爆炸或梯度消失的潜在风险 [54]。因此，[55] 提出了 LSTM。凭借着门控机制和记忆单元，LSTM 能够有效的捕捉序列中相邻数据的上下文信息。LSTM 的结构如图2.1

可以看到 LSTM 的单元利用三个控制门不断的更新着隐藏状态  $h$  和细胞状态  $c$ ，这三个门包括输入门  $i$ ，遗忘门  $f$  和输出门  $o$ ，分别用来丢弃、保留或者放大信号，从而达到信息控制和转换的效果。在时间  $t$  时刻一个 LSTM 单元的计算过程为

$$\begin{aligned} i &= F_s(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci} \otimes c_{t-1} + b_i) \\ f &= F_s(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf} \otimes c_{t-1} + b_f) \\ g &= F_h(W_{xg}x_t + W_{hg}h_{t-1} + b_g) \\ c_t &= i \otimes g + f \otimes c_{t-1} \\ o &= F_s(W_{xo}x_t + W_{ho}h_{t-1} + W_{co} \otimes c_t + b_o) \\ h_t &= o \otimes F_h(c_t) \end{aligned} \quad (2.5)$$

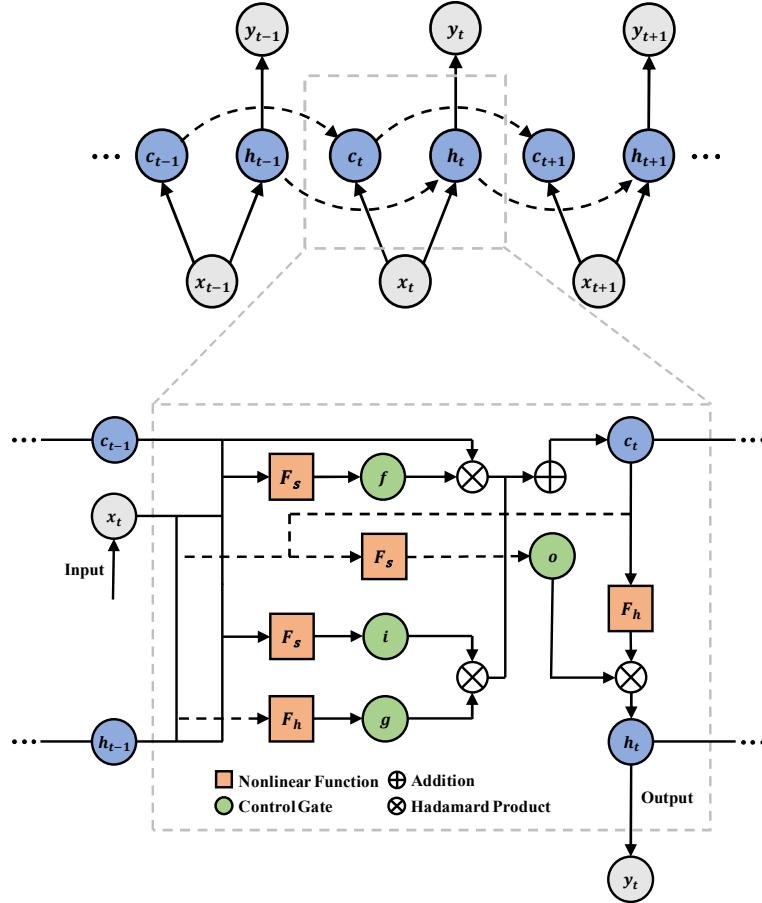


图 2.1: LSTM 的结构。

上式中  $x_t$ ,  $h_t$  和  $c_t$  分别表示  $t$  时刻的输入值, 隐藏状态和细胞状态。

$\{(W_{\odot i}, b_i), (W_{\odot f}, b_f), (W_{\odot o}, b_o), \odot = x, h, c\}$  是输入门  $i$ , 遗忘门  $f$  和输出门  $o$  的权值矩阵和偏置向量。 $\{(W_{\odot g}, b_g), \odot = x, h\}$  是过渡值  $g$  的权重和偏置。我们用  $F(\cdot)$  表示非线性映射, 通常采用 sigmoid 函数  $s(x) = 1 / (1 + \exp(-x))$  和双曲正切函数  $\tanh(x) = (\exp(x) - \exp(-x)) / (\exp(x) + \exp(-x))$ , 并分别表示为  $F_s$  和  $F_h$ 。“ $\otimes$ ”是哈达玛积, 是逐像素乘法。

LSTM 可以很好的捕捉光谱域中相邻波段存在的上下文信息。但是它依然受限于空间上下文信息的提取。虽然 [37] 已经尝试将 LSTM 用于获取空间信息, 然而, 由于输入值形状的限制, 原始图像不得不展平为 1 维向量, 这样数据固有的空间结构不可避免的被破坏掉了。因此, 有必要开发一种不限尺寸的多维输入模型实现空间上下文信息的提取。

### 2.2.2 多维序列模型

上文提到 LSTM 的弱点主要是源于其中的全连接层，它限制了输入数据的尺寸。最近，由于局部感知能力和同时保持相邻像素空间关系的优良特性，卷积操作在高光谱分类领域被广泛的用于空间信息编码。因此，自然能够想到把卷积核和 LSTM 结合起来以提取多维数据的上下文信息。这个想法已经被香港理工大学的施等人 [48] 实现，他们构建了 ConvLSTM 并用于降水预测。

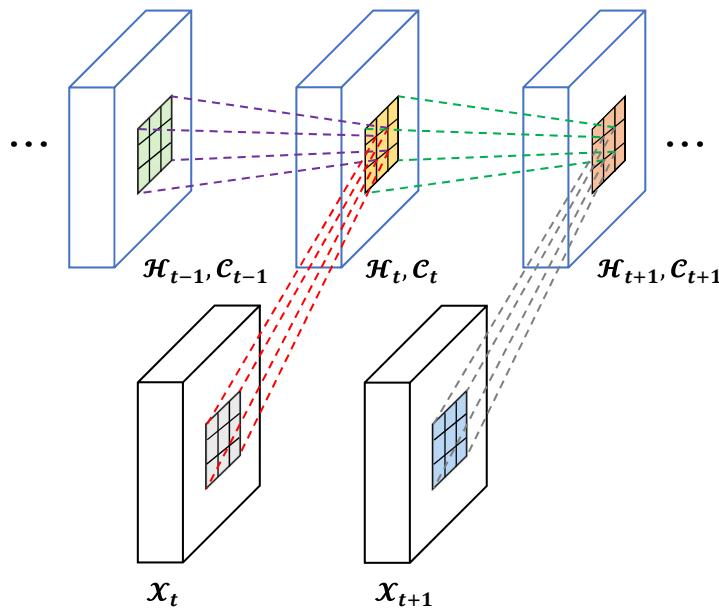


图 2.2: ConvLSTM 的示意图。

图2.2展示了 ConvLSTM 的结构，与 LSTM 相比，ConvLSTM 的输入、隐藏状态和记忆单元均为多维数据，并且整个过程通过卷积操作实现。因此，ConvLSTM 可以对空间序列数据建模。 $t$  时刻 ConvLSTM 的计算过程可以公式化为

$$\begin{aligned}
 i &= F_s(W_{xi} * \mathcal{X}_t + W_{hi} * \mathcal{H}_{t-1} + W_{ci} \otimes \mathcal{C}_{t-1} + b_i) \\
 f &= F_s(W_{xf} * \mathcal{X}_t + W_{hf} * \mathcal{H}_{t-1} + W_{cf} \otimes \mathcal{C}_{t-1} + b_f) \\
 g &= F_h(W_{xg} * \mathcal{X}_t + W_{hg} * \mathcal{H}_{t-1} + b_g) \\
 \mathcal{C}_t &= i \otimes g + f \otimes \mathcal{C}_{t-1} \\
 o &= F_s(W_{xo} * \mathcal{X}_t + W_{ho} * \mathcal{H}_{t-1} + W_{co} \otimes \mathcal{C}_t + b_o) \\
 \mathcal{H}_t &= o \otimes F_h(\mathcal{C}_t)
 \end{aligned} \tag{2.6}$$

其中  $\mathcal{X}_t, \mathcal{H}_t, \mathcal{C}_t$  表示  $t$  时刻多维的输入、隐藏状态和细胞状态。“ $*$ ”是卷积操作，其他

变量和符号与标准的 LSTM 具有类似的含义。很显然 ConvLSTM 将原始的矩阵乘法替换为了卷积操作并保持了原有的门口机制，这提供了数据在层内和层间的交流和转换。因此，ConvLSTM 可以用来对多维序列数据建模，尤其是空间信息。

本文中 RNN 主要用于第四章的空间上下文信息感知模块，LSTM 则被第五章的光谱子网络所采用以提取相邻波段间的序列信息。第五章的空间子网络则灵活的借助了 ConvLSTM 以实现上下文信息的获取。

## 第 3 章 多尺度混合补偿网络

在传统的 CNN 中，随着信息在网络中的不断传播，高层次语义信息不断丰富的同时，底层的细节信息却逐渐消失掉了，并且这一现象随着网络深度的加深而愈发严重。在高光谱分类网络中，特征图上携带的对应于原始输入空间补丁的信息可能会在中途卷积、池化或其他操作中损失，这可能会影响后续的分类质量。另外，正如前文提到的，传统 CNN 的识别能力通常被提取到特征的单一尺度所限制。

为了解决上述问题，本章主要介绍两个基于 CNN 的多尺度空间特征提取端到端模型：多尺度补偿网络 (Multi-Scale Compensation Network, MSHCN)、多尺度混合补偿网络 (Multi-Scale Hybrid Compensation Network, MSHCN)，其多尺度源自于所利用的不同尺寸的空间补丁和在信息传播不同阶段的操作。

### 3.1 数据预处理

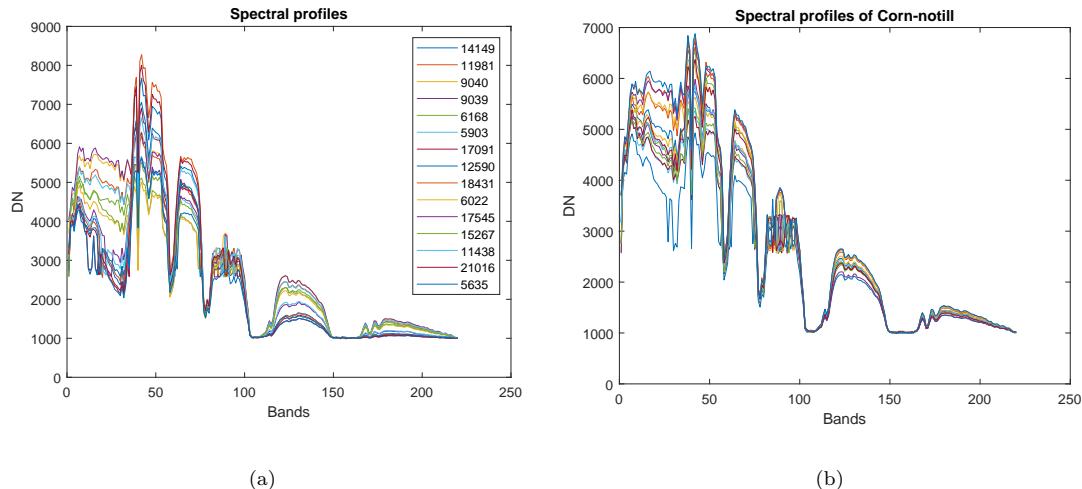


图 3.1: 典型的高光谱图像地物光谱曲线: (a) 随机挑选的不同地物 (b) 同类地物

高光谱图像的光谱曲线往往存在许多微小的扰动，这些扰动可以看做是噪声。图3.1展示了两组从 Indian Pines 数据共 21025 个像素中挑选的 15 个像素的光谱曲线，(a) 是随机挑选的不同地物，图例是该 15 个像素的编号，而 (b) 则是从 Corn-notill 类中随

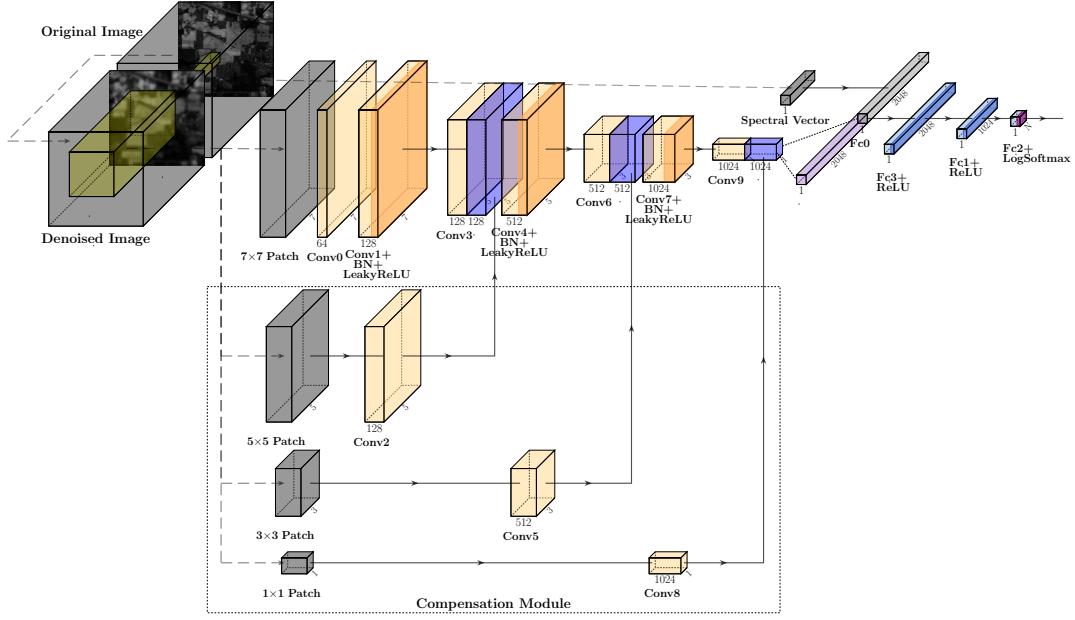


图 3.2: 多尺度混合信息补偿网络

机挑选的 15 个像素。可以观察到不管是同类还是异类地物，这些噪声使得光谱曲线不够平滑，且这些光谱曲线往往表现出相似的模式，这意味着光谱向量存在很大的冗余，这些冗余不仅对分类没有帮助，其在同类地物上造成同物异谱现象，并可能对分类器造成误导。因此有必要对图像进行增强处理。最典型的技术是进行滤波操作，已经有多种滤波器被应用于高光谱分类中，如引导滤波 [17, 18]，边缘保持滤波 [19] 等等。

受上述工作的启发，在预处理过程中，我们首先对图像进行高斯平滑消除细微噪声，其依赖于利用一种离散化的窗口进行滑动卷积，该窗口被称为高斯模板，其计算公式为

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \quad (3.1)$$

其中， $x, y$  分别是模板像素坐标相对于模板原点在  $X, Y$  轴的差值， $\sigma$  是标准差。此外，在进行高斯平滑前，原始高光谱图像需要在边界进行对称填充操作以便于模拟边界像素周围的实际情形。

## 3.2 多尺度混合补偿网络

### 3.2.1 空间特征提取：MSCN

图3.2展示了本文提出的整体框架。为了更充分的使用空间信息，包括地物本身的细节和周围的环境信息。目标像素周围不同尺寸的空间补丁块被同时使用进行像素级分类。我们主要运用了  $7 \times 7$ ,  $5 \times 5$ ,  $3 \times 3$  和  $1 \times 1$  四种尺寸的空间补丁，即选取目标像素周围的一至三圈像素。不过，考虑到高光谱图像的空间分辨率，像素间语义关联随着距离加大的不断减弱，这可能对分类带来不利影响，因此我们不使用更多的像素。

归一化后的  $7 \times 7$  大小的补丁首先被输入网络，其第一层是用于降维和跨通道信息交互 [56] 的一个  $1 \times 1$  卷积层，这种卷积可以去除高光谱图像中的冗余数据同时也减小了计算复杂度。紧接着， $3 \times 3$  卷积层作为特征提取器，批标准化 (Batch Normalization, BN) [57] 和非线性激活函数被用分别来避免过拟合现象的发生和增强模型的数据表示能力。

考虑到前向传播时的细节信息丢失，我们在通道维连接了含有原始细节信息的  $5 \times 5$  空间补丁与此时的特征图。此外，由于需要做尺寸匹配和信息集成， $1 \times 1$  和  $3 \times 3$  的卷积操作被分别应用到  $5 \times 5$  的空间补丁和与之连接的特征图。这个看似简单的操作不仅通过添加细节信息改善了提取到的特征，还使得合并后的特征具有了多种尺度因为此时特征图上同时具有了不同感受野的像素。接下来的步骤和先前的操作相同： $3 \times 3$  和  $1 \times 1$  的补丁在不同的阶段被分别输入网络中，这也提供了关于目标像素更加多层次的信息并且提升了所提取的特征的多样性。在这个过程中，半监督学习策略也被应用到因为没有标记的数据也在这些输入的补丁中。整个的编码过程持续到网络中的特征图尺寸变为  $1 \times 1$ ，然后这个特征图将会与原始的  $1 \times 1$  补丁连接。这个融合后的特征将会被展平并送入全连接层。接着多层感知机 (Multilayer Perceptron, MLP) 和对数 softmax 函数被用来生成最终的分类结果。在全连接层中，随机丢弃策略 (Dropout) [58] 被用来应对过拟合。本文所提出方法的具体流程如下。

假设输入的  $7 \times 7$  补丁为  $x_{77}$ ，相应的， $x_{55}, x_{33}$  和  $x_{11}$  则分别表示归一化后的  $5 \times 5$ ,  $3 \times 3$  和  $1 \times 1$  原始补丁，并且所有补丁被逐通道的归一化到  $[-1,1]$ 。另外，为了保留源自  $[-1,1]$  归一化后的负值，我们提出网络中的卷积层中使用 LeakyReLU，这可

能对分类有积极影响。如果用  $f_L(\cdot)$ , 则  $f_L(x) = \max(0, x) + \alpha \min(0, x)$ ,  $\alpha$  通常是一个很小的数, 本文设置  $\alpha = 0.01$ 。设  $W_{ij}^{(0)}(\text{"(0)"} \text{ 对应于图3.2中的 "Conv0", 下同})$  表示  $n_0$  个卷积核中的第  $i$  个的第  $j$  层, 则  $W^{(0)}$  卷积后的结果为

$$F_1(x_{77}, W^{(0)}) = W^{(0)} * x_{77} \quad (3.2)$$

其中  $F_1(x_{77}, W^{(0)})_i = \sum_j W_{ij}^{(0)} * x_{77}^j$ , “\*”是卷积操作,  $F_1(x_{77}, W^{(0)})$  有  $n_0$  层特征图,  $F_1(x_{77}, W^{(0)})_i$  是  $F_1(x_{77}, W^{(0)})$  的第  $i$  个波段,  $x_{77}^j$  是  $x_{77}$  的第  $j$  个通道。接着, 经过之后  $n_1$  个  $3 \times 3$  卷积核的特征图可以通过计算

$$F_2(x, W^{(1)}, b^{(1)}) = f_L(B(W^{(1)} * x + b^{(1)})) \quad (3.3)$$

得到。上式中,  $x = F_1(x_{77}, W^{(0)})$ ,  $F_2(x, W^{(1)}, b^{(1)})_i = f_L\left(B\left(\sum_j W_{ij}^{(1)} * x_j + b_i^{(1)}\right)\right)$ ,  $x_j$  是  $x$  的第  $j$  个通道。 $F_2(x, W^{(1)}, b^{(1)})$  包含  $n_1$  个特征图,  $F_2(x, W^{(1)}, b^{(1)})_i$  是  $F_2(x, W^{(1)}, b^{(1)})$  的第  $i$  个波段,  $b_i^{(1)}$  是这  $n_1$  个卷积核中第  $i$  个滤波器的偏置和  $b^{(1)}$  的第  $i$  个数,  $B(\cdot)$  表示了 BN 操作。需要连接的特征  $x_1, x_2$  可以通过  $x_{55}, F_2(x, W^{(1)}, b^{(1)})$  分别与两个卷积层  $W^{(2)}, W^{(3)}$  按照式3.2得到。融合的特征可以表示为  $x_f^1 = C(x_1, x_2)$ ,  $C$  是连接操作。

接下来的步骤和之前提到的相同直到全连接层。在应用  $x_{33}$  和  $x_{11}$  之后, 最终可以得到  $1 \times 1$  大小的连接后的特征  $x_f^3$ , 该特征可以被展平为 1 维向量并送入 MLP 中。为了提高计算效率, 我们将 MLP 中的激活函数设置为 ReLU 而非先前的 LeakyReLU, MLP 的输出可以表示为

$$F_3(x_f^3) = D\left(f_R\left(W_{fc}^{(1)} * D(x_f^3) + b_{fc}^{(1)}\right)\right) * W_{fc}^{(2)} + b_{fc}^{(2)} \quad (3.4)$$

其中  $D(\cdot)$  表示了 dropout,  $f_R(\cdot)$  则是 ReLU 函数, 全连接层中的权值矩阵和偏置被表示为  $W_{fc}, b_{fc}$ 。为了得到最终的分类结果, 如果我们令  $S = F_3(x_f^3)$ , 需要 log softmax 函数

$$P(S)_{ij} = \log\left(\frac{\exp(S_{ij})}{\sum_j \exp(S_{ij})}\right) \quad (3.5)$$

其中,  $S_{ij}$  是样本  $S_i$  的第  $j$  个元素, 损失函数采用负对数似然损失

$$\text{Loss}(S) = -\frac{1}{M} \sum_i \sum_j \mathbf{I}(y_{ij} = 1) P(S)_{ij} \quad (3.6)$$

上式中的  $\mathbf{I}$  是指示函数

$$\mathbf{I}(x) = \begin{cases} 1 & \text{if } x \text{ is True} \\ 0 & \text{otherwise} \end{cases} \quad (3.7)$$

$M$  是训练样本的个数,  $y_{ij}$  是第  $i$  个标签  $y_i$  进行独热编码后的第  $j$  个数。

梯度下降 (Gradient Descent, GD) 已经被广泛的应用于网络的训练。在后向传播的学习过程中, 带有动量的随机梯度下降 (Stochastic Gradient Descent with Momentum, SGDM) 方法被用来加速收敛, 网络中每个卷积核中的权值和偏移通过下式更新

$$\begin{aligned} v_{i+1} &= \beta \cdot v_i + \frac{\partial \text{Loss}(\mathbf{S})}{\partial \theta_i} \\ \theta_{i+1} &= \theta_i - \lambda \cdot v_{i+1} \end{aligned} \quad (3.8)$$

$\beta$  是动量, 本文中  $\beta = 0.9$ 。学习率  $\lambda$  被设置为  $1e-3$ , 并按照多步衰减策略进行衰减。衰减率被设置为 0.5。 $\theta$  代表了网络中每一个卷积核中的每个权值和偏置,  $v_i$  和  $\theta_i$  是第  $i$  个 epoch 的速度和参数。 $\frac{\partial \text{Loss}(\mathbf{S})}{\partial \theta_i}$  可以通过链式法则计算得到, 整体训练过程如算法1。

---

### Algorithm 1 MSCN 的训练

---

```

1: while 训练轮次 < 最大轮次 do
2:    $i \leftarrow 1$ ;
3:   计算  $x = F_1(x_{77}, W^{(0)})$  由式 (3.2);
4:   for  $x$  in  $\{x_{55}, x_{33}, x_{11}\}$  do
5:     计算  $x_{i-1} = F_2(x, W^{(i)}, b^{(i)})$  由式 (3.3);
6:     计算  $x_i = F_1(x_{i-1}, W^{(i+1)}, b^{(i+1)})$  由式 (3.2);
7:     计算  $x_{i+1} = F_1(x_i, W^{(i+2)}, b^{(i+2)})$  由式 (3.2);
8:      $x \leftarrow C(x_i, x_{i+1})$ ;
9:    $i \leftarrow i + 3$ ;
10:  end for
11:  计算  $S = F_3(x)$  由式 (3.4);
12:  计算  $P(S)$  由式 (3.5);
13:  计算  $\text{Loss}(S)$  由式 (3.6), (3.7);
14:  更新  $\theta$  由式 (3.8);
15: end while

```

---

### 3.2.2 空谱融合

在 MSCN 的基础上, 光谱信息也被考虑在内进而构建了 MSHCN。假设 MSHCN 光谱部分的输入是原始或归一化后的光谱向量  $z$ , 然后其通过一个全连接层映射到特定长度, 得到  $z_1 = W_{fc}^{(0)} * z + b_{fc}^{(0)}$ , 这一特征和  $x_f^3$  等长。接着,  $x_f^3$  和  $z_1$  通过  $C(\cdot)$  连接得到  $x_f^4$ , 其可以被送入 MLP 并且可以得到  $T = f_R \left( W_{fc}^{(3)} * D(x_f^4) + b_{fc}^{(3)} \right)$ , MLP 的输出为

$$S_1 = F_3(T) \quad (3.9)$$

$p(S_1)$ ,  $\text{Loss}(S_1)$  和  $\theta$  可以通过计算式 (3.5), (3.6), (3.7) 和 (3.8) 得到.

表 3.1: MSCN 和 MSHCN 的结构参数

网络层	卷积核尺寸	步长/填充	通道数	MSCN	MSHCN
Conv0( $W^{(0)}$ )	$1 \times 1$	1/0	64	✓	✓
Conv1( $W^{(1)}, b^{(1)}$ )	$3 \times 3$	1/1	128	✓	✓
Conv2( $W^{(2)}$ )	$1 \times 1$	1/0	128	✓	✓
Conv3( $W^{(3)}$ )	$3 \times 3$	1/0	128	✓	✓
Conv4( $W^{(4)}, b^{(4)}$ )	$3 \times 3$	1/1	512	✓	✓
Conv5( $W^{(5)}$ )	$1 \times 1$	1/0	512	✓	✓
Conv6( $W^{(6)}$ )	$3 \times 3$	1/0	512	✓	✓
Conv7( $W^{(7)}, b^{(7)}$ )	$3 \times 3$	1/1	1024	✓	✓
Conv8( $W^{(8)}$ )	$1 \times 1$	1/0	1024	✓	✓
Conv9( $W^{(9)}$ )	$3 \times 3$	1/0	1024	✓	✓
Fc0( $W_{fc}^{(0)}, b_{fc}^{(0)}$ )	—	—	2048	—	✓
Fc3( $W_{fc}^{(3)}, b_{fc}^{(3)}$ )	—	—	2048	—	✓
Fc1( $W_{fc}^{(1)}, b_{fc}^{(1)}$ )	—	—	1024	✓	✓
Fc2( $W_{fc}^{(2)}, b_{fc}^{(2)}$ )	—	—	$N^1$	✓	✓

<sup>1</sup> 输出类别数.

表3.1展示了所提出网络的细节化参数, 这些模型都是使用 PyTorch 框架在单张 NVIDIA RTX2080 Ti GPU 上完成的。

### 3.3 实验及分析

本文采用两种公共的高光谱图像分类数据集“University of Houston”(UH)和“Pavia University”(PU)来评估提出的方法。UH 图像大小为  $349 \times 1905$  像素，包含 144 个光谱波段和 15 类。这个数据集也被应用于 2013 年 IEEE GRSS 数据融合大赛。PU 数据集大小为  $610 \times 340$  像素，有 103 个波段和 9 种土地覆盖地物。

我们将提出的方法与 FEFCN-ELM [59], DPRN [60] 和 SSRN [32] 进行比较。在第一个方法中，光谱首先通过大尺寸补丁优化 ( $48 \times 48$ )，然后采用极限学习机 (Extreme Learning Machine, ELM) 分类。DPRN 是建立在 ResNet [61] 的基础上，与 SSRN 有相近的结构。DPRN, SSRN 和我们提出的方法均是基于  $7 \times 7$  补丁。除了 MSCN，去掉补偿模块的 MSCN(CNN)，输入光谱向量用 ICA 或 PCA 处理到 50 个波段的 (ICA-MSHCN, PCA-MSHCN)，以及采用原始光谱的 MSHCN(ORI-MASHCN) 也被用做对比实验。

表 3.2: 不同算法在 UH 数据集上的精度

Indices	Sample per class	Method							
		CNN	FEFCN-ELM	SSRN	DPRN	MSCN	ICA-MSHCN	PCA-MSHCN	ORI-MSHCN
1	100	99.02	97.48	97.68	97.04	98.64	96.26	96.17	99.04
2	100	98.14	98.61	99.72	99.36	99.79	98.47	99.68	94.94
3	100	99.35	100.00	100.00	98.31	99.82	99.68	99.95	99.92
4	100	94.04	98.43	99.84	96.44	96.18	97.33	97.05	96.78
5	100	98.63	100.00	99.62	99.72	100.00	99.96	99.96	99.98
6	100	99.73	100.00	98.77	96.09	99.69	99.60	99.11	99.82
7	100	89.29	94.43	98.10	93.12	95.80	94.87	96.70	93.99
8	100	91.72	80.42	99.25	94.07	92.96	91.04	93.22	92.47
9	100	81.79	94.10	98.06	94.37	95.44	96.93	96.93	96.44
10	100	84.28	99.73	97.38	95.94	93.55	94.88	95.52	98.32
11	100	78.54	99.65	98.86	96.08	97.96	96.18	97.65	93.41
12	100	83.65	95.85	98.40	97.25	95.16	98.83	98.39	98.62
13	100	95.80	95.12	94.14	92.18	96.37	97.07	97.89	94.01
14	100	99.27	100.00	100.00	99.89	100.00	100.00	100.00	100.00
15	100	99.39	100.00	98.28	99.89	100.00	100.00	99.98	100.00
OA	—	91.29	96.36	<b>98.59</b>	96.55	96.97	96.91	<b>97.49</b>	96.78
	—	$\pm 3.29$	$\pm 1.17$	$\pm 0.46$	$\pm 1.73$	$\pm 1.39$	$\pm 1.31$	$\pm 0.88$	$\pm 1.23$
AA	—	92.84	95.39	<b>98.54</b>	96.65	97.42	97.41	<b>97.88</b>	97.18
	—	$\pm 2.61$	$\pm 1.01$	$\pm 0.59$	$\pm 1.75$	$\pm 1.06$	$\pm 1.05$	$\pm 0.72$	$\pm 1.09$
Kappa	—	90.56	96.06	<b>98.47</b>	96.27	96.72	96.67	<b>97.28</b>	96.51
	—	$\pm 3.57$	$\pm 1.43$	$\pm 0.50$	$\pm 1.87$	$\pm 1.50$	$\pm 1.42$	$\pm 0.96$	$\pm 1.33$

实验的相关信息，包括训练样本的设置、每类的精度和评价指标诸如不同算法在 10 次重复实验的总体精度 (Overall Accuracy, OA)、平均精度 (Average Accuracy, AA)

表 3.3: 不同算法在 PU 数据集上的精度

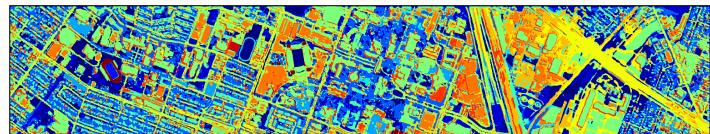
Indices	Sample per class	Method						
		CNN	FEFCN-ELM	SSRN	DPRN	MSCN	ICA-MSHCN	PCA-MSHCN
1	100	88.96	90.45	99.38	92.22	91.09	97.01	97.57
2	100	69.85	86.83	99.73	99.68	97.19	98.33	99.61
3	100	99.30	90.50	93.27	93.27	90.40	97.05	97.50
4	100	98.31	88.80	92.69	99.49	98.68	99.09	97.44
5	100	100.00	90.20	99.99	100.00	100.00	100.00	100.00
6	100	99.92	97.65	95.02	99.33	99.80	98.64	98.32
7	100	93.90	98.54	96.68	91.16	99.11	97.89	98.13
8	100	82.81	88.58	96.01	93.86	98.58	94.39	95.11
9	100	100.00	94.33	99.98	99.87	100.00	100.00	100.00
OA	—	89.09	91.50	<b>97.66</b>	97.40	97.17	97.57	97.49
	—	± 9.05	± 1.31	<b>± 1.34</b>	± 5.16	± 0.69	± 0.43	± 0.71
AA	—	92.66	91.92	96.97	96.54	97.50	<b>97.70</b>	97.69
	—	± 3.52	± 0.57	± 1.70	± 6.12	± 0.89	<b>± 0.57</b>	± 0.52
Kappa	—	86.17	88.84	<b>96.90</b>	96.60	96.24	96.76	96.66
	—	± 10.76	± 1.64	<b>± 1.74</b>	± 6.73	± 0.91	± 0.58	± 0.94

和 Kappa 系数的平均数和标准差在表3.2和表3.3中展示，其中第一名和第二名分别用粗体和蓝色字体显示。所有的算法均使用相同数量样本进行训练。在这两种数据集中，我们从每类随机挑选 100 个样本来组成训练集，剩下的数据用做测试。很显然，我们提出的方法和其他优秀方法相比取得了有竞争力的效果。尤其是在 PU 数据上，这一结果验证了提出的补偿模块和空谱融合框架的效果。不过，不管是采用原始光谱还是经过 PCA 或者 ICA 降维处理后的光谱，精度变化不大，这说明，PCA 和 ICA 能够有效捕捉光谱向量中的主要信息，这意味着可以在减小计算复杂度的同时依然可以保证分类质量。而在 UH 数据集上，SSRN 精度更高，这是因为相比于 MSHCN 中直接级联光谱向量的操作，SSRN 中的 3-D 卷积能够进一步挖掘出光谱向量中通道间蕴含的上下文信息。本文将在第五章对这种信息的提取贡献出两种方案，不同于简单的应用 3-D 卷积获得的单尺度特征，这两种方案均可进一步提取出具有多尺度效应的光谱上下文信息。

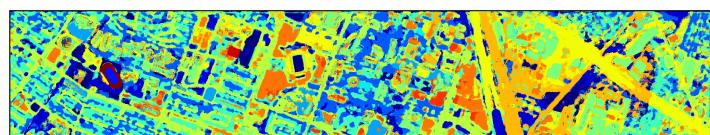
图3.3与3.4展示了原始图像与使用上述方法得到的分类结果。和其他使用小补丁的方法比较，采用大补丁的 FEFCN-ELM 方法产生了过平滑现象，对象的边界有所丢失，这种情况在 PU 数据集上尤为明显。通过 DPRN 方法获得的图3.4视觉效果较差。此外，如图3.4(d) 所示，图像中部连续的房顶被严重错分到其他类别，而且边缘也没能保持。受益于信息补偿模块得到的多尺度特征，我们提出的 MSCN 方法平滑了分类图，由于光谱信息的引入，MSHCN 进一步提高了分类质量，因为它错分更少 (相比于



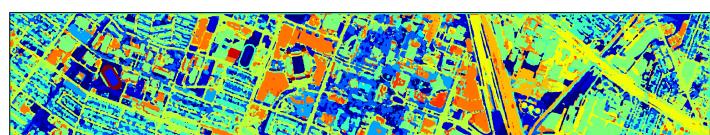
(a)



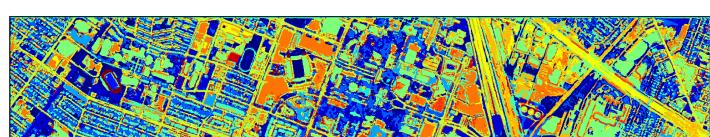
(b)



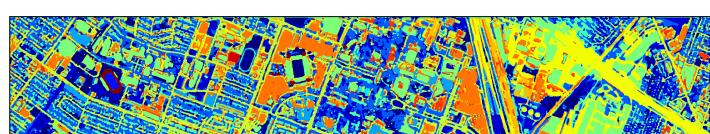
(c)



(d)



(e)



(f)

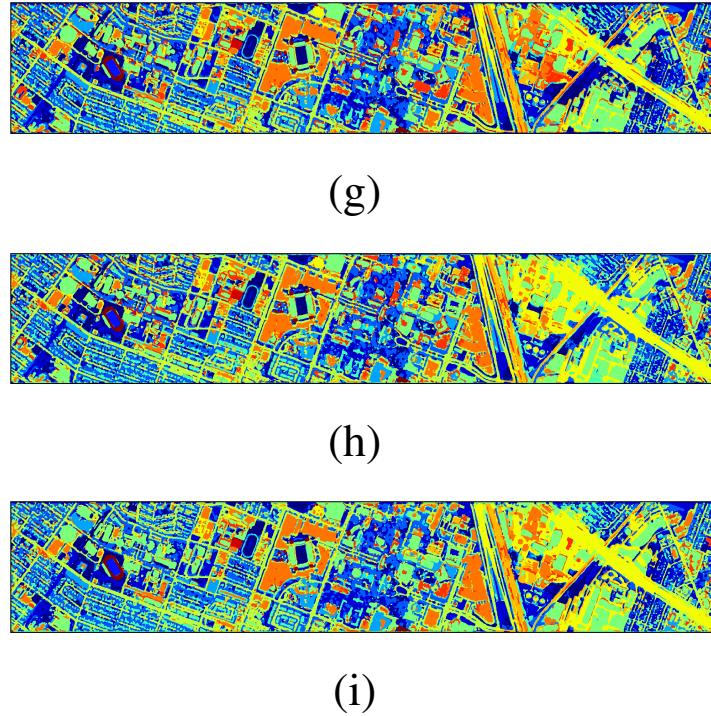


图 3.3: 不同方法在 UH 数据集上的分类结果: (a) 原始图像 (b) CNN (c) FEFCN-ELM (d) SSRN (e) DPRN (f) MSCN (g) ICA-MSHCN (h) PCA-MSHCN (i) ORI-MSHCN.

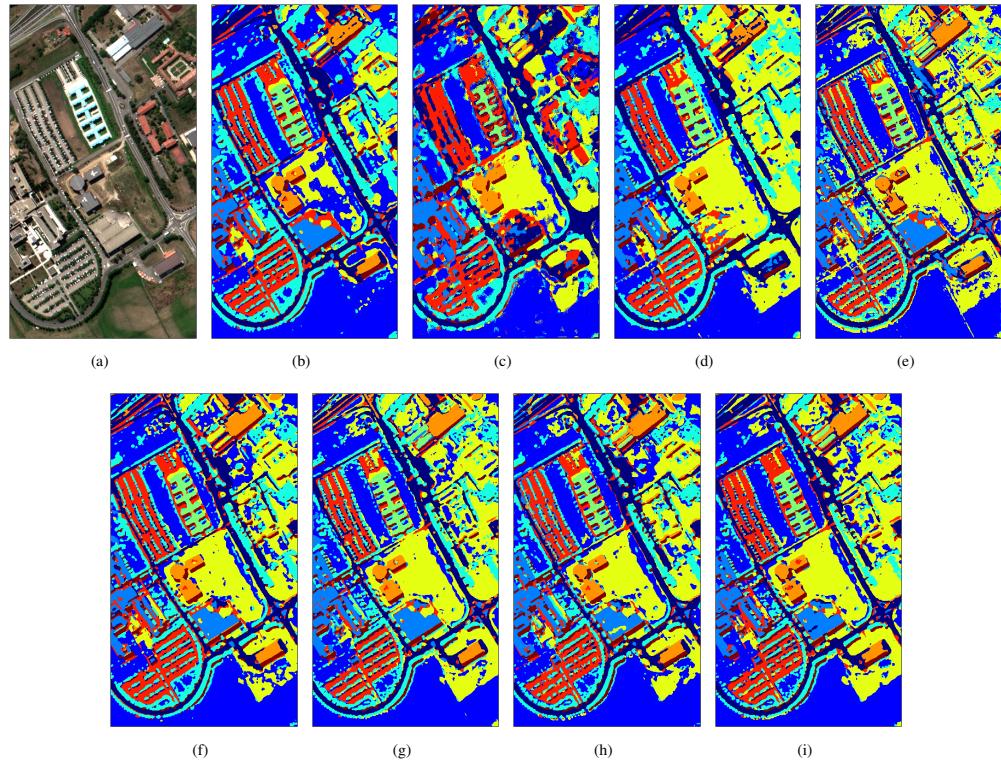


图 3.4: 不同方法在 PU 数据集上的分类结果: (a) 原始图像 (b) CNN (c) FEFCN-ELM (d) SSRN (e) DPRN (f) MSCN (g) ICA-MSHCN (h) PCA-MSHCN (i) ORI-MSHCN.

图3.4(f), PU 图像上部的线性地物在图3.4(g) 被感知), 而且 MSHCN 也保留了边缘。需要说明的是, 本次实验直接采用 matplotlib 中的 jet 颜色映射进行成图, 各类别对应颜色按照默认的映射顺序而定。在后续方法的实验中, 我们将为地物的每一类指定一种特定的颜色。

表 3.4: 不同方法在 UH 数据集上的时间消耗

Indices	Method			
	SSRN	DPRN	MSCN	PCA-MSHCN
Training time(s)	968.76	253.62	123.29	126.66
Testing time(s)	6.94	1.74	1.13	1.07

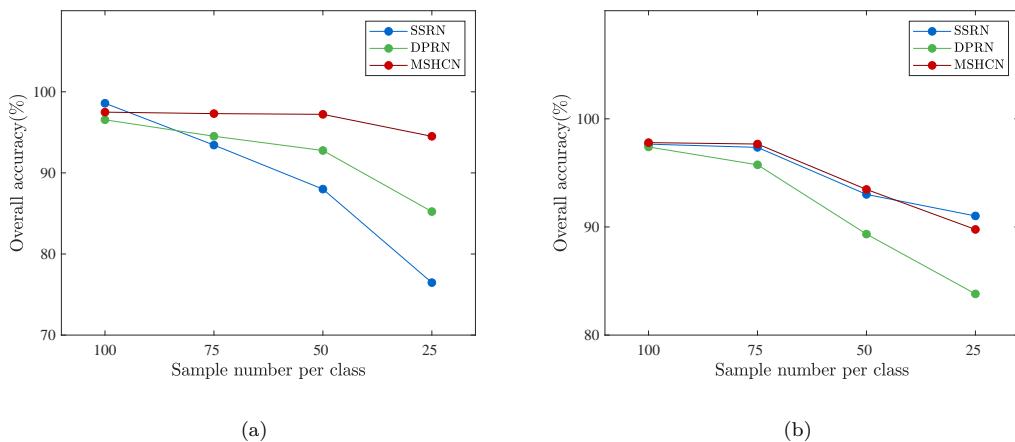


图 3.5: 不同方法不同训练样本个数下的总体精度. (a) UH dataset. (b) PU dataset.

另外, 不同样本下各种方法的总体精度 (UH 数据集采用 PCA-MSHCN, PU 数据集上采用 ORI-MSHCN) 在图3.5中呈现。可以看到与其他两种方法相比, MSHCN 具有竞争力, 且在小样本情况下更为稳定。

最后, 为了更加综合全面进行评估, 我们还记录了表3.2中几个有竞争力的方法在 UH 数据集上 10 次实验的平均训练和测试时间。其结果如表3.4。与使用了 3-D 卷积的 SSRN 相比, 虽然我们提出的仅仅依靠 2-D 卷积的方法在 UH 数据集上的精度略有不足, 但是我们的方法极大的缩小了训练和预测的时间, 这在实际应用中具有重要意义, 尤其是在高光谱图像的实时处理方面。

### 3.4 章末总结

本章主要介绍了一种基于 CNN 的高光谱图像多尺度特征提取模块，其旨在补偿前向信息传播时细节信息的丢失同时保留了语义信息。多尺度来源于特征上的不同感受野，以及卷积操作更加多样化了提取的特征。在此模块基础上，一个端到端的空间信息提取网络被提出用来进行高光谱分类。除了空间信息，光谱信息也被考虑在内，然后一个空谱联合分类框架被开发从而进一步提高高光谱图像分类的精度。实验结果表明，受益于信息补偿模块和空谱特征的融合，提出的方法和其他先进方法比仍有良好的表现。下一章将在本章提出的多尺度模块的基础上进一步探索，并介绍一种初步改进的能够获取空间上下文信息的方法。

## 第 4 章 混合信息感知网络

凭借局部感知和权重共享的良好特性, CNN 已经被高光谱分类领域广泛采用来获取更加抽象的空间和光谱特征。然而, CNN 不能很好的表示相邻像素或波段间的连续性或序列信息。RNN 的序列建模能力已经在 NLP 领域被证实, 因此最近有许多基于 RNN 进行高光谱图像分类的工作被提出。在 [36, 38–40], RNN、LSTM、GRU 和级联 RNN 被分别用来提取波段间的上下文信息。在空间特征提取方面, [47] 将 RNN 应用于经过设计的局部空间序列特征提取。然而, 蕴含于地物间十分重要且有判别力进而可以作为线索用来区分不同类别的空间上下文信息 (Spatial Contextual Information, SCI) 毫无例外的被以上方法忽略掉了。虽然 [47] 使用 RNN 在局部空间序列特征的帮助下提取空间信息, 这可以被视为获取 SCI 的雏形。然而, SCI 中十分关键的相对位置信息 (Relative Position Information, RPI) 和方位信息 (Directional Information, DI) 依旧损失掉了。

为了解决上述问题, 本章介绍了一种新颖的基于 RNN 的获取 SCI, 尤其是 RPI 和 DI 的模块。紧接着我们开发出了基于 CNN 的混合信息感知网络 (Hybrid Perception Network, HPN) 和空谱混合信息感知网络 (Spectral-Spatial Hybrid Perception Network, SS-HPN)。混合不仅意味着 CNN 和 RNN 的混合, 更是综合利用多种信息的含义。需要说明的是, 与 [38, 40] 将 RNN 建立在 CNN 获取的特征的基础上不同, 我们的方法则将 RNN 嵌入 CNN 中作为一个单独的模块, 下面进行详述。

### 4.1 准备工作

图4.1展示了我们提出方法的整体流程: 首先, 对原始影像直接采用 PCA 进行降维以克服维数灾难带来的影响。接着是逐通道的归一化。我们从经过上述处理后的数据获取尺寸为  $w \times w \times p$  的补丁作为标签是中心像素类别的训练样本,  $p$  是降维后的波段数。不过如果直接按照从原始图像大小的数据上取补丁对边界像素可能不太公平。因此, 有必要进行边界扩展。然而, 传统的 0 值补全可能严重的破坏从边界像素取领

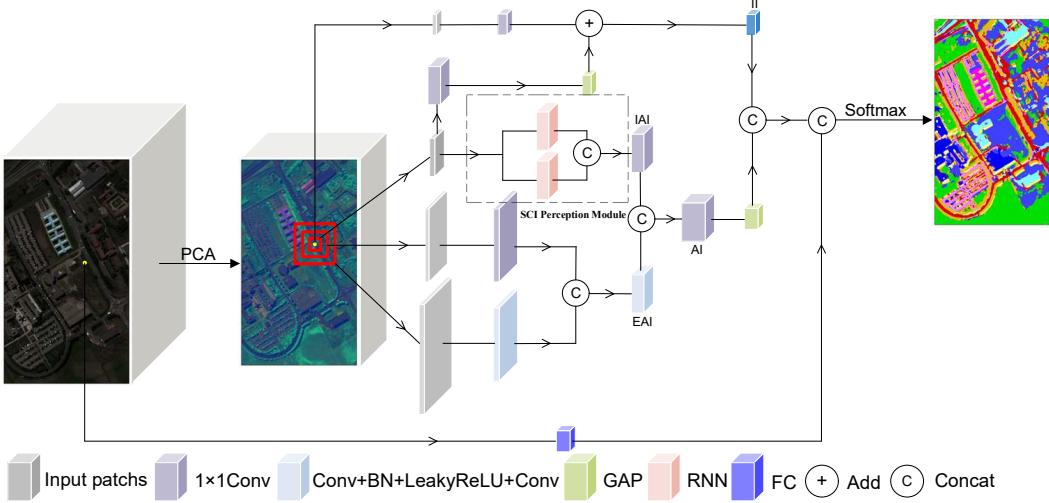


图 4.1: 网络整体流程

域得到的补丁的数据平衡性，这一现象随着  $w$  的增大而愈发严重因为这些补丁中约有一半的像素为 0，这对分类不仅毫无帮助，还可能带来负面影响。因此，在获取补丁前我们在边界进行对称扩展以模拟边界像素周围的真实环境，然后这些补丁就可以被送入网络了。

## 4.2 SCI 感知模块

图4.2展示了我们提出的 SCI 感知模块在从左到右方向上的运行过程。输入该模块数据的尺寸是  $3 \times 3$  因为我们主要考虑位于目标像素附近的局部 SCI。这种情况可以被视作提取 SCI 最不复杂且最直接的方式，需要的计算资源也最少。双通道的结构表示我们从微观 (像素级) 和宏观 (区域级) 两种水平提取 SCI，相当于从两种尺度研究局部环境。

1) 像素级 SCI 感知：上支路表示了微观 SCI 在从左到右方向的提取。事实上，我们总共考虑了八种方向，这对应于中心像素的八个相邻像素。在每个方向的路径中，中心像素都在中间的时间步。例如，在从左到右的方向，左侧的输入对应于目标中心像素左侧最近的像素，右侧的输入则是目标中心像元右侧最近的像素。相应的，八个 RNN 被采用并且最后一个时间步输出的将作为当前方向提取到的 SCI。经过上述操作，原始  $3 \times 3$  输入的八个边界像元被替换成八个方向分别提取到的 SCI。因为中心像元和八个方向均有关系，因此将中心像元替换成边界改变的八个像素的任

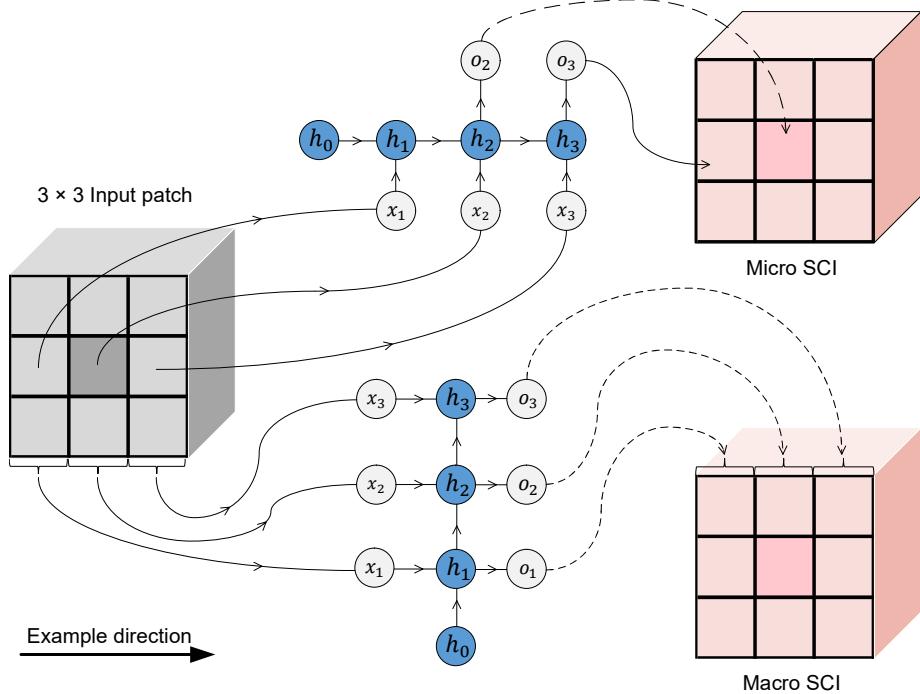


图 4.2: 空间上下文信息提取模块 (示例方向: 从左到右)

意一个都是不合适的。因此，我们使用八个 RNN 的中间时间步输出的平均值来替换原始的中心像元因为它可以认为是中心像素与周围八个方向的八个邻接像素的交集，这样我们就得到了上方支路的结果。假设原始  $3 \times 3$  输入为  $\mathbf{x}_0 \in \mathbb{R}^{3 \times 3 \times d_0}$ ，上方支路的输出为  $\mathbf{x}_1 \in \mathbb{R}^{3 \times 3 \times d_1}$ ， $d_1$  由隐层神经元数目确定。

- 2) 区域级 SCI 感知: 宏观 SCI 的提取展示了下支路, 方向同样是从左到右。现在, 我们使用四个而非八个 RNN 来捕捉竖直和水平方向上的上下文信息, “四”表示着水平情况下的从左到右, 从右到左和竖直情况下的从上到下与从下到上总共四个方向。与微观情况类似, 提取宏观 SCI 时的 RNN 也是三个时间步。依然以从左到右的方向为例, 第一个时间步是左边的输入, “左”表示中间列左边的最近的列。换句话说, 第一列需要展平然后才可以作为 RNN 的第一个输入值, 三列对应三个时间步。对于所有三个时间步长的输出, 应将处于隐藏状态的神经元数量设置为三的倍数, 这些输出构成当前方向的宏观 SCI 感知的结果。在从左到右的方向, 再次假设原始的  $3 \times 3$  输入  $\mathbf{x}_0 \in \mathbb{R}^{3 \times 3 \times d_0}$ ,  $\mathbf{x}_0 = [\mathbf{x}_0^{:1}, \mathbf{x}_0^{:2}, \mathbf{x}_0^{:3}]$ ,  $\mathbf{x}_0^{:1}$  是原始补丁的第一列。展平后的  $\mathbf{x}_0^{:1}$  是第一个输入值, 此时的其输出值可以被变换形状为  $\mathbf{x}_{2 \rightarrow}^{:1} \in \mathbb{R}^{3 \times 1 \times d_2}$ , 从而代替了原始补丁的第一列, 其余时间步执行相同的操作, “ $\rightarrow$ ”表示从左到右的方向。经过上述过程, RNN 的输出值为  $\mathbf{x}_{2 \rightarrow} \in \mathbb{R}^{3 \times 3 \times d_2}$ , 四个 RNN 中隐层神经元个数均为

$3 \cdot d_2$ 。最后，可以得到分别代表四个方向的四个 RNN 的输出  $[\mathbf{x}_{2\rightarrow}, \mathbf{x}_{2\leftarrow}, \mathbf{x}_{2\uparrow}, \mathbf{x}_{2\downarrow}]$ 。宏观 SCI 感知的结果是这四个输出值的平均值。

经过上述过程，可以看到这两个路径分别具有在  $3 \times 3$  补丁上捕捉像素或区域级上下文信息的能力。所以这个模块可以感知到宏观和微观的 SCI。此外，这个模型可以捕捉到 DI 因为 RNN 是沿特定的方向设计的。输入值被视为序列数据，当考虑方向时这些数据包含 RPI，因为这些位于遥感影像上的像素表示了实际地物表面。可以相信这个 SCI 感知模块可以进一步提升分类表现凭借 RPI 和 DI 的感知能力。在我们的实验中， $d_1$  和  $d_2$  均被设为 256， $d_0 = p$ ，这些 RNN 的激活函数均被设为 ReLU 以提高计算效率。

### 4.3 混合信息感知网络

混合信息感知网络的构建是建立在 SCI 感知模块的基础上，这个网络接收大小为  $w \times w \times p$  的补丁作为输入。为了达到多尺度特征提取的效果，中等尺寸、 $3 \times 3$  和  $1 \times 1$  尺寸的补丁也被使用，其中中尺寸是  $(w + 1)/2$ 。这三种尺寸的补丁均可以从  $w \times w \times p$  的输入数据上获得，在我们的实验中， $w = 21$ ， $p = 10$ 。

首先，大尺度信息通过两个卷积层提取。为了避免过拟合且实现非线性映射，BN [57] 和 LeakyReLU  $g_l(x) = \max(0, x) + 0.01 \cdot \min(0, x)$  用在两个卷积层之间。使用 LeakyReLU 的原因是为了保持负值，因为数据被归一化到了  $[-1, 1]$ 。两个卷积层的卷积核大小分别为  $7 \times 7 \times 48$  和  $5 \times 5 \times 64$ 。接着获得的特征图与经过  $1 \times 1$  卷积后的通道数相同的中尺寸特征图连接。 $1 \times 1$  卷积具有跨通道整合信息的能力 [56] 同时完成通道数匹配。然后另外两个卷积层被分别用来编码刚才得到的具有更大感受野的特征图和中尺度信息，这两种信息分别对应于先前的特征图和原始的中尺寸补丁。这两个卷积核大小被分别设为  $5 \times 5 \times 192$  和  $5 \times 5 \times 192$ ，BN 和 LeakyReLU 也被使用。现在特征图大小为  $3 \times 3$ ，它可以被视为外部辅助信息 (External Auxiliary Information, EAI) 因为它包含了两个相对较大尺度的表示目标像素周围环境的不变和内在特征。EAI 更适合作为补充信息而不是直接输入分类器中因为当理解目标地物时它更关注语义角度而非细节信息。此外，通过连接宏观和微观角度并通过一个  $1 \times 1 \times 256$  卷积层整合信息提取得到的 SCI 被考虑作为内部辅助信息 (Internal Auxiliary Information, EAI) 因为 RNN 更关注像素之间的关系而不是同时感知局部区域的所有像素。因此，我们

连接了 EAI 和 SCI 从而得到了辅助信息 (Auxiliary Information, AI)。待分类像素的内在信息 (Intrinsic Information, II) 由原始的  $3 \times 3$  和  $1 \times 1$  补丁产生因为它们在小尺度下获得，更接近于原始情形。 $1 \times 1$  卷积只用作尺寸匹配。为了维持原始信息，全局平均池化 (Global Average Pooling, GAP) 用在  $3 \times 3$  补丁上，这减少了计算复杂度。II 通过将处理后的原始  $3 \times 3$  和  $1 \times 1$  补丁相加获得。然而，AI 的尺寸是  $3 \times 3$ ，这与 II 不匹配。因此 GAP 也需要用在 AI 上并尽可能的保持有效信息。上述过程中所有的卷积核均没有边界填充且步长均为 1。

HPN 最终提取得到的特征是经过 GAP 后的 AI 与 II 的连接，连接后的特征在经过一个全连接层和 softmax 层后可以用来分类。HPN 使用交叉熵损失函数，定义如下

$$L = -\frac{1}{N} \sum_{i=1}^N [z_i \log(y_i) + (1 - z_i) \log(1 - y_i)] \quad (4.1)$$

上式中， $N$  是训练样本数， $z_i$  和  $y_i$  分别表示真实和预测标签。这个网络通过小批量随机梯度下降测量优化，RNN 部分中的参数通过基于时间的反向传播 (Back-propagation through time, BPTT) [62] 算法更新。

进一步，光谱信息也应该被考虑，光谱向量源自于归一化后的但没有做降维处理的原始图像。这些光谱向量仅仅通过一层全连接层放缩然后获得与连接 AI 与 II 后的特征相同长度的特征。接着空谱联合分类也可以通过一层全连接层和 softmax 层完成。这样就实现了一个新的框架：SS-HPN。这个网络也使用交叉熵损失函数和同样的优化方法。

另外，Dropout 也被用克服过拟合现象。它主要用在一些具体信息构建完成后，包括在 HPN 中形成 EAI, SCI 和 II，以及光谱信息和空间信息的连接。这种配置可以提取更加均衡的不同种类的特征同时避免了过拟合问题。上述方法均是使用 Pytorch 框架在一块 NVIDIA RTX 2080Ti GPU 上实现的。

## 4.4 实验及分析

本章节采用 Indian Pines (IP) 数据集和 PU 数据集进行算法验证。移除水吸收波段后，前者大小为  $145 \times 145 \times 200$ ，有 16 类，后者大小为  $610 \times 340 \times 103$ ，有 9 类。

我们将提出的方法与一些基于 CNN 的先进方法比较，这些方法有 3DCNN [31]

,SSRN [32], 因为后者也使用了残差连接。此外, 为了评估我们提出的 SCI 感知模块, LSS-RNN 结构也被考虑在内。为了更公平的比较, 我们将输入改为了经过 PCA 降维且归一化后的数据而非 [47] 提及的底层特征, 只是将 HPN 中的 SCI 感知模块替换成 LSS-RNN。另外, 我们还测试了 RNN 单元的影响, 除了最基本的 RNN 单元 (HPN-RNN, SS-HPN-RNN), LSTM 和 GRU 也被包含到我们的实验中 (HPN-LSTM, HPN-GRU, SS-HPN-LSTM, SS-HPN-GRU)。除了 IP 数据集中的 alfalfa, grass-pasture-mowed, oats 和 stone-steel-towers 类别中的样本被分别设定为 33, 20, 14 和 75 以外。这两个数据集中的其余类别均随机选择 100 个样本。包括 OA, AA, 和 kappa 系数等一些实验细节在表4.1和4.2中展示并且当前指标下的第一、二、三名被标记为粗体、红色和黑色, 所有的实验均随机测试 10 次。 从表中可以看到我们提出的模块在多尺度

表 4.1: 不同算法在 IP 数据集上的分类精度

Class	3DCNN	SSRN	CNN	LSS -RNN	HPN -RNN	HPN -LSTM	HPN -GRU	SS-HPN -RNN	SS-HPN -LSTM	SS-HPN -GRU
OA	95.80	<b>97.26</b>	96.20	96.68	96.71	96.83	96.75	96.81	<b>97.30</b>	<b>97.03</b>
	±0.70	<b>±0.63</b>	±0.79	±0.64	±0.59	±0.52	±0.42	±0.41	<b>±0.63</b>	<b>±0.47</b>
AA	97.98	96.20	98.47	98.64	98.70	98.74	98.74	<b>98.76</b>	<b>98.86</b>	<b>98.79</b>
	±0.26	±1.10	±0.39	±0.18	±0.28	±0.17	±0.13	<b>±0.14</b>	<b>±0.29</b>	<b>±0.17</b>
Kappa	95.15	<b>96.83</b>	95.61	96.15	96.19	96.33	96.24	96.31	<b>96.87</b>	<b>96.57</b>
	±0.80	<b>±0.72</b>	±0.91	±0.74	±0.69	±0.60	±0.48	±0.47	<b>±0.73</b>	<b>±0.54</b>
1	100.00	100.00	98.46	99.23	100.00	100.00	100.00	100.00	100.00	100.00
2	93.12	96.99	95.12	93.70	95.78	95.17	94.06	94.83	95.50	95.29
3	96.05	97.02	98.69	99.10	98.71	97.74	98.99	98.67	98.45	98.89
4	99.93	95.73	100.00	99.93	100.00	100.00	100.00	100.00	100.00	100.00
5	97.49	99.33	99.32	99.09	99.40	99.43	98.85	99.43	98.77	99.43
6	98.81	99.00	99.65	99.68	99.90	99.65	99.84	99.81	99.86	99.63
7	100.00	95.71	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
8	99.87	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	95.47	91.82	96.62	96.49	96.09	96.72	96.79	96.77	97.00	96.86
11	93.27	98.30	91.08	93.63	93.15	93.61	93.43	93.66	94.97	94.13
12	97.69	96.69	98.07	98.60	98.86	98.40	98.66	98.42	98.19	98.38
13	100.00	98.39	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
14	97.81	99.46	99.18	98.96	97.98	99.13	99.26	98.58	99.09	98.73
15	99.75	95.06	99.90	99.79	99.90	99.93	100.00	100.00	100.00	99.90
16	98.33	75.68	99.44	100.00	99.44	100.00	100.00	100.00	100.00	99.44

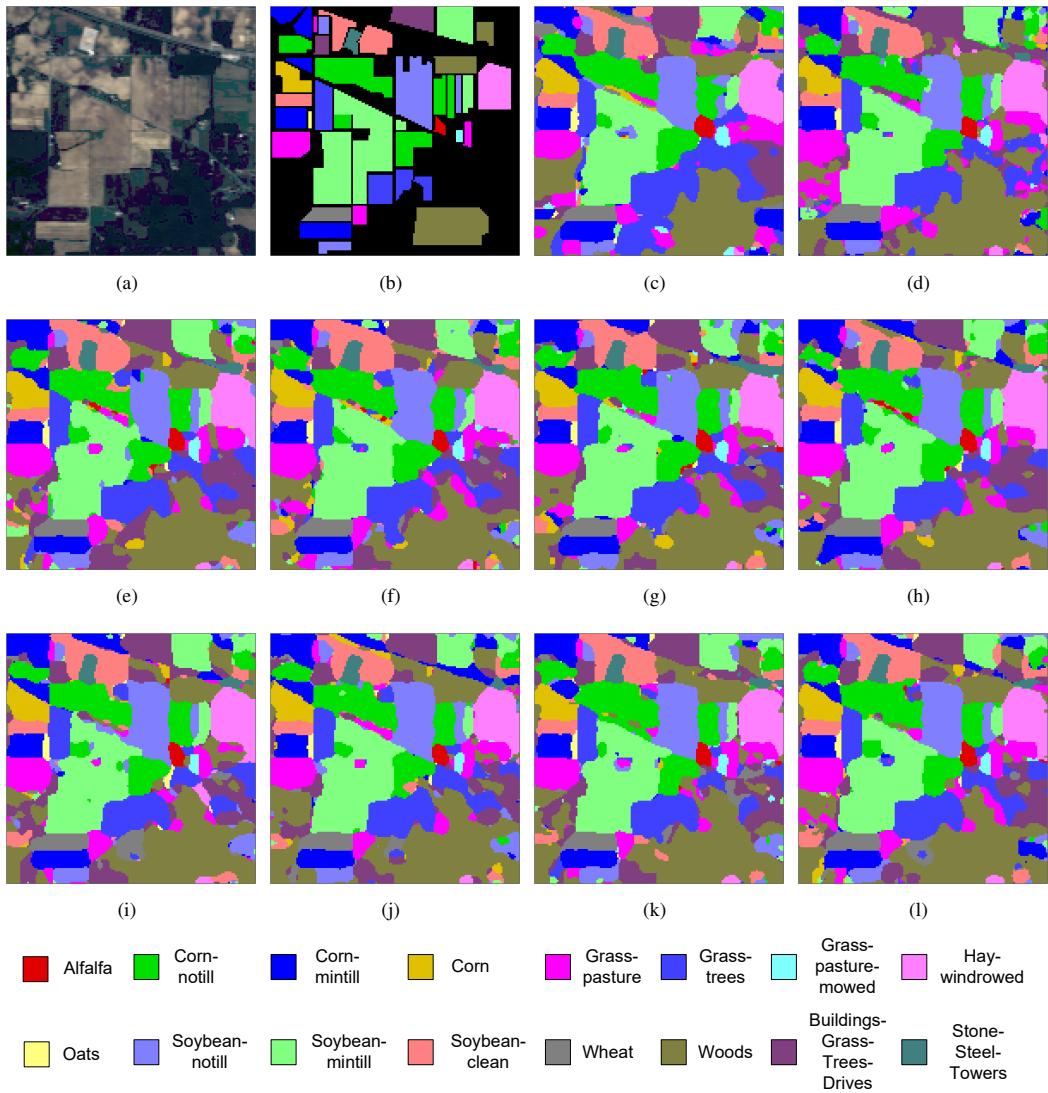


图 4.3: 不同方法在 IP 数据集上的分类结果图: (a) 原始图像. (b) 地物标签. (c) 3DCNN. (d)SSRN. (e) CNN. (f) LSS-RNN. (g) HPN-RNN. (h) HPN-LSTM. (i) HPN-GRU. (j) SS-HPN-RNN. (k) SS-HPN-LSTM. (l) SS-HPN-GRU.

CNN 的基础是进一步提高了精度，并且这一模块较 LSS-RNN 更为优越，因为它考虑了 SCI，尤其是 RPI 和 DI。另外，与其他先进方法比，我们的方法具有竞争性。在三种 RNN 单元中，LSTM 取得了最高的精度。定量实验结果也表明随着光谱信息的引入，分类精度再次有所提高，这显示了特征多元化的重要性。

图4.3和4.4展示了不同方法的分类结果。与其他方法相比，我们的方法均有更好的视觉效果。可以看到提出的模块减少了误分类现象，尤其是在 PU 数据集上。当引入光谱信息后，土地覆盖的边界和一些小的地物被很好的保持下来。

表 4.2: 不同算法在 PU 数据集上的分类精度

Class	3DCNN	SSRN	CNN	LSS -RNN	HPN -RNN	HPN -LSTM	HPN -GRU	SS-HPN -RNN	SS-HPN -LSTM	SS-HPN -GRU
OA	87.71	97.66	97.46	97.98	98.08	<b>98.21</b>	97.99	<b>98.09</b>	<b>98.22</b>	98.00
	$\pm 1.75$	$\pm 1.34$	$\pm 0.79$	$\pm 0.73$	$\pm 0.58$	<b><math>\pm 0.49</math></b>	$\pm 0.67$	<b><math>\pm 0.58</math></b>	<b><math>\pm 0.66</math></b>	$\pm 0.72$
AA	90.46	96.97	97.76	98.23	<b>98.28</b>	98.19	<b>98.28</b>	<b>98.29</b>	<b>98.27</b>	97.88
	$\pm 0.89$	$\pm 1.70$	$\pm 0.62$	$\pm 0.83$	<b><math>\pm 0.25</math></b>	$\pm 0.52$	<b><math>\pm 0.34</math></b>	<b><math>\pm 0.35</math></b>	<b><math>\pm 0.65</math></b>	$\pm 0.83$
Kappa	83.81	96.90	96.63	97.31	97.44	<b>97.61</b>	97.32	<b>97.45</b>	<b>97.62</b>	97.33
	$\pm 2.03$	$\pm 1.74$	$\pm 1.05$	$\pm 0.97$	$\pm 0.77$	<b><math>\pm 0.66</math></b>	$\pm 0.88$	<b><math>\pm 0.77</math></b>	<b><math>\pm 0.87</math></b>	$\pm 0.96$
1	86.80	99.38	94.63	95.36	94.97	96.57	95.12	95.01	94.71	95.19
2	88.65	99.73	98.24	98.63	98.98	99.19	98.69	98.95	99.38	99.35
3	83.22	93.27	95.66	95.71	96.87	96.08	96.62	97.14	96.69	94.81
4	96.61	92.69	98.35	98.97	98.57	98.58	98.60	98.42	98.42	98.44
5	99.98	99.99	100.00	99.99	99.97	99.99	100.00	99.94	100.00	99.99
6	71.53	95.02	98.16	98.70	98.49	98.63	98.63	98.94	98.58	97.91
7	95.55	96.68	98.98	99.85	99.61	98.24	99.69	99.41	99.37	99.30
8	92.12	96.01	95.97	97.07	97.14	96.78	97.21	96.81	97.33	96.09
9	99.73	99.98	99.88	99.80	99.96	99.67	99.95	99.96	99.98	99.85

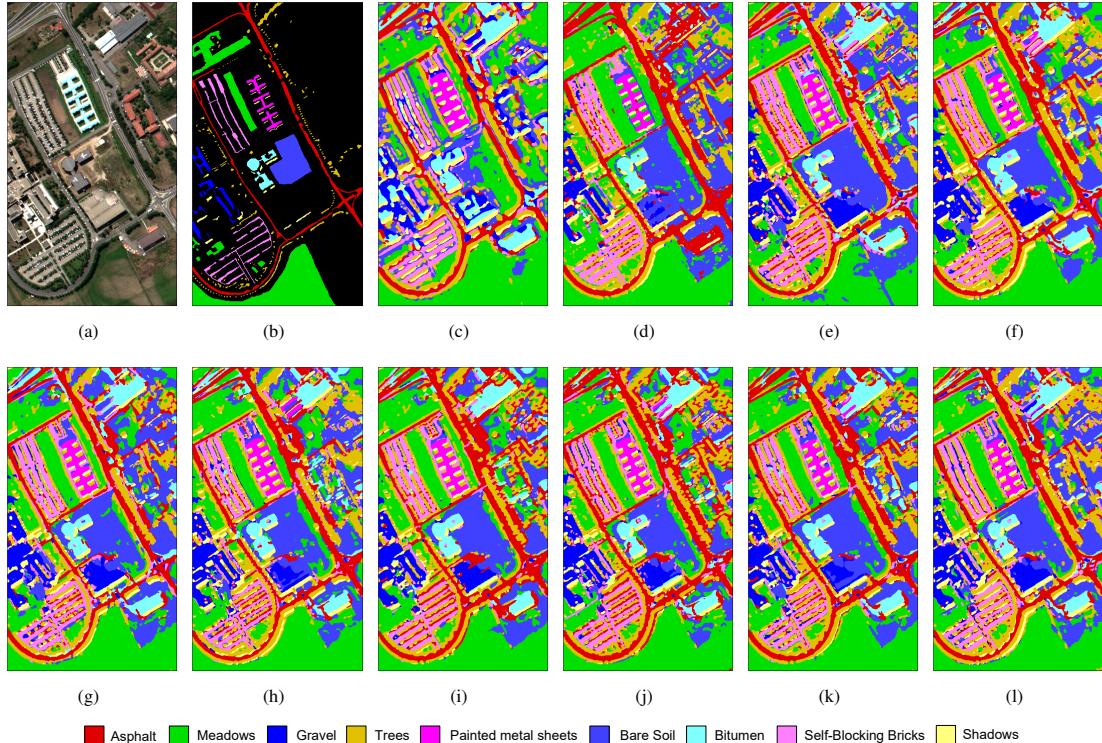


图 4.4: 不同方法在 PU 数据集上的分类结果图: (a) 原始图像. (b) 地物标签. (c) 3DCNN. (d)SSRN. (e) CNN. (f) LSS-RNN. (g) HPN-RNN. (h) HPN-LSTM. (i) HPN-GRU. (j) SS-HPN-RNN. (k) SS-HPN-LSTM. (l) SS-HPN-GRU.

## 4.5 章末总结

本章介绍了一种可以端到端获取多特征用于高光谱空谱联合分类的网络。除了不同尺寸的补丁被用来提取多尺度信息外，一个新颖的基于 RNN 的模块也被设计以从空间层面感受上下文信息，尤其是 RPI 和 DI。接着，光谱信息也被考虑以进行空谱融合。实验结果表明，与其他方法相比我们提出的方法表现良好，分类结果也展示了所提出的 SCI 感知模块的有效性，以及在高光谱分类中信息多元化的意义。下一章将在前两章的基础上，介绍一种能够在光谱和空间方面同时挖掘更为强力的多尺度上下文信息从而能够对高光谱图像进行高精度分类的模型。

## 第 5 章 自适应空谱多尺度网络

前文提到的几乎所有方法除了 [39] 都忽略了特征的尺度问题，因为它们最后提取出用于分类的特征总是单尺度的，这同时意味着这些特征都具有同样的感受野，这限制了分类精度。然而，在 [39] 中，光谱部分也仍然是一个单尺度模型。受 [39] 提出的 MSCNN 的启发，自然能够想到进行多尺度序列学习，从而感知和产生更丰富的信息，这可能对分类产生有利影响。此外，之前的方法通常不能同时实现多尺度特征提取和空间上下文信息感知。例如 3-DCNN 和 MSCNN 通过滑动卷积核操纵 CNN 中的特征图，这并不能获得当前尺度下的空间上下文信息。

为了解决上述问题，在第三、四章工作的基础上，本章介绍了一种的名为“自适应空谱多尺度网络”(Adaptive Spectral-Spatial Multiscale Network, ASSMN) 的方法用于端到端的高光谱图像分类。它包含两个子网络：光谱多尺度网络 (Spectral Multiscale Network, SeMN) 和空间多尺度网络 (Spatial Multiscale Network, SaMN) 分别用于光谱和空间信息的提取。“自适应”意为从这两个部分中提取到的特征被合理的结合起来而且相关参数可以通过端到端的学习自动确定。通过共享一个统一的损失函数，这两部分被纳入了一个统一的框架并通过反向传播进行训练。在光谱信息提取方面，多尺度的光谱变异通过灵活的运用 LSTM 感知得到。受 [49, 50] 的启发，我们在空间信息感知网络部分采用 ConvLSTM 来实现多尺度特征提取和上下文信息的捕捉，本章的主要内容概括如下。

- 1) 提出了一个名为 ASSMN 的自适应端到端分类网络。与现有的空谱融合框架相比，我们的模型不仅完全利用了多尺度特征，还具有自适应学习的能力，从而可以更有效的结合光谱和空间特征。
- 2) 为了获得更有判别力的特征，我们设计了两种多尺度方案 (scheme)，方案 1 在原始光谱向量上操作，另一个方案则具有卷积层特征。基于这两种方案的两个相应的网络也被进一步开发出来。
- 3) 提出了一种新颖的基于 ConvLSTM 的空间信息提取网络，它可以同时完成多尺度的特征提取与上下文信息的捕捉。

4) 提出的方法在三个被普遍在高光谱图像分类社区中公共基准数据集中均取得了令人满意的效果。定量结果和分类结果图均展现了我们提出的方法相比于其他先进方法的竞争力。

## 5.1 光谱多尺度子网络

尽管前人的工作已经将序列模型用于获取光谱信息 [36, 37, 39, 40]，特征的尺度这一问题却被忽略并且现有方法中提取的用于分类特征仅仅只有单一种类的感受野。前文证明了多尺度空间特征有助于分类，因此自然而然可以将类似想法扩展到光谱特征提取领域。为了提取多尺度光谱特征，我们探索了在宽泛或者狭窄的波长范围内的不同反射率之间的关系。

经过归一化的预处理过程后，原始的光谱向量可以被送入 SeMN。探索不同尺度的最简单的方式就是多样化 LSTM 的输入数据。我们提出了两种分别从波长和感受野角度入手的方案来达到多尺度效果。本文提出的 SeMN 分类方法流程由图5.1示意

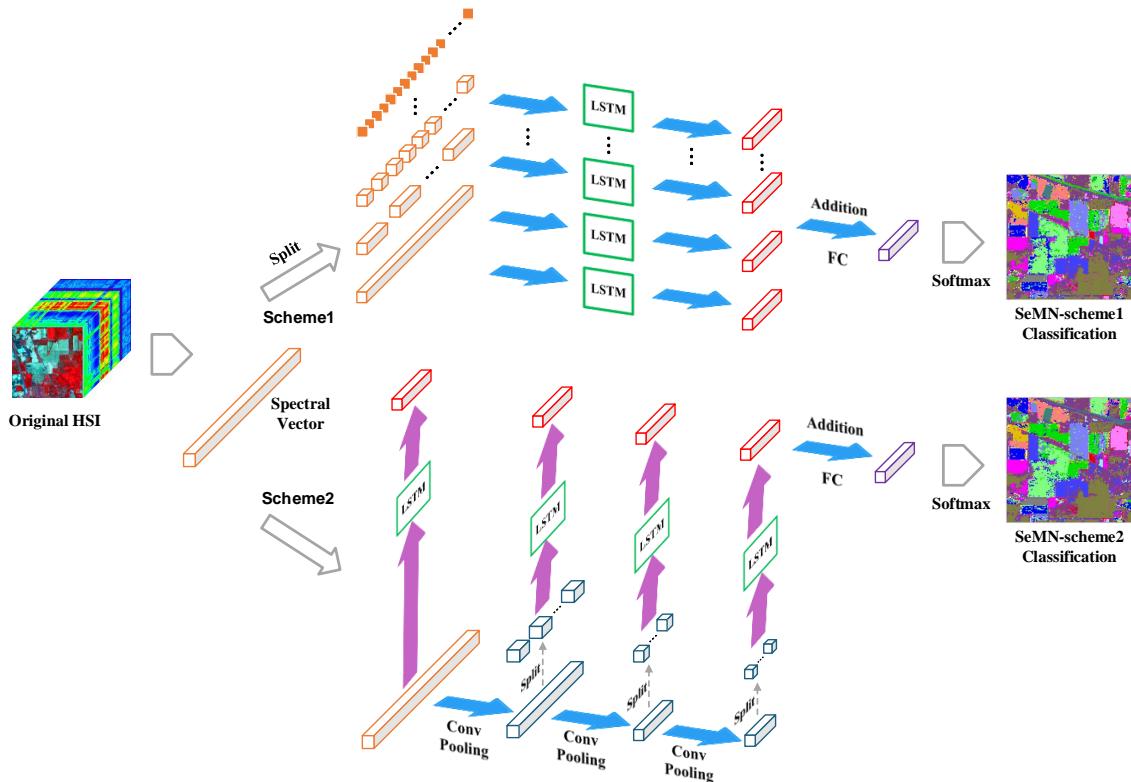


图 5.1: 本文提出的 SeMN，包含两条路径分别代表两种方案。“Split” 表示将光谱向量分为几组作为 LSTM 不同时间步的输入。Conv、Pooling 和 FC 表示卷积层，最大池化层和全连接层。

### 5.1.1 方案一：波长多尺度

在早期基于 LSTM 的光谱分类方法中，当前像素每个波段的反射率值在每个时间步被依次输入 LSTM [36]。然而，这样可能会产生长序依赖 [55]，带来梯度爆炸或者梯度消失的问题，因为高光谱图像中具有数百个波段。一种可能的解决办法是通过对光谱向量分组从而一次性使用更多的波段。[39] 提出了两种新颖的分组策略。然而，这些策略仍然是在单尺度上操作的，这限制了 LSTM 的性能。

一种获取位于不同波长范围的多个光谱段的自然想法重复使用分组策略若干次，这就是图5.1中方案 1 的含义。最简单的方法就是不断的对分光谱向量，然后把这些结果输入不同的 LSTM。最终的光谱特征通过结合所有 LSTM 的输出来得到。例如，一个具有两个时间步的 LSTM 用于第一次对分的光谱向量。再次把这两部分对分，一个具有四部分的光谱将会将被送入另一个 LSTM，并且获得当前尺度下的特征。为了获得全局信息，整个光谱向量也被直接输入到一个 LSTM。最后，所有的 LSTM 输出的特征通过逐点相加合并。

### 5.1.2 方案二：感受野多尺度

由于考虑到交替分组，[39] 进一步提高了所提取特征的判别能力。其本质是减小了相邻时间步的输入在特征空间中的距离，使得这些向量过渡更为自然。受益于这一想法，方案二旨在产生相邻时间步距离更近的序列数据。这一方案可以实现多尺度特征提取并同时保证输入序列的连贯性。如图5.1下方支路所示，通过使用多个卷积层，可以得到若干个特征图。事实上，这些不同层的特征图具有多尺度信息。因为深度网络具有层级学习的特性，深层的感受野大于浅层的。如果卷积核的滑动步长小于卷积核大小，那每组中相邻通道的数据的感受野一定存在交集。获得这些特征图后，与方案一中的分组方法类似，这些特征图被分为若干部分并分别送入不同的 LSTM，不过此时这些特征图只需要分离一次。方案二中需要调整的超参数是在每个特征图上操作的不同 LSTM 的时间步数。换句话说，不同的特征图可能被分为不同的组。这些参数将会在 5.4.3 1) 和 2) 中进行讨论。此外，这个方案也采用了全局信息和加法操作。

最后，SeMN 使用了一个全连接层和 softmax 函数将方案一和方案二中融合后的特征转换为分类图。

## 5.2 空间多尺度子网络

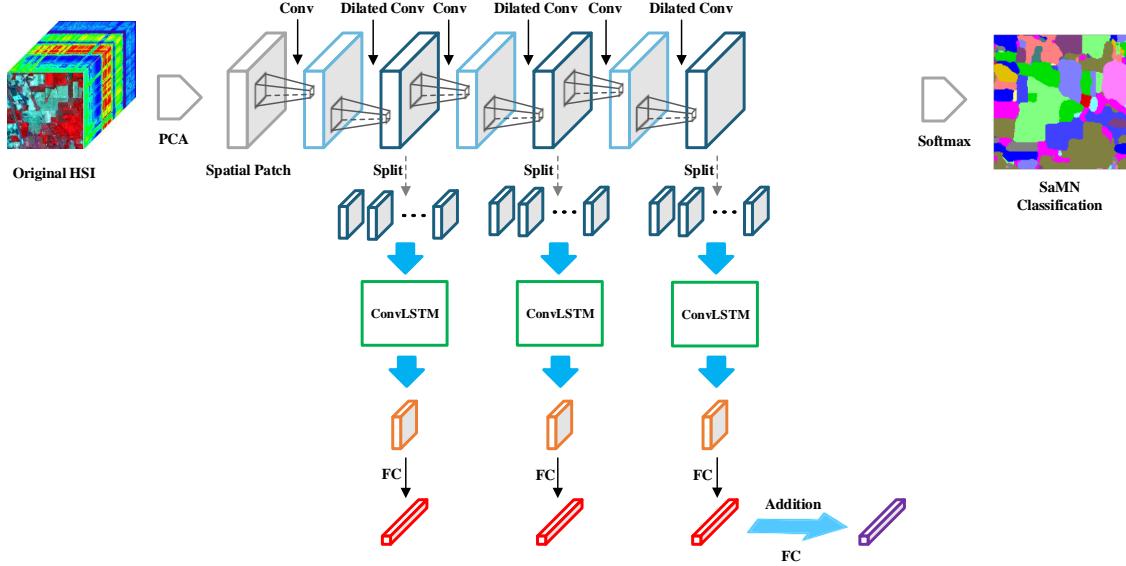


图 5.2: 本文提出的空间分类子网络 SaMN。在 PCA 降维之后，空间补丁被送入这个网络。Conv、Dilated Conv 和 FC 分别是卷积、扩张卷积和全连接层。“Split”表示将特征图分为在不同时间步输入 ConvLSTM 的若干组。

图5.2展示了本文中提出的 SaMN 的工作流程。首先，PCA 用于原始高光谱图像进行降维处理。在归一化操作后，取目标中心像素周围尺寸为  $H \times W \times C$  的空间补丁并输入网络。 $H, W$  和  $C$  是降维后这些补丁的高、宽及通道数。为了方便起见，本文中宽高相等。

[39] 提出的 MSCNN 提供了一种多尺度空间信息提取的示例。然而，这个结构仍然需要改进。具体来说，MSCNN 中的池化层主要用于增大感受野从而获得多尺度特征。这不可避免的缩小了特征图的尺寸并丢失了空间信息。为了达到这一目标，我们将所有的池化层替换了扩张卷积 [44, 63–65] 并后接 ReLU 激活函数。扩张卷积通过在卷积核内嵌入空洞来解决上述问题，从而在增大感受野的同时保持特征图的分辨率。ReLU 激活函数用来实现之前最大值池化操作中的非线性映射。此外，SaMN 也考虑了空间上下文信息，弥补了 MSCNN 不能感知关系型信息的弱点。

如图5.2所示，额外的空间上下文信息提取分支被添加在扩张卷积层后，这与 MSCNN 收集多尺度信息的结构类似。我们这样设计目的是在获得空间上下文信息的同时在主干网络保证层次化的学习。因此提取到的特征同样具有多尺度特性，它们需要在空间维进行分组。对于图5.3中展示的尺寸为  $H \times W \times C$  的空间补丁，假设每个时间步输入 ConvLSTM 的数据尺寸为  $s \times s \times C$ ，这意味着在行和列方向分别有  $H/s$  和  $W/s$

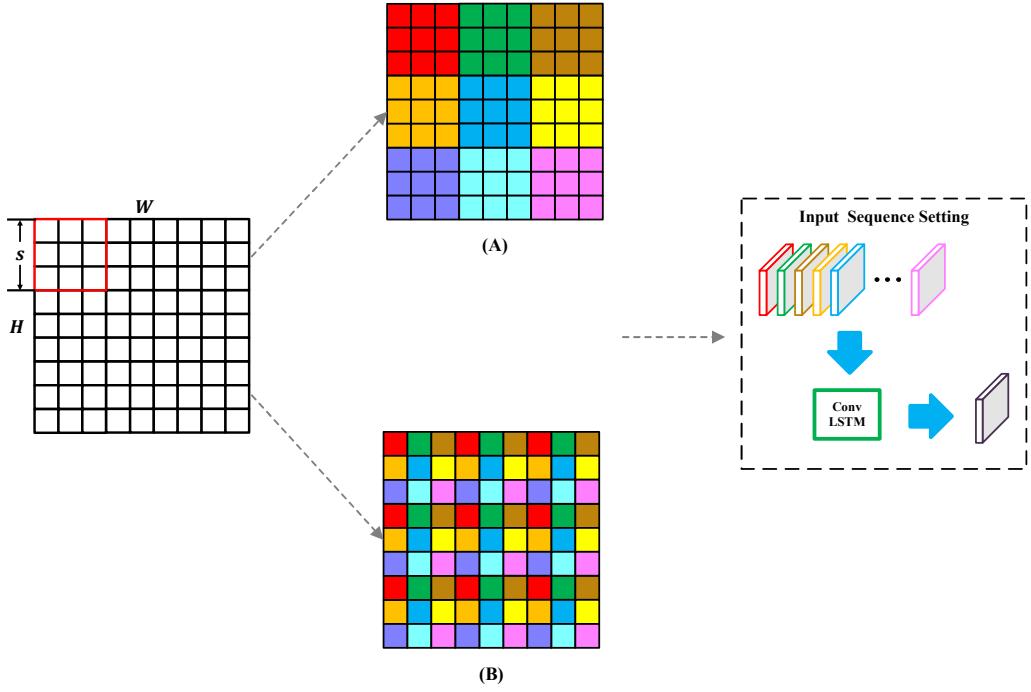


图 5.3: 两种典型的空间分组策略: 传统分组策略 (A) 和我们提出的策略 (B), 在相同时间步输入的 ConvLSTM 的像素被表示为相同的颜色。

一个时间步。通过这种方式, 我们可以在这两个方向获得  $H/s$  和  $W/s$  组, 因此每个补丁一共可以得到  $H \cdot W/s^2$  个组。为了完成目标, 一个自然的想法是首先在一个方向对这些像素按顺序分组, 然后在另一个方向重复这一过程, 相应的结果如图5.3(A) 所示。然而, 不同于在狭窄波长范围内获得的彼此之间具有强相关性的光谱。因为受到高光谱图像空间分辨率的限制, 空间区域之间可能关联较弱, 而且很难具有有意义的空间上下文信息如果它们不邻接的话, 尤其是大尺度区域。例如, 对于计算机来说很难辨识出图5.3(A) 中红色与棕色区域之间的关联, 因为它们彼此相距较远。

受 [39] 用来处理光谱向量所提出的一维交替分组策略的启发, 我们通过在行列方向依次执行这一策略将这一想法拓展到多维对空间补丁进行分组以导入 ConvLSTM 中, 如图5.3(B) 所示。可以看到我们提出的方法确保不同时间步的输入能够更好的交流, 因为除了边界像素以外几乎所有的像素都与其他所有时间步的输入像素相邻, 表明不同时间步的数据在特征空间有着更紧密的分布, 从而便于 ConvLSTM 提取邻接信息。此外, 与策略 (A) 相比, 我们提出的策略 (B) 使得所有时间步的数据能够覆盖更大的范围, 更有助于 ConvLSTM 感受全局信息。在完成分组后, 得到的空间序列数据可以被送入 ConvLSTM。一般的方法是将这些分组后的数据依次输入 ConvLSTM,

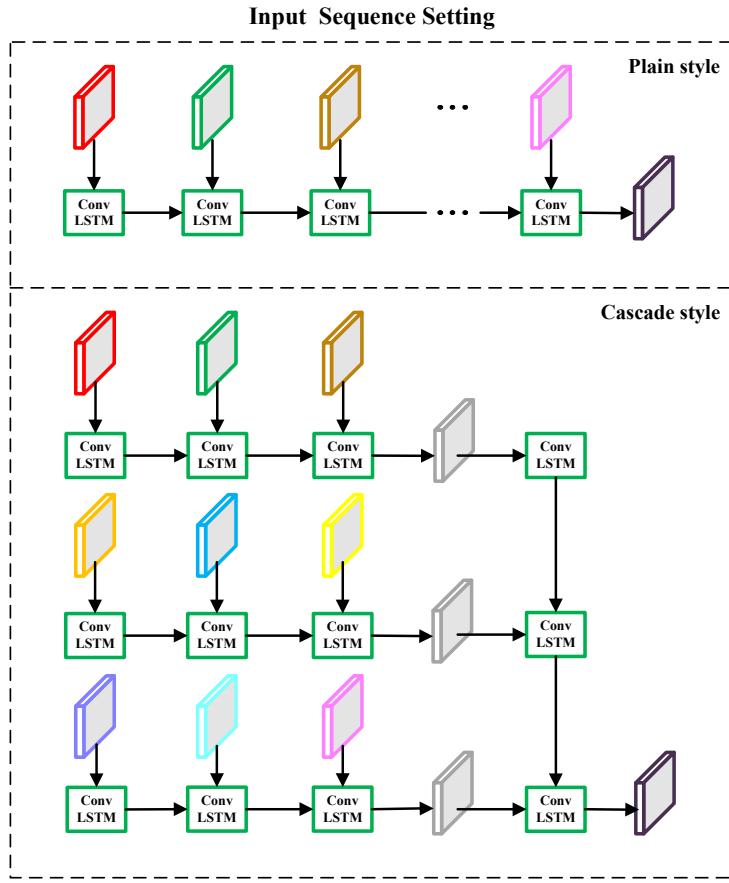


图 5.4: 两种输入序列: 朴素式是将分组后的补丁直接输入 ConvLSTM 的传统方法; 本文提出的级联式可以通过层次化的结构提取特征。

如图5.4所示，时间步数与分组数相同。然而，这种方法忽略了本质的空间结构并且导致 ConvLSTM 仅能探测到一个方向空间上下文信息。换句话说，另一个方向的空间上下文信息丢失掉了。例如，如果 ConvLSTM 的输入序列是像图5.4的朴素式的话，那么 ConvLSTM 将不能在行方向有效的提取到图5.3中红色和橙色区域的关系。为了获得在这两个方向上的上下文信息，可以分别在这两个方向上直接应用朴素式方案。但是它不可避免的增加了计算的负担因为大量数据被重复使用。另外，这种长序列数据可能会导致梯度爆炸或者梯度消失，并有过拟合的风险。

为了解决这一问题，我们提出了一种叫“级联式”的解决方案，它可以平衡计算效率和两个方向的关系型信息提取，如图5.4所示。如果行方向有  $H/s$  个组的话，这个方案将会首先使用  $H/s$  个时间步数为  $W/s$  的 ConvLSTM 来捕捉序列信息，并得到  $H/s$  个上下文特征。接着使用另一个具有  $H/s$  时间步的 ConvLSTM 在行方向来提取这  $H/s$  个列方向上上下文特征的关联。经过这个级联的操作，原始输入空间补丁的整

体上下文信息被逐步获得。它的本质在于采用层级化结构从先前小范围的特征来获得更大范围下的特征。

接下来的步骤与 MSCNN 相同，在所有分支提取出的空间上下文特征将会被展平并送入相应的全连接层中。每个分支的结果通过相加结合。最终的分类图则通过另一个全连接层和 softmax 函数获得。

然而，ConvLSTM 中存在大量的线性和非线性计算，具有复杂的数据转换，并有过拟合的风险。因此，有必要采用正则化约束。BN 和 Dropout 是广泛使用的正则化器，并且它们已经应用在本文提出的 SaMN 中。Dropout 用在展平的特征后以加速计算和并使得训练随机化。BN 层则被嵌入到所有分支的 ConvLSTM 中的卷积层后，因为网络计算过程主要存在于 ConvLSTM。另外，ConvLSTM 中所有的双曲正切函数  $F_h$  被替换为 ReLU 以提供更简单的映射并避免过拟合。

### 5.2.1 特征融合模块

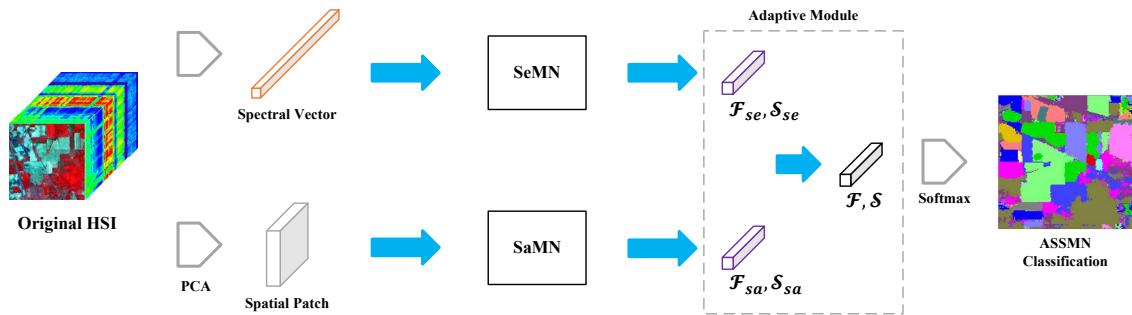


图 5.5: 本文提出的自适应空谱多尺度框架，包括光谱和空间子网络 SeMN 和 SaMN。提取到的特征通过一个自适应模块几何并产生最终的分类图。

在分别通过上述的子网络 SeMN 和 SaMN 得到多尺度光谱和空间上下文特征后，可以在这些特征的基础上构建出一个自适应的空谱联合分类网络 ASSMN，其核心问题在于如何有效的结合这些特征。“自适应”的意思是这个结合的过程是自动的。图5.5展示了本文提出的 ASSMN 和通过 SeMN 及 SaMN 中的加法操作及全连接层之后得到的特征。此外，这些特征已经在图5.1和图5.2中用紫色表示。在这些特征中， $\mathcal{F}_{se}, \mathcal{F}_{sa}$  是 SeMN 及 SaMN 中获得的特征，而  $\mathcal{S}_{se}, \mathcal{S}_{sa}$  则是相应的长度与类别数相同在 softmax 函数之前的分数向量。每个神经元的值反映了对应类别的重要性。通常，最终的联合分数向量  $\mathcal{S}$  按如下方式获得：

- 特征连接

这种方式在 [39] 中被使用，联合特征  $\mathcal{F}$  通过在通道维直接连接  $\mathcal{F}_{se}$  和  $\mathcal{F}_{sa}$  得到， $\mathcal{S}$  通过全连接层获得。

$$\begin{aligned}\mathcal{F} &= \text{Concat}(\mathcal{F}_{se}, \mathcal{F}_{sa}) \\ \mathcal{S} &= W\mathcal{F} + b\end{aligned}\tag{5.1}$$

上式中  $W, b$  是全连接层的权值矩阵和偏置 (下同)。

- 特征加权相加

这个方法由 [66] 提出，通过学习一个权重参数  $m$  得到联合特征  $\mathcal{F}$ ，分数向量  $\mathcal{S}$  通过与特征连接相同的方法获得。

$$\begin{aligned}\mathcal{F} &= F_s(m) \cdot \mathcal{F}_{se} + (1 - F_s(m)) \cdot \mathcal{F}_{sa} \\ \mathcal{S} &= W\mathcal{F} + b\end{aligned}\tag{5.2}$$

上式中 sigmoid 函数  $F_s$  用来将权值参数  $m$  变换到 (0,1) 范围内。事实上，特征相加可以看做这种方式的一种特殊情形，此时两边权重相等。

然而，我们认为特征加权相加似乎看起来并不能比特征连接表现更好，这个结论可以简单的证明如下：特征连接过程可以展开为

$$\begin{aligned}\mathcal{S} &= W\mathcal{F} + b \\ &= W_1\mathcal{F}_{se} + W_2\mathcal{F}_{sa} + b_1 + b_2 \\ &= A_1 + b_1 + A_2 + b_2\end{aligned}\tag{5.3}$$

其中， $W = \text{Concat}(W_1, W_2)$ ,  $b = b_1 + b_2$ ,  $W_1, W_2$  是两个可以学到的不同的矩阵，而特征加权相加为

$$\begin{aligned}\mathcal{S} &= W\mathcal{F} + b \\ &= F_s(m) \cdot W\mathcal{F}_{se} + (1 - F_s(m)) \cdot W\mathcal{F}_{sa} + b\end{aligned}\tag{5.4}$$

以上两式显示，特征加权相加与特征连接形式很类似，或许可能效果逊于后者，因为两个不同的特征  $\mathcal{F}_{se}, \mathcal{F}_{sa}$  通过一个相同的矩阵  $W$  进行转换而非两个相应的矩阵  $W_1, W_2$ ，这意味着得到的融合特征  $\mathcal{F}$  与特征连接相比可能数据表示能力较弱。

考虑到上述问题，并注意到光谱和空间部分不同的重要性。我们认为应该“分数加权”而非特征加权。这可以理解为联合分数向量  $\mathcal{S}$  通过对  $\mathcal{S}_{se}$  和  $\mathcal{S}_{sa}$  加权得到，这

一过程可以公式化为

$$\begin{aligned}
 \mathcal{S} &= F_s(m)\mathcal{S}_{se} + (1 - F_s(m))\mathcal{S}_{sa} \\
 &= F_s(m) \cdot W_{se}\mathcal{F}_{se} + (1 - F_s(m)) \cdot W_{sa}\mathcal{F}_{sa} + \\
 &\quad F_s(m) \cdot b_{se} + (1 - F_s(m)) \cdot b_{sa} \\
 &= \lambda_1(A_1 + b_1) + \lambda_2(A_2 + b_2)
 \end{aligned} \tag{5.5}$$

上式中  $\lambda_1 + \lambda_2 = 1$ ,  $W_{se}, W_{sa}, b_{se}, b_{sa}$  是  $\mathcal{F}_{se}, \mathcal{F}_{sa}$  的转换矩阵和偏置向量。与式 (5.4) 和式 (5.5) 相比, 可以看到我们提出的方式不仅保留了两个不同的矩阵, 也强调了光谱和空间两部分不同的重要性, 并且所有参数均可以通过网络训练学习到从而自适应的确定这两部分的比重。这可以看作特征连接的高级版本, 因此赋予了模型比特征加权相加更强的数据表征能力。

最后, 我们使用了一个多任务交叉熵损失函数来优化提出的 ASSMN, 该函数定义如下

$$\begin{aligned}
 L &= L_{\mathcal{S}} + L_{\mathcal{S}_{se}} + L_{\mathcal{S}_{sa}} = \\
 &- \frac{1}{N} \sum_{i=1}^N [z_i^{\mathcal{S}} \log(y_i) + (1 - z_i^{\mathcal{S}}) \log(1 - y_i)] \\
 &- \frac{1}{N} \sum_{i=1}^N [z_i^{\mathcal{S}_{se}} \log(y_i) + (1 - z_i^{\mathcal{S}_{se}}) \log(1 - y_i)] \\
 &- \frac{1}{N} \sum_{i=1}^N [z_i^{\mathcal{S}_{sa}} \log(y_i) + (1 - z_i^{\mathcal{S}_{sa}}) \log(1 - y_i)]
 \end{aligned} \tag{5.6}$$

上式中,  $N$  是训练样本数,  $L_{\bullet}$  和  $z_i^{\bullet}$  分别表示损失和概率向量 ( $\bullet = \mathcal{S}, \mathcal{S}_{se}, \mathcal{S}_{sa}$ ), 它可以通过 softmax 函数计算。例如,  $z_{ij}^{\mathcal{S}} = \frac{\exp(\mathcal{S}_{ij})}{\sum_j \exp(\mathcal{S}_{ij})}$ .  $y_i$  是第  $i$  个样本的真实标签。

### 5.3 实验及分析

本节将在 IP, PU 和 Kennedy Space Center (KSC) 三个公共的高光谱图像分类数据集上对提出的方法进行一系列定性和定量的综合性分析。

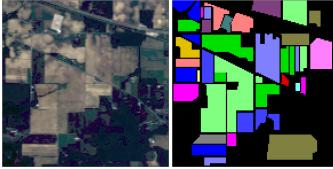
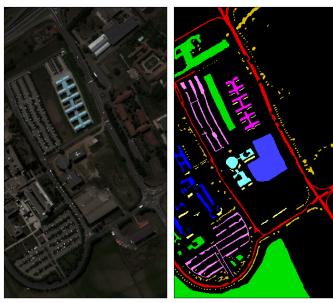
Dataset ID	Original Image and Ground Truth	Class ID	Color	Class Name	Training	Testing	Total
(a)		1		Alfalfa	33	13	46
		2		Corn-notill	100	1328	1428
		3		Corn-mintill	100	730	830
		4		Corn	100	137	237
		5		Grass-pasture	100	383	483
		6		Grass-trees	100	630	730
		7		Grass-pasture-mowed	20	8	28
		8		Hay-windrowed	100	378	478
		9		Oats	14	6	20
		10		Soybean-notill	100	872	972
		11		Soybean-mintill	100	2355	2455
		12		Soybean-clean	100	493	593
		13		Wheat	100	105	205
		14		Woods	100	1165	1265
		15		Buildings-Grass-Trees-Drives	100	286	386
		16		Stone-Steel-Towers	75	18	93
		Total			1342	8907	10249
(b)		1		Asphalt	100	6531	6631
		2		Meadows	100	18549	18649
		3		Gravel	100	1999	2099
		4		Trees	100	2964	3064
		5		Painted metal sheets	100	1245	1345
		6		Bare Soil	100	4929	5029
		7		Bitumen	100	1230	1330
		8		Self-Blocking Bricks	100	3582	3682
		9		Shadows	100	847	947
		Total			900	41876	42776
(c)		1		Scrub	33	728	761
		2		Willow swamp	23	220	243
		3		CP hammock	24	232	256
		4		CP/Oak	24	228	252
		5		Slash pine	15	146	161
		6		Oak/Broadleaf	22	207	229
		7		Hardwood swamp	9	96	105
		8		Graminoid marsh	38	393	431
		9		Spartina marsh	51	469	520
		10		Catiai marsh	39	365	404
		11		Salt marsh	41	378	419
		12		Mud flats	49	454	503
		13		Water	91	836	927
		Total			459	4752	5211

图 5.6: 三个数据集的信息: (a) IP, (b) PU, (c) KSC, 包含假彩色图, 相应的地面真值以及训练和测试集设置。

### 5.3.1 数据集描述

- 1) IP: 该数据集由机载可见/红外成像光谱仪 (Airborne Visible/Infrared Imaging Spectrometer, AVIRIS) 于 1992 年收集于印第安纳州西北部, 有 200 个波段, 尺寸为  $145 \times 145$ , 在去除了水吸收波段后空间分辨率为 20m, 波长范围  $0.4\text{-}2.5\mu\text{m}$ 。该场景涉及 16 个植被类别, 相关信息 (如原始图像可视化, 地面真实情况及训练测试样本的配置) 已经显示在图5.6(a) 中显示。
- 2) PU: 该数据集是通过反射光学系统成像仪 (Reflective Optics System Imaging Spectrometer, ROSIS) 于 2001 年在意大利北部的帕维亚大学获得的, 该数据集有 103 个波段, 尺寸为  $610 \times 340$ , 空间分辨率为 1.3m, 波长范围为  $0.43\text{-}0.86\mu\text{m}$ 。该数据集包括 9 个类别, 如图5.6(b) 所示。
- 3) KSC: 该场景由 AVIRIS 传感器于 1996 年在佛罗里达州的肯尼迪航天中心收集。在去除水吸收和低 SNR 波段后, 剩下的 176 个波段波长范围位于  $0.4\text{-}2.5\mu\text{m}$ 。该数据集中包含  $512 \times 614$  个像素, 空间分辨率为 18m。定义了该场景下 13 个表示各

种土地覆盖的类别。详细信息如图5.6(c) 所示。

### 5.3.2 模型和实验设置

在定量分析之前，有必要介绍关于所提出方法和接下来实验的一些重要的参数和设置，这些设置将全程固定不变。

#### 5.3.2.1 光谱多尺度子网络设置

表 5.1: SeMN 中方案二的设置

网络层	卷积核数量	大小	步长	填充	激活函数
卷积层	1	3	1	1	No
最大池化层	/	2	2	/	No
全连接层	128	128	/	/	ReLU

在导入 SeMN 之前，原始的光谱向量均归一化到 [-1,1]。所有 LSTM 隐层的神经元数均设置为 128。另外，方案二中使用一维卷积神经网络来获得多尺度信息，相关信息如表5.1。步长是卷积或池化层中滑动窗口在通道维每步移动的像素数。填充则维持了一维特征图的尺寸。此外，卷积核数量和全连接层的大小均被设为 128，因为全连接层可以被视为一个大小等于一维输入特征的神经元数的一维卷积层，并且输出的一维特征的长度是卷积核数量。卷积层后没有加非线性映射 ReLU 因为卷积操作仅用于扩大感受野。

#### 5.3.2.2 空间多尺度子网络设置

SaMN 的基本设置几乎与 MSCNN [39] 相同，经过 PCA 降维后的四个主成分被归一化到 [-1,1]，并输入 SaMN。另外，补丁的大小设为  $27 \times 27$ (MSCNN 是  $28 \times 28$ ) 因为我们考虑了空间补丁相对于目标中心像素的对称性，补丁如果尺寸为偶数则不能满足这一性质。

表5.2列出了 SaMN 不同种类网络层的设置。扩张率是计算卷积时两个卷积核内部两个相邻像素之间的内部距离 (标准卷积层中该值为 1)。另外，全连接层的激活函数被标为 “ReLU/No” 因为 ReLU 函数仅用在相加之后的全连接层。

表 5.2: SaMN 的设置

网络层	卷积核数量	大小	步长	填充	扩张率	激活函数
卷积层	32	$3 \times 3$	1	1	1	ReLU
扩张卷积层	32	$3 \times 3$	1	2	2	ReLU
全连接层	128	/ <sup>1</sup>	/	/	/	ReLU/No
ConvLSTM	32	$3 \times 3$	1	1	1	No <sup>2</sup>

<sup>1</sup> 全连接层的大小主要由输入特征的长度确定。这个参数由于  $H/s$  是可变的所以还不能确定,  $H/s$  的确定将在之后的内容讨论。

<sup>2</sup> ReLU 函数是嵌入 ConvLSTM 中而非加在后边激活 ConvLSTM 的输出。

### 5.3.2.3 实验设置

一些重要的参数在所有实验中都是固定不变的。在 SeMN 中, 学习率除了方案二中在 IP 数据集中被设置为 1e-3, 其余情形均为 1e-2, 而 SaMN 和 ASSMN 中所有的学习率均为 1e-2。这些学习率每过 30 个 epoch 减半。SeMN 和 SaMN 分别使用小批量 Adam 和 SGDM 算法优化, LSTM 中的参数使用 BPTT 算法更新。这种安排充分利用了 Adam 和 SGDM 各自的快速收敛和高精度特性 [67]。因为前者更适于在少量数据中找到一个相对较好的局部最优, 而 SGDM 用于 SaMN 是为了精细调整从而获得高精度。另外, SGDM 算法也用于 ASSMN, 因为空间部分在高光谱图像空谱联合分类中更重要 [39]。训练时的 epoch 设置为 200, batch size 设置为 16。所有的实验均重复运行 10 次。每次尝试中训练样本都是随机选取, 剩下的组成测试集。OA, AA 和 Kappa 系数 ( $\kappa$ ) 用作定量评估指标, 均值  $\mu$  和标准差  $\sigma$  均被记录。所有的实验运行在 Inter Xeon Sliver 4114 2.20GHz 处理器和一张 NVIDIA RTX 2080 Ti GPU 上。

### 5.3.3 光谱多尺度子网络分析

#### 5.3.3.1 方案一 V.S. 方案二

考虑到高光谱图像的波段数, SeMN 的方案一通过融合 LSTM 输出的特征实现, 这些特征源于八等分、四等分、二等分和一等分(原始)的光谱向量, 这一方案被称为方案一 (8-4-2)。为了更公平的对比这两个方案, 池化层后的特征图也被依次等分成八部分、四部分和两部分并送入 LSTM。原始光谱本体也被输入另一个 LSTM 并且输出的特征与前述三个特征融合用于分类。这一方案叫做方案二 (8-4-2)。

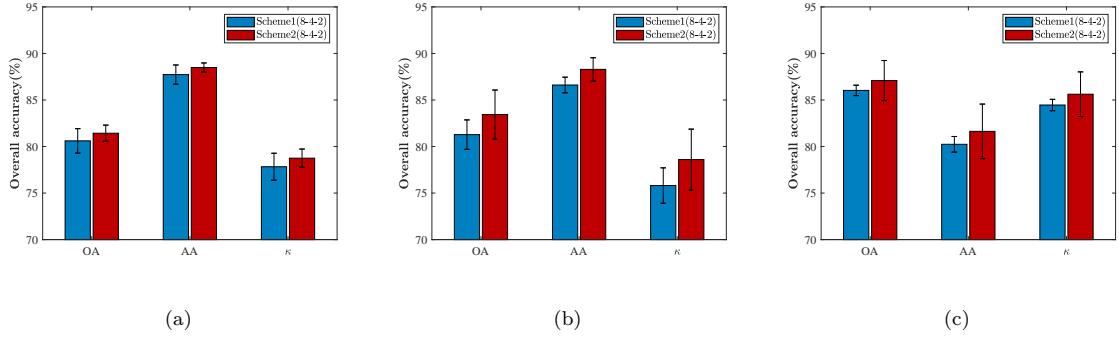


图 5.7: 方案一 (8-4-2) 与方案二 (8-4-2) 在三个数据集上的性能比较. (a) IP. (b) PU. (c) KSC.

图5.7展示了上述两方案在固定参数下的定量评估结果。显然方案二 (8-4-2) 优于方案一 (8-4-2) 因为前者在所有的测度和所有的数据集下表现更好，因为相邻时间步输入的感受野存在交集，意味着特征空间中更近的距离，从而便于 ConvLSTM 提取上下文信息。

然而，与方案一 (8-4-2) 相比，方案二 (8-4-2) 具有大幅度波动，这表明后者不太稳定。这些波动由图5.7的黑线显示，其表示了精度浮动的上下界  $\mu \pm \sigma$ 。

通常在分组过程中配置所有特征图的分组数是笨拙且不明智的。让所有特征的分组数相同，或者说让所有的 LSTM 均有相同的时间步数则更为简便。为了探究出一种对于方案二更有效的设置，它不仅优于方案二 (8-4-2)，同时更稳定也更简便，我们接下来分析了时间步数  $T_{se}$  和分类质量之间的关系。

### 5.3.3.2 SeMN 中方案二的参数分析

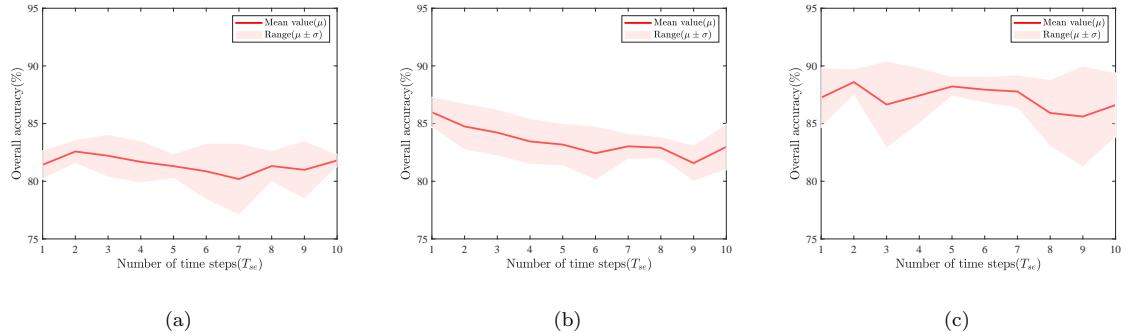


图 5.8: 不同时间步数下方案二在三个数据集上的表现. (a) IP. (b) PU. (c) KSC.

图5.8展示了 OA 随方案二中不同时间步数  $T_{se}$  在三个数据集上的变化。可以看到

随着  $T_{se}$  增加, OA 均值  $\mu$  在 IP 和 KSC 数据集上变化并不明显并有小幅度波动。而在 PU 数据集上, OA 随着  $T_{se}$  增加而下降。因此, 方案二中的  $T_{se}$  被设置为 2. 另外, 相比于其他数值, 当  $T_{se}$  设置为 2 时, OA 的波动范围也相对较小, 这表示方案二在当前设置下更稳定, 此时也是符合上述要求的最合适的选择。这时的模型可以表示为方案二 (2-2-2)。此外, 额外的实验也展示了方案二 (2-2-2) 优于方案二 (8-4-2), 结果如表5.3。

表 5.3: 方案二 (2-2-2) 相比于方案二 (8-4-2) 的结果变化

性能提升	IP		PU		KSC	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
OA(%)	+1.06	-0.01	+2.18	-0.61	+0.70	-0.32
AA(%)	+0.24	+0.81	+0.66	-0.31	+1.11	-0.31
$\kappa(\times 100)$	+1.19	-0.01	+2.69	-0.77	+0.80	-0.36

表5.3定量的描述了方案二 (2-2-2) 相比于方案二 (8-4-2) 在这三个数据集上 OA, AA 和  $\kappa$  的精度变化。可以看到几乎所有的  $\mu$  得到了提升而  $\sigma$  有所下降, 这表明方案二 (2-2-2) 可以取得高精度同时具有强鲁棒性。方案二 (2-2-2) 将会作为后续实验中 ASSMN 的光谱部分。

### 5.3.4 空间多尺度子网络分析

#### 5.3.4.1 参数估计

SaMN 有两个参数  $H/s$  和  $p$  需要调整。 $H/s$  是每个空间补丁在行列方向获得的组数,  $H, s$  分别是输入补丁和分组后各部分的尺寸。 $H/s$  事实上表示 ConvLSTM 的时间步数, 为了方便, 这个参数用  $T_{sa}$  表示。另一个参数  $p$  是 Dropout 的概率, 它是特征融合前全连接层中不工作的神经元占全部神经元的比例。因为太多或者太少的神经元会导致过拟合或欠拟合现象, 有必要使用适当数量的神经元用于特征转换。

图5.9展示了不同  $T_{sa}$  和  $p$  下的 OA 变化,  $T_{sa}$  和  $p$  的变化范围分别设置为 [1,10] 和 [0,0.9], 步长分别为 1 和 0.1。除此以外, 为了更好地展示这些参数对于模型表现的影响, 图片通过三次插值平滑。可以清晰地看到高精度区域在这些图片中均位于右下部。在图片上部, SaMN 不稳定且经常表现较差, 尤其是在右上角。此外, 相比于其

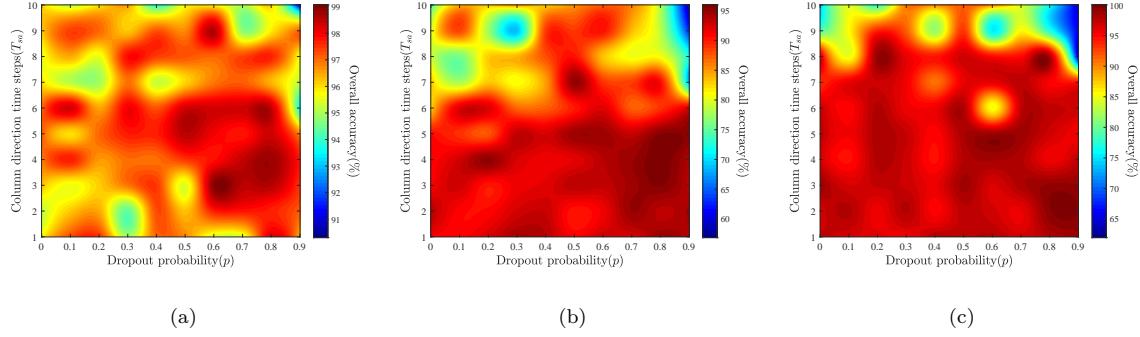


图 5.9: 不同的列方向时间步数和 Dropout 概率下 SaMN 在三个数据集上的表现. (a) IP. (b) PU. (c) KSC.

他数据集, SaMN 在 IP 数据集上具有更高的 OA。本文中  $T_{sa}$  和  $p$  分别设置为 3 和 0.8, 该配置下的 SaMN 作为后续实验中 ASSMN 的空间部分。

#### 5.3.4.2 分组策略和输入序列设置

除了估计  $T_{sa}, p$ , 我们也测试了提出的分组策略和输入序列的设置, 并与传统的分组策略及朴素式进行对比。结果如图5.10所示。我们提出的交替分组策略表示为“交替”(Alter), 而传统分组策略则被表示为“传统”(Trad)。“朴素”(Plain) 和“级联”(Cas) 分别用来表示输入序列设置中的朴素式和级联式。图5.10表明, 本文提出的策略取得

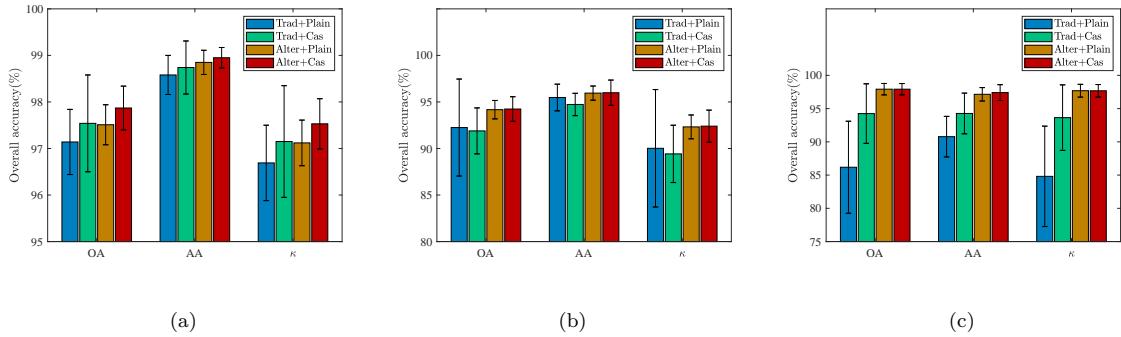


图 5.10: 不同的空间分组和输入序列设置下 SaMN 在三个数据集上的表现 (a) IP. (b) PU. (c) KSC.

最好的效果。与传统策略相比, 它具有更高的均值和更低的标准差, 尤其是在 PU 和 KSC 数据集上。这些结果验证了模型在提出的交替策略下更有效更稳健的结论, 因为 ConvLSTM 在每个时间步的当前输入能够覆盖更大的范围, 便于 ConvLSTM 从相邻时间步中提取上下文信息。另外, 凭借着层次性的上下文特征提取, 当输入序列使用级联式时分类精度进一步提高。

### 5.3.4.3 消融实验

表 5.4: 不同设置下 SaMN 的总体精度 (%)

方法	IP	PU	KSC
主干网络	95.72	84.42	89.23
主干网络 +ConvLSTM(无 BN 层)	92.47	86.01	94.64
主干网络 +ConvLSTM+BN	96.53	90.82	96.25
主干网络 +ConvLSTM+Dropout	95.17	85.61	96.04
主干网络 +ConvLSTM+BN+Dropout	97.46	94.56	98.11

为了探索 SaMN 中提出的上下文信息感知部分对分类的贡献及正则化的影响，我们进行了一系列的消融实验。相应的结果如表5.4所示。“主干网络”意为扩张卷积后被激活的特征图将被直接展平并送入全连接层而非 ConvLSTM。

表5.4的第二行显示了单纯使用 ConvLSTM 确实能够提高在 PU 和 KSC 数据集上的分类结果。在 BN 的帮助下，所有数据集上的精度均进一步提高，意为 BN 很好的解决了包含大量复杂转换的 ConvLSTM 中的过拟合问题。在 BN 的基础上使用 Dropout 比单独使用时表现更好。

### 5.3.5 自适应空谱多尺度网络分析

#### 5.3.5.1 特征融合评价

表 5.5: ASSMN 中不同特征融合方式的总体精度 (%)

方法	IP	PU	KSC
特征连接	98.29	95.74	98.35
特征加权相加	98.17	95.80	97.88
分数加权相加	98.30	96.26	98.44

在这部分，我们比较了不同的特征融合方式，相应的结果如表5.5所示。实验结果证明了我们的假设即“特征连接”优于“特征加权相加”。可以观察到与其他两个方法相比，凭借着最强的表示能力，我们的方法在全部三个数据集上表现均达到最好，这显示了分数融合的合理性与优越性。

### 5.3.5.2 分类精度评定

表 5.6: 光谱组不同方法在 IP 数据集上的实验表现

类别	光谱组					
	SVM	1-DCNN	LSTM-策略一	LSTM-策略二	SeMN-方案一	SeMN-方案二
OA (%)	79.02±0.92	75.40±4.18	75.99±0.87	77.96±2.04	80.61±1.31	<b>82.49±0.86</b>
AA (%)	87.08±1.19	85.05±1.77	82.55±1.36	84.55±1.74	87.73±1.03	<b>88.73±1.30</b>
$\kappa \times 100$	76.04±1.04	72.06±4.56	72.63±0.99	74.81±2.29	77.83±1.45	<b>79.94±0.97</b>
$T_{trn}$ (s)	<b>0.14±0.00</b>	255.91±3.16	119.45±2.75	118.95±3.16	177.19±11.44	218.75±2.64
$T_{tes}$ (s)	<b>1.68±0.02</b>	0.46±0.34	2.05±0.83	1.90±0.71	0.53±0.07	0.52±0.07
1	94.62	89.23	81.54	87.69	92.31	90.00
2	75.02	72.97	69.98	72.15	74.74	77.68
3	76.51	74.01	73.51	75.66	76.90	76.03
4	90.00	84.31	84.31	85.62	88.25	89.05
5	92.98	90.99	89.90	88.83	92.09	93.52
6	95.11	94.14	91.16	92.57	94.54	95.62
7	90.00	95.00	88.75	90.00	91.25	90.00
8	97.86	97.94	95.34	95.48	97.96	98.28
9	90.00	85.00	73.33	81.67	95.00	93.33
10	79.64	78.97	79.79	81.69	85.38	87.02
11	64.64	59.71	63.49	68.06	69.54	72.41
12	84.89	82.76	75.88	83.57	84.91	87.81
13	98.95	99.14	98.48	96.29	98.19	99.24
14	88.87	77.48	85.05	82.11	87.12	88.44
15	76.40	79.65	73.04	71.92	78.85	81.33
16	97.78	99.44	97.22	99.44	96.67	100.00

我们将本文提出的方法 SeMN, SaMN 和 ASSMN 同时与几个先进的深度学习方法进行比较, 包括一维卷积神经网络 (1-DCNN), 二维卷积神经网络 (2-DCNN), 三维卷积神经网络 (3-DCNN) [31], LSTM-策略一, LSTM-策略二, MSCNN, SSUN [39], SSRN [32]。SeMN 的两个方案均在对比实验中进行了详细的评估。除了这些深度学习方法, 作为传统方法代表的 SVM 也被评估。我们使用归一化到 [-1,1] 之间的原始光谱向量进行 SVM 分类。本文中, SVM 使用径向基核, 并通过网格搜索及五折交叉验证在  $\{2^{-5}, 2^{-4}, \dots, 2^{19}\}$  和  $\{2^{-15}, 2^{-14}, \dots, 2^5\}$  的范围 [31] 内确定两种重要的超参数  $C$  和  $\gamma$ 。整个过程利用 LIBSVM [68] 实现。SSRN, SSUN 和它的子网络 (LSTM-策略一, LSTM-策略二, MSCNN) 使用 Tensorflow 1.14.0 实现, 其他的深度学习方法则利

表 5.7: 空间组、联合组不同方法在 IP 数据集上的实验表现

类别	空间组			联合组			
	2-DCNN	MSCNN	SaMN	3-DCNN <sup>1</sup>	SSRN	SSUN	ASSMN
OA (%)	86.20±1.50	96.58±0.79	<b>97.46±0.52</b>	96.57±1.12	96.62±0.75	96.75±0.65	<b>98.30±0.51</b>
AA (%)	94.58±0.75	98.52±0.30	<b>98.85±0.16</b>	98.53±0.41	95.87±1.18	98.64±0.23	<b>99.09±0.36</b>
$\kappa \times 100$	84.20±1.69	96.05±0.90	<b>97.05±0.60</b>	96.04±1.29	96.09±0.86	96.24±0.75	<b>97.03±0.59</b>
$T_{trn}$ (s)	<b>57.42±2.80</b>	100.70±5.57	1301.53±57.29	34686.45±1262.40	869.76±31.28	<b>216.98±3.28</b>	1430.76±43.07
$T_{tes}$ (s)	<b>1.17±0.59</b>	1.83±0.64	5.14±1.34	144.42±10.11	4.02±0.33	<b>2.77±0.69</b>	4.75±1.07
1	100.00	100.00	100.00	100.00	100.00	100.00	99.23
2	82.27	95.86	94.33	94.65	95.26	95.74	96.48
3	92.92	97.16	98.55	97.30	96.96	97.55	98.68
4	99.64	99.85	100.00	100.00	95.20	100.00	99.71
5	93.58	98.62	99.37	98.98	98.88	97.96	98.88
6	96.52	99.02	99.37	99.57	98.51	99.67	99.97
7	100.00	100.00	100.00	100.00	95.71	100.00	98.75
8	99.63	99.95	100.00	100.00	100.00	100.00	100.00
9	100.00	100.00	100.00	100.00	100.00	100.00	100.00
10	89.39	95.67	96.59	96.66	87.79	95.92	98.20
11	69.85	93.46	96.41	93.88	99.01	93.71	97.29
12	96.31	98.50	97.81	98.34	96.59	99.23	99.23
13	99.90	100.00	100.00	100.00	99.89	100.00	99.52
14	93.93	98.55	99.23	97.76	99.54	98.52	99.45
15	99.37	99.72	100.00	99.97	94.10	100.00	100.00
16	100.00	100.00	100.00	99.44	76.45	100.00	100.00

<sup>1</sup> 根据 [31] 的设置，在三个 GPU 上运行因为要求高显存（仅针对 IP 数据集）。

用 Pytorch 1.1.0 完成。

从输入数据和提取到的特征的种类来看，这些方法可以分为三组。像素级 SVM, 1-DCNN, LSTM-策略一, LSTM-策略二和我们提出的 SeMN-方案一, SeMN-方案二可以归于光谱组。2-DCNN, MSCNN 和我们提出的 SaMN 则被划分到空间组，因为他们主要在空间域感知目标地物周围的信息。其余的方法，使用三维卷积同时提取空谱信息的方法 (3-DCNN, SSRN) 或者利用特征融合生成空谱特征 (SSUN, ASSMN) 构成了联合组。

表5.6-5.11定量的展示了不同方法下每个类别的 OA, AA,  $\kappa$ , 训练时间  $T_{trn}$  和测试时间  $T_{tes}$ 。每组中最好的结果用粗体显示。很显然空间组借助周围地物信息表现比光谱组要好，而联合组则自然地在所有数据集的每个指标下均取得了最优。在光谱组，我们提出的方法相比于其他单尺度特征提取网络（例如 1-DCNN 和两种分钟策略下的 LSTM）表现更好，因为 SeMN-方案一和 SeMN-方案二更充分的使用了多尺度信息。另外，由于相邻时间步的输入在特征空间中彼此距离更近且过渡更自然，更便于 LSTM 提取上下文信息。因此，SeMN-方案二进一步超过了 SeMN-方案一。需要说明的是，

表 5.8: 光谱组不同方法在 PU 数据集上的实验表现

类别	光谱组					
指标	SVM	1-DCNN	LSTM-策略一	LSTM-策略二	SeMN-方案一	SeMN-方案二
OA (%)	<b>88.28±0.61</b>	75.06±2.62	71.46±6.71	67.49±6.57	81.28±1.58	85.62±2.02
AA (%)	<b>90.67±0.56</b>	83.65±0.53	80.96±1.86	72.52±6.49	86.61±0.84	88.95±0.94
$\kappa \times 100$	<b>84.65±0.79</b>	68.44±2.74	64.35±6.71	58.97±7.54	75.81±1.90	81.29±2.50
$T_{trn}$ (s)	<b>0.03±0.00</b>	58.64±3.38	87.60±2.92	86.27±3.37	126.42±5.57	167.19±6.60
$T_{tes}$ (s)	<b>1.66±0.06</b>	1.68±0.76	4.98±0.96	4.34±0.28	2.61±0.20	2.92±0.26
1	81.11	72.81	71.33	76.99	81.32	84.53
2	88.83	68.73	63.40	63.97	78.42	84.14
3	82.85	78.76	69.16	55.98	80.33	81.40
4	95.58	89.89	86.45	91.82	94.13	94.07
5	99.37	99.65	99.33	99.51	98.75	99.04
6	89.57	75.44	72.70	63.29	75.96	84.96
7	93.49	90.39	87.89	55.67	91.42	90.89
8	85.31	77.46	78.64	45.85	79.24	81.68
9	99.95	99.74	99.85	99.63	99.93	99.81

表 5.9: 空间组、联合组不同方法在 PU 数据集上的实验表现

类别	空间组			联合组			
指标	2-DCNN	MSCNN	SaMN	3-DCNN	SSRN	SSUN	ASSMN
OA (%)	71.67±12.12	94.26±1.03	<b>94.56±1.67</b>	92.98±1.01	96.17±3.27	95.36±0.72	<b>96.26±1.08</b>
AA (%)	80.76±4.01	96.01±0.50	<b>96.25±1.10</b>	95.13±0.49	96.30±2.91	97.09±0.42	<b>98.12±0.32</b>
$\kappa \times 100$	64.72±12.43	92.45±1.32	<b>92.82±2.15</b>	90.76±1.28	95.00±4.21	93.89±0.93	<b>95.06±1.40</b>
$T_{trn}$ (s)	<b>46.92±1.54</b>	55.93±6.29	892.33±20.92	2602.84±164.34	579.45±13.06	<b>157.72±4.41</b>	976.86±22.31
$T_{tes}$ (s)	<b>1.61±0.82</b>	3.26±0.67	21.43±2.19	161.40±8.13	21.02±0.78	<b>8.37±0.86</b>	20.21±1.99
1	70.94	91.52	94.22	90.16	99.21	93.73	96.80
2	66.66	92.85	93.45	92.23	99.61	93.74	94.06
3	75.10	93.23	91.51	92.62	95.11	94.33	97.95
4	89.02	97.49	98.49	96.39	89.56	98.49	99.21
5	99.97	99.54	99.97	99.94	99.99	100.00	100.00
6	51.23	98.05	93.03	91.63	92.97	98.53	97.92
7	84.74	97.77	99.33	97.88	97.77	98.96	99.54
8	90.07	95.25	96.81	95.57	92.57	96.56	97.68
9	99.08	98.41	99.47	99.75	99.95	99.47	99.94

表 5.10: 光谱组不同方法在 KSC 数据集上的实验表现

类别	光谱组					
	SVM	1-DCNN	LSTM-策略一	LSTM-策略二	SeMN-方案一	SeMN-方案二
OA (%)	<b>90.84±0.72</b>	78.44±2.54	78.88±6.18	77.86±4.45	86.03±0.56	87.79±1.83
AA (%)	<b>86.48±0.68</b>	69.14±3.11	70.32±9.35	68.13±7.57	80.24±0.84	82.74±2.62
$\kappa \times 100$	<b>89.80±0.79</b>	75.99±2.76	76.45±6.91	75.24±5.00	84.46±0.61	86.42±2.04
$T_{trn}$ (s)	<b>0.02±0.00</b>	68.18±2.27	49.27±3.00	49.28±4.45	64.14±3.79	81.69±4.34
$T_{tes}$ (s)	<b>0.17±0.01</b>	1.03±0.51	1.42±0.70	1.32±0.55	0.30±0.05	0.33±0.10
1	89.73	79.59	83.54	88.01	83.05	84.23
2	89.14	87.73	72.50	83.59	84.14	88.00
3	91.68	37.89	56.77	35.39	83.49	86.85
4	78.07	29.69	46.89	29.04	56.01	63.11
5	56.85	41.37	40.55	44.93	51.03	58.97
6	60.82	40.00	37.44	40.19	51.55	53.04
7	83.85	51.67	59.38	53.44	74.79	76.25
8	90.59	70.66	59.11	53.79	84.86	85.06
9	95.95	87.33	79.53	81.60	90.55	93.35
10	97.10	92.49	93.62	91.34	94.77	96.11
11	96.64	95.56	97.04	96.85	96.56	96.88
12	93.79	84.87	87.80	87.51	92.58	93.79
13	100.00	99.92	99.99	100.00	99.93	99.95

表 5.11: 空间组、联合组不同方法在 KSC 数据集上的实验表现

类别	空间组			联合组			
	2-DCNN	MSCNN	SaMN	3-DCNN	SSRN	SSUN	ASSMN
OA (%)	93.05±0.77	97.75±0.68	<b>98.11±0.62</b>	97.92±0.76	98.30±1.18	98.13±0.74	<b>98.44±0.92</b>
AA (%)	91.40±1.09	97.13±0.93	<b>97.49±0.87</b>	97.31±1.09	97.51±1.47	97.80±0.67	<b>98.00±1.03</b>
$\kappa \times 100$	92.26±0.86	97.49±0.76	<b>97.90±0.69</b>	97.68±0.84	98.11±1.31	97.91±0.82	<b>98.27±1.02</b>
$T_{trn}$ (s)	<b>24.25±1.17</b>	40.30±5.15	391.65±3.38	4616.92±204.81	332.95±11.31	<b>85.85±4.74</b>	472.07±7.84
$T_{tes}$ (s)	<b>0.98±0.46</b>	1.15±0.45	3.05±1.48	90.30±8.19	2.48±0.18	<b>1.96±0.35</b>	2.27±0.30
1	86.15	95.67	96.24	95.71	99.67	95.71	97.09
2	92.68	93.73	95.95	95.91	98.91	96.14	96.91
3	92.33	96.21	95.60	93.58	98.02	95.95	93.88
4	78.60	90.44	93.55	93.42	86.92	93.11	93.25
5	96.44	97.53	93.70	98.08	96.66	98.15	98.15
6	73.33	97.58	97.68	95.27	94.97	98.60	97.73
7	86.04	95.94	97.60	95.83	94.76	97.81	98.85
8	94.78	97.53	98.07	98.93	98.81	97.99	98.98
9	97.10	99.64	99.70	98.83	99.86	99.06	99.72
10	98.58	99.81	99.92	99.92	99.73	99.21	99.97
11	97.49	100.00	100.00	100.00	99.62	100.00	100.00
12	94.69	98.55	99.38	99.49	99.78	99.71	99.47
13	100.00	100.00	100.00	100.00	99.90	100.00	100.00

在光谱组中 SVM 在 PU 和 KSC 两个数据集的每个指标上均取得最好的成绩，表明 SVM 可以在较短时间达到相对较高的精度，显示了传统方法的优势。但是 SVM 很难作为特征提取器去获取与空间特征融合的光谱特征。在空间组中，在设计好的交替分组策略、级联式输入序列和正则化的帮助下，ConvLSTM 成功提取出了空间上下文信息，并且获得的特征是多尺度的。因此，我们的 SaMN 与其他方法相比表现更好。最短的运行时间由 2-DCNN 取得，然而，因为仅有一个主成分作为网络的输入 [31]，它的表现并不令人满意，尤其是在 PU。在联合组，所有的方法表现都较好，并且我们提出的方法在所有数据集上取得了最佳结果。从运行时间上看，基于三维卷积的方法需要更多的计算和更长的时间。注意到有许多非线性转换和激活函数在 ConvLSTM 中，这限制了 ASSMN 的效率，最短的运行时间由 SSUN 取得。

### 5.3.5.3 分类结果图

图5.11-5.13展示了不同方法的分类图，数据集原图及它们的地面真值。很显然，光谱分类方法趋向于产生强烈的椒盐噪声，使得地物内部表面很粗糙，因为这些方法只考虑目标像素的光谱特性而忽略了周围的环境。另一方面，这些方法通常具有清晰的轮廓并保持了地物的形状因为分类时不受其他像素干扰。这个现象在 IP 和 PU 数据集上尤为明显。借助于周围像素包含的环境信息，通过空间分类方法，椒盐噪声被有效抑制并且获得了更干净和平滑的图像。和光谱组最好的结果相比，我们提出的 SaMN 在三个数据集上分别提高了 14.97%，6.28% 和 7.27%。然而，这些基于二维卷积的空间特征提取方法通常受困于过平滑现象，因为提取到的空间特征的大感受野致使小型地物被严重干扰甚至错分。例如 IP 数据上部的狭窄道路和 KSC 数据集中许多渺小孤立的“CP hammocks”在 MSCNN 和 SaMN 中消失了。根据上述讨论，为了获得相邻地物和纯净目标可以清晰辨认的分类图，有必要同时使用光谱和空间特征。受益于上述两种特性，联合组的方法获得了更高精度和更好的分类图。与 3-DCNN，SSRN 和 SSUN 相比，我们提出的 ASSMN 很好的保持了地物的边缘，这可以通过具有区分度的边角观察到，尤其是 IP 数据中部的紫色和绿色目标。在 PU 数据集，ASSMN 克服了出现在 3-DCNN 和 SSUN 的过平滑问题，它保持了许多小型地物并产生了令人满意的分类图。

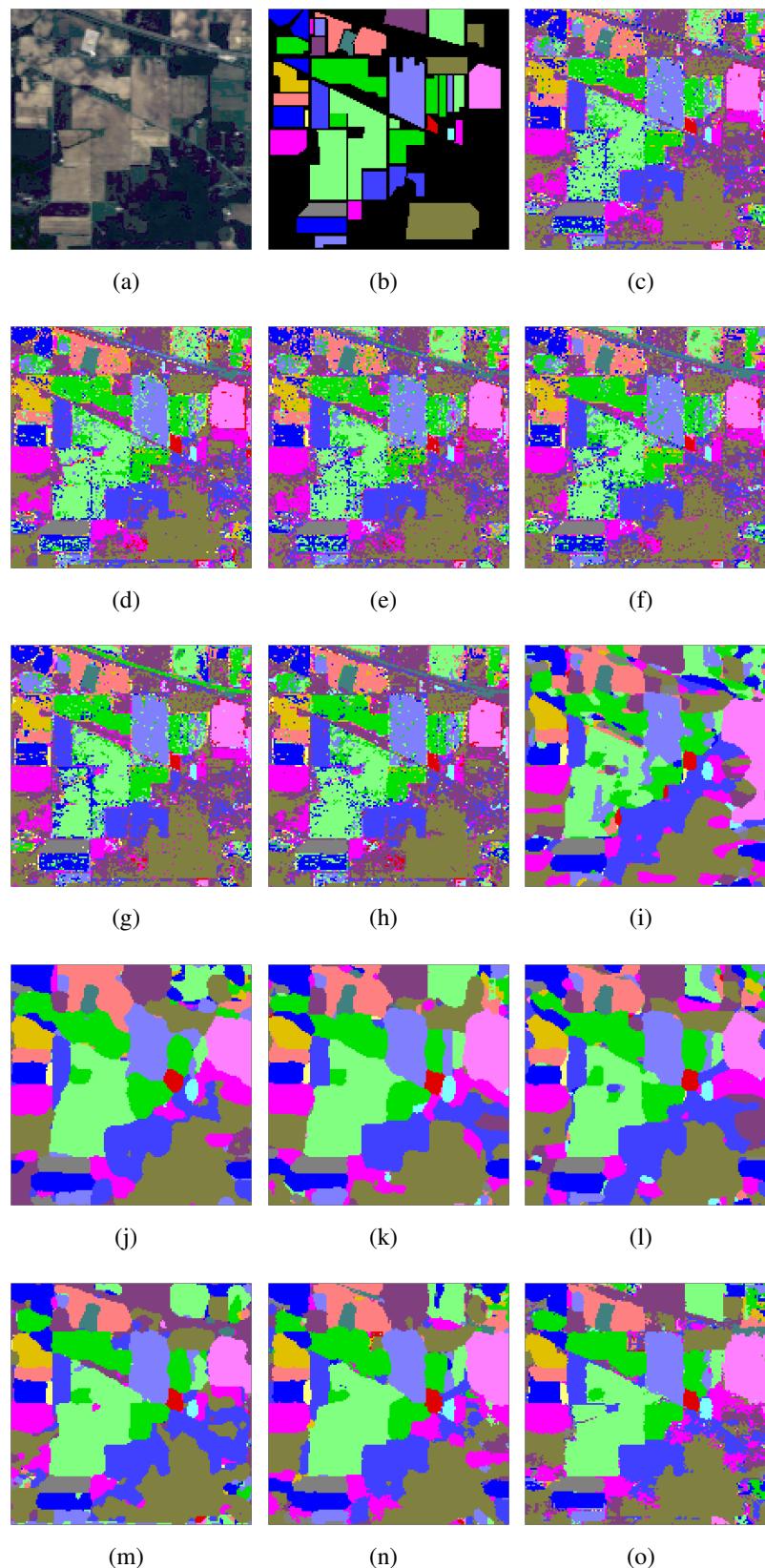


图 5.11: 不同方法在 IP 数据集的分类图。 (a) False color image. (b) Ground truth. (c) SVM. (d) 1-DCNN. (e) LSTM-strategy1. (f) LSTM-strategy2. (g) SeMN-scheme1. (h) SeMN-scheme2. (i) 2-DCNN. (j) MSCNN. (k) SaMN. (l) 3-DCNN. (m) SSRN. (n) SSUN. (o) ASSMN.

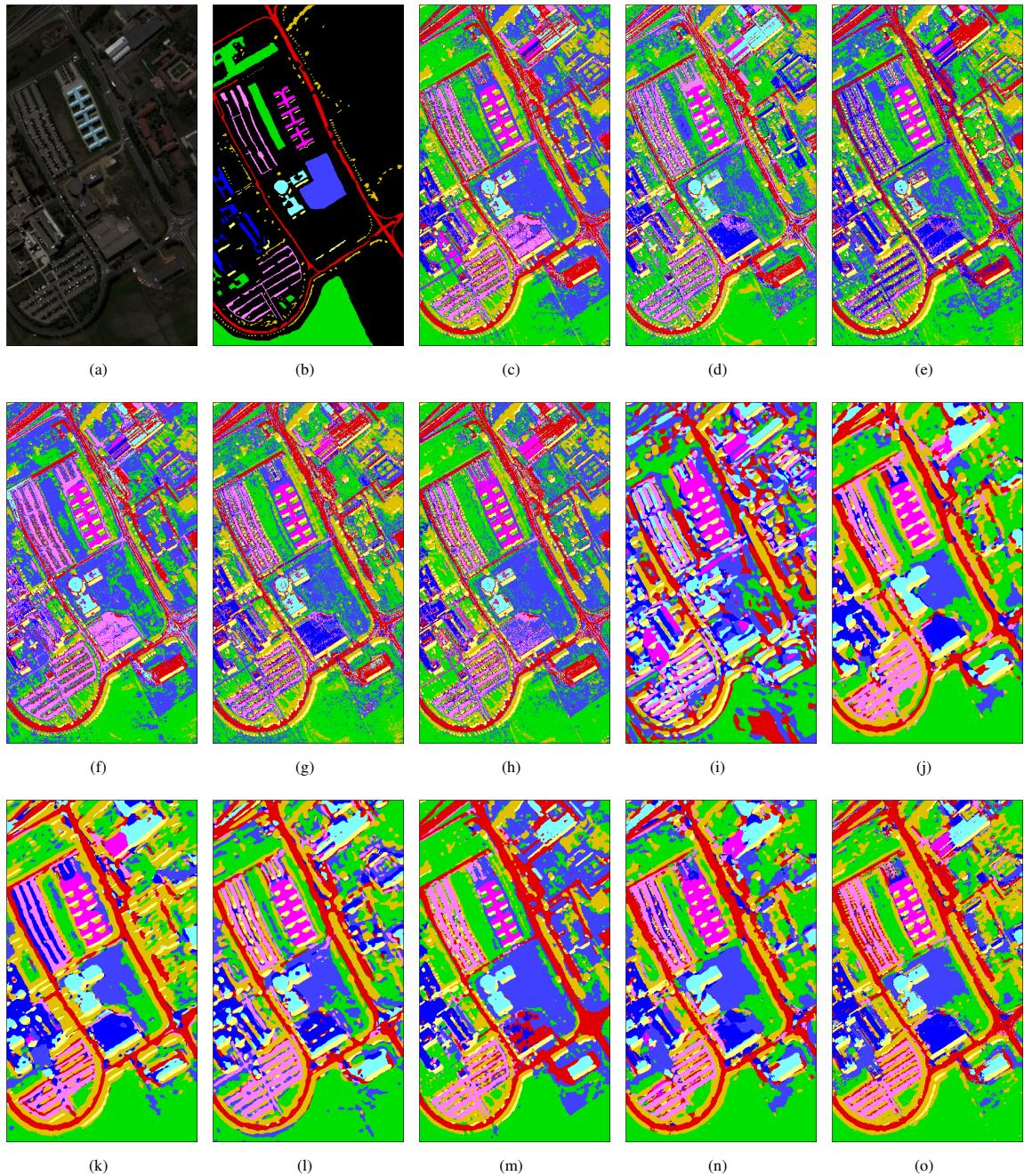


图 5.12: 不同方法在 PU 数据集的分类图。 (a) False color image. (b) Ground truth. (c) SVM. (d) 1-DCNN. (e) LSTM-strategy1. (f) LSTM-strategy2. (g) SeMN-scheme1. (h) SeMN-scheme2. (i) 2-DCNN. (j) MSCNN. (k) SaMN. (l) 3-DCNN. (m) SSRN. (n) SSUN. (o) ASSMN.

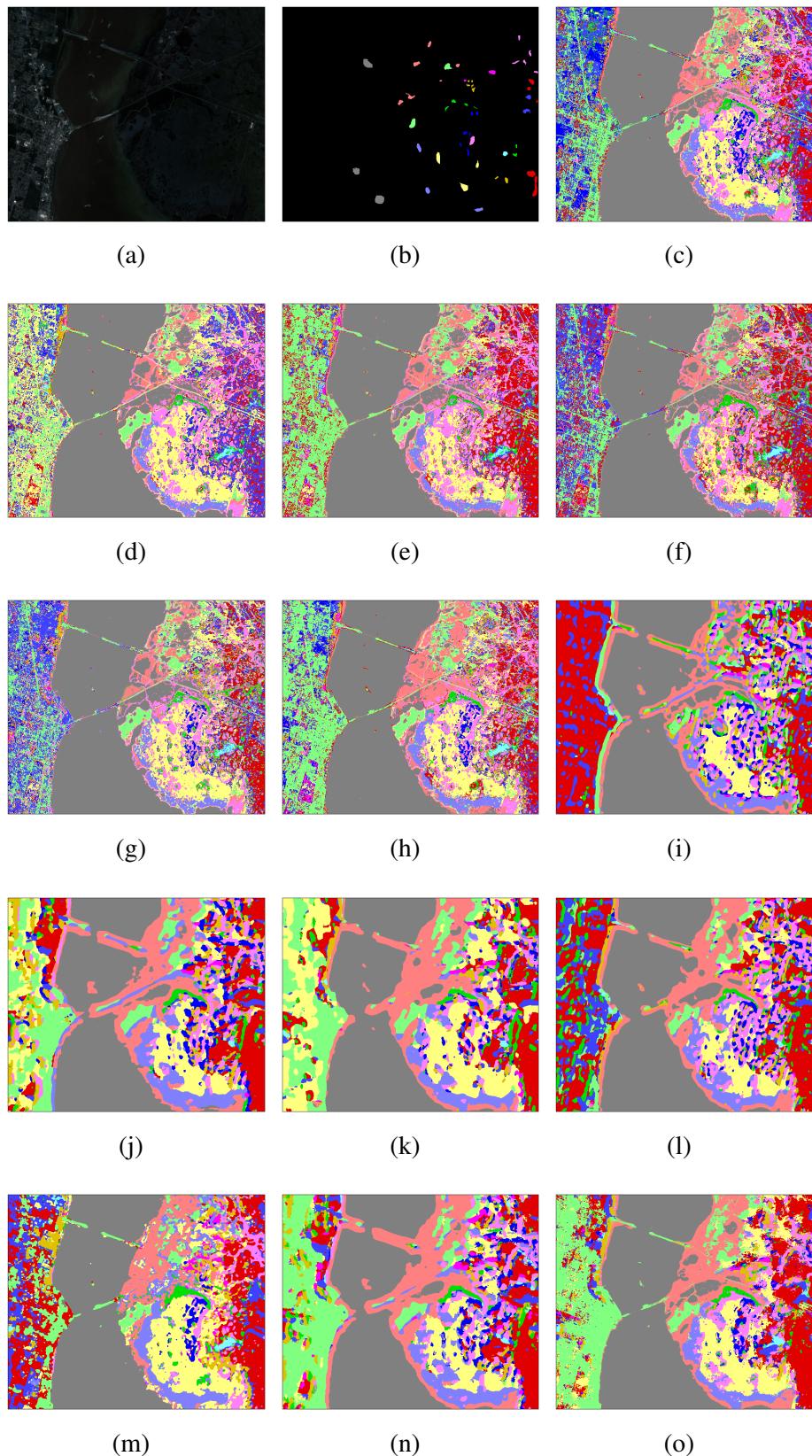


图 5.13: 不同方法在 KSC 数据集的分类图。 (a) False color image. (b) Ground truth. (c) SVM. (d) 1-DCNN. (e) LSTM-strategy1. (f) LSTM-strategy2. (g) SeMN-scheme1. (h) SeMN-scheme2. (i) 2-DCNN. (j) MSCNN. (k) SaMN. (l) 3-DCNN. (m) SSRN. (n) SSUN. (o) ASSMN.

## 5.4 章末总结

本章介绍了一种端到端用来提取多尺度上下文信息进行高光谱图像分类的自适应分类网络 ASSMN, 它包含光谱特征提取和空间特性提取两个子网络。与先前获得单尺度光谱信息的方法相比, 借助于波段分组策略的帮助, 在 SeMN 中开发出基于波长或感受野的两种多尺度方案。在 SaMN 中, 扩张卷积和激活层用来替代最大池化层。这种安排同时增大了感受野并保持了特征图的分辨率。在已有多尺度结构的基础上, 凭借 ConvLSTM 获得了空间上下文信息, 并且多维交替分组策略使得不同时间步的输入具有更充分的交流。另一种设计, 输入序列设置中的级联式平衡了计算效率和同时在行列方向的上下文信息感知。在 SeMN 和 SaMN 获得特征后, 一种名为“分数加权的”自适应特征融合模块被提出, 它考虑了光谱和空间部分不同的重要性。

在 IP, PU 和 KSC 数据集上的实验显示我们的方法受益于光谱和空间特征的双重特性, 相比于其他先进方法, 获得了更高精度和在地物之间以及地物内部具有区分度边缘以及连续表面的分类图。另外, 本章也证明了处理 ConvLSTM 中过拟合问题时正则化的有效性。

## 第 6 章 模型对比及分析

### 6.1 模型性能比较

为了更加完善的评价所提出模型的性能，本章在三四五章分别挑选出三个表现最好且最具有代表性的网络 ORI-MSHCN、SS-HPN-LSTM 以及 ASSMN，并在三个前述实验中未曾用过的高光谱数据集：Washington DC Mall (WDC), Salinas (SA) 及 Pavia Centre (PC) 上进行实验，相关实验设置均与前文相同。在分析实验结果之前，首先对这三个数据集予以介绍：

- 1) WDC: 该数据集由高光谱数字影像收集 (Hyperspectral Digital Imagery Collection Experiment, HYDICE) 传感器于华盛顿特区广场获得，包含  $307 \times 1280$  个像素。该数据集在  $0.4\text{-}2.5\mu\text{m}$  波长范围内具有 220 个光谱波段，并且在移除噪声波段后，剩余 191 个通道。该数据集共有 7 个地物类别。
- 2) SA: 该场景由 224 波段的 AVIRIS 传感器在加利福尼亚州的 Salinas Valley 获得，并且它具有 3.7m 的高空间分辨率。整个区域大小为  $512 \times 217$ 。该数据集在摒弃了 20 个水吸收波段后剩余 204 个波段用于分类。该数据集包含植被，裸土，葡萄园等 16 个类别。
- 3) PC: 该场景由 ROSIS 传感器在意大利北部的帕维亚市中心拍摄。它具有 102 个光谱波段，大小为  $1096 \times 715$ 。该数据集几何分辨率为 1.3m，共有 9 个类别。

表6.1-6.3分别展示了 ORI-MSHCN, SS-HPN-LSTM 和 ASSMN 在这三种数据集上的定量实验结果。表中结果显示，相比于 ORI-MSHCN 和 SS-HPN-LSTM，ASSMN 取得了更高的分类精度，尤其是在 SA 数据集上。在 WDC 和 PC 数据集上，三者差距不大。特别需要注意的是，虽然在 PC 数据集上 ORI-MSHCN 比 ASSMN 的总体精度表现略高，但是受制于 Trees 类的精度，平均精度却不如 ASSMN，并且 ASSMN 精度高的类较 ORI-MSHCN 要多，表明 ASSMN 的性能更稳定，因而 ASSMN 的表现要更好。以上实验结果说明，与单纯进行空间多尺度特征提取的 ORI-MSHCN 和利用 LSTM 进行逐方向的空间局部上下文信息探索的 SS-HPN-LSTM 相比，能够同时

表 6.1: 三种网络在 WDC 数据集上的分类结果

类别	Street	Grass	Water	Path	Trees	Shadow	Roofs	OA(%)	AA(%)	$\kappa$ (%)
颜色										
训练样本	100	100	100	100	100	100	100			
测试样本	3199	2975	2782	917	1927	993	5711			
ORI-MSHCN	99.39	99.97	99.18	99.54	99.94	98.18	99.07	99.35 ±0.18	99.33 ±0.27	99.20 ±0.23
SS-HPN-LSTM	98.96	99.96	99.58	99.62	99.89	99.04	99.64	99.56 ±0.09	99.53 ±0.17	99.46 ±0.11
ASSMN	98.74	99.89	99.93	99.54	99.85	99.54	99.73	99.61 ±0.11	99.60 ±0.13	99.52 ±0.14

提取并采用“分数加权融合”模块结合空谱多尺度上下文特征且空间上下文信息是利用 ConvLSTM 并采用“交替分组策略”和“级联序列设置”在更大范围内挖掘得到的 ASSMN 结合并进一步拓展了 ORI-MSHCN 和 SS-HPN-LSTM 各自的优点，因而在这三者中取得了最有力的分类结果。另外还需注意到的是，正是因为 ORI-MSHCN 和 SS-HPN-LSTM 侧重点不同，因而无法简单评判两者表现优劣。ORI-MSHCN 在 WDC 和 PC 数据集上较 SS-HPN-LSTM 表现更优，而在 SA 数据集，由于 Grapes-untrained 类别的糟糕表现，其综合性能远不如 SS-HPN-LSTM。

图6.1-6.3展示了分别展示了 ORI-MSHCN, SS-HPN-LSTM 和 ASSMN 在这三种数据集上的分类图。可以看到，因为在测试集中取得的高精度的保证，这三种网络均得到了相当高质量的分类图。尽管如此，依然能从一些关键区域进一步比较这三种方法的优劣：相比于 ORI-MSHCN 和 ASSMN，SS-HPN-LSTM 分类图较为平滑，这使得一些小型地物消失，其中对比较为明显的是 WDC 图像左上部分横穿水域连接左右两岸的桥，以及 PC 数据集中夹在细长道路绿色植被间的 Bitumen 类别，这是因为 ORI-MSHCN 采用了小补丁。尽管 ASSMN 采取的补丁比 SS-HPN-LSTM 更大，这会带来更大的过平滑风险，不过由于 ASSMN 能够提取出更高质量的多尺度光谱特征，而这些特征能够很好的保持地物边缘，因而相比于 SS-HPN-LSTM，反而取得了更精细的分类图。当然，大尺度补丁还是会不可避免的带来过平滑现象，例如在 PC 数据集上，与 ORI-MSHCN 和 SS-HPN-LSTM 相比，ASSMN 未能将图像中部的网格状 Self-

表 6.2: 三种网络在 SA 数据集上的分类结果

类别	Brocoli-green-weeds-1	Brocoli-green-weeds-2	Fallow	Fallow-rough-plow	Fallow-smooth	Stubble	Celery	Grapes-untrained	Soil-vinyard-develop	Corn-senesced-green-weeds
颜色										
训练样本	100	100	100	100	100	100	100	100	100	100
测试样本	1009	2726	1876	1294	2578	3859	3479	11171	5203	2278
ORI-MSHCN	100.0	99.97	100.0	99.82	99.26	100.0	99.96	79.26	99.99	98.64
SS-HPN-LSTM	99.99	100.0	99.95	99.78	99.75	100.0	100.0	92.55	100.0	99.65
ASSMN	99.98	99.98	99.85	99.93	99.50	100.0	99.95	95.03	99.88	99.61

类型	Lettuce-romaine-4wk	Lettuce-romaine-5wk	Lettuce-romaine-6wk	Lettuce-romaine-7wk	Vinyard-untrained	Vinyard-vertical-trellis	OA(%)	AA(%)	$\kappa$ (%)
颜色									
训练样本	100	100	100	100	100	100			
测试样本	968	1827	816	970	6268	1707			
ORI-MSHCN	100.0	100.0	99.99	99.87	96.41	99.44	94.95	98.29	94.38
SS-HPN-LSTM	100.0	99.99	99.96	99.98	97.14	99.89	$\pm 1.69$	$\pm 0.47$	$\pm 1.86$
ASSMN	99.90	100.0	99.96	99.94	96.14	99.49	97.98	99.29	97.75
							$\pm 0.44$	$\pm 0.13$	$\pm 0.49$
							98.32	99.32	98.13
							$\pm 0.52$	$\pm 0.23$	$\pm 0.58$

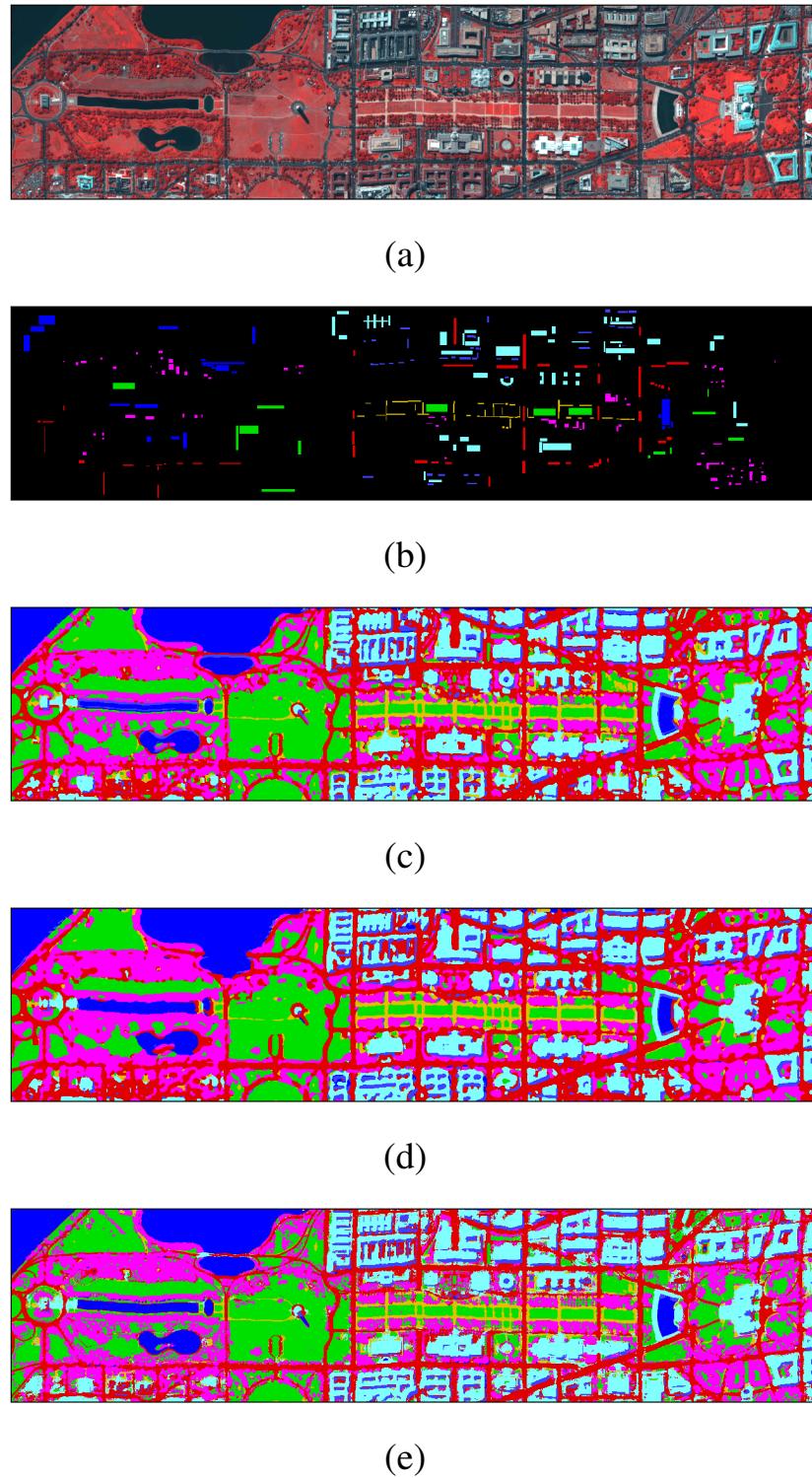


图 6.1: 三种网络在 WDC 数据集上的分类图。 (a) False color image. (b) Ground truth. (c) ORI-MSHCN. (d) SS-HPN-LSTM. (e) ASSMN.

表 6.3: 三种网络在 PC 数据集上的分类结果

类别	Water	Trees	Asphalt	Self-Blocking Bricks	Bitumen	Tiles	Shadows	Meadows	Bare Soil	OA(%)	AA(%)	$\kappa$ (%)
颜色												
训练样本	100	100	100	100	100	100	100	100	100			
测试样本	65871	7498	2990	2585	6484	9148	7187	42726	2763			
ORI-MSHCN	100.0	95.85	97.51	99.64	99.39	98.96	97.92	99.80	99.80	99.48	98.76	99.25
SS-HPN-LSTM	99.53	96.71	95.05	98.87	97.67	98.55	94.99	99.31	99.45	98.85	97.79	98.38
ASSMN	99.84	97.05	98.40	99.91	99.71	98.91	98.49	99.59	99.95	99.47	99.09	99.25
										$\pm 0.10$	$\pm 0.20$	$\pm 0.14$
										$\pm 1.21$	$\pm 1.59$	$\pm 1.70$
										$\pm 0.11$	$\pm 0.10$	$\pm 0.16$

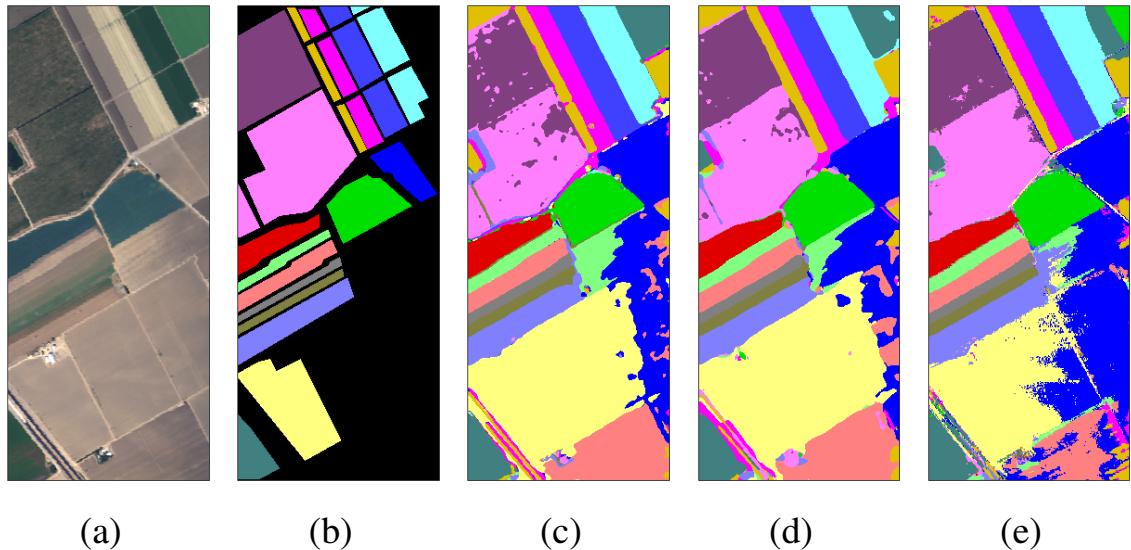


图 6.2: 三种网络在 SA 数据集上的分类图。(a) False color image. (b) Ground truth. (c) ORI-MSHCN. (d) SS-HPN-LSTM. (e) ASSMN.

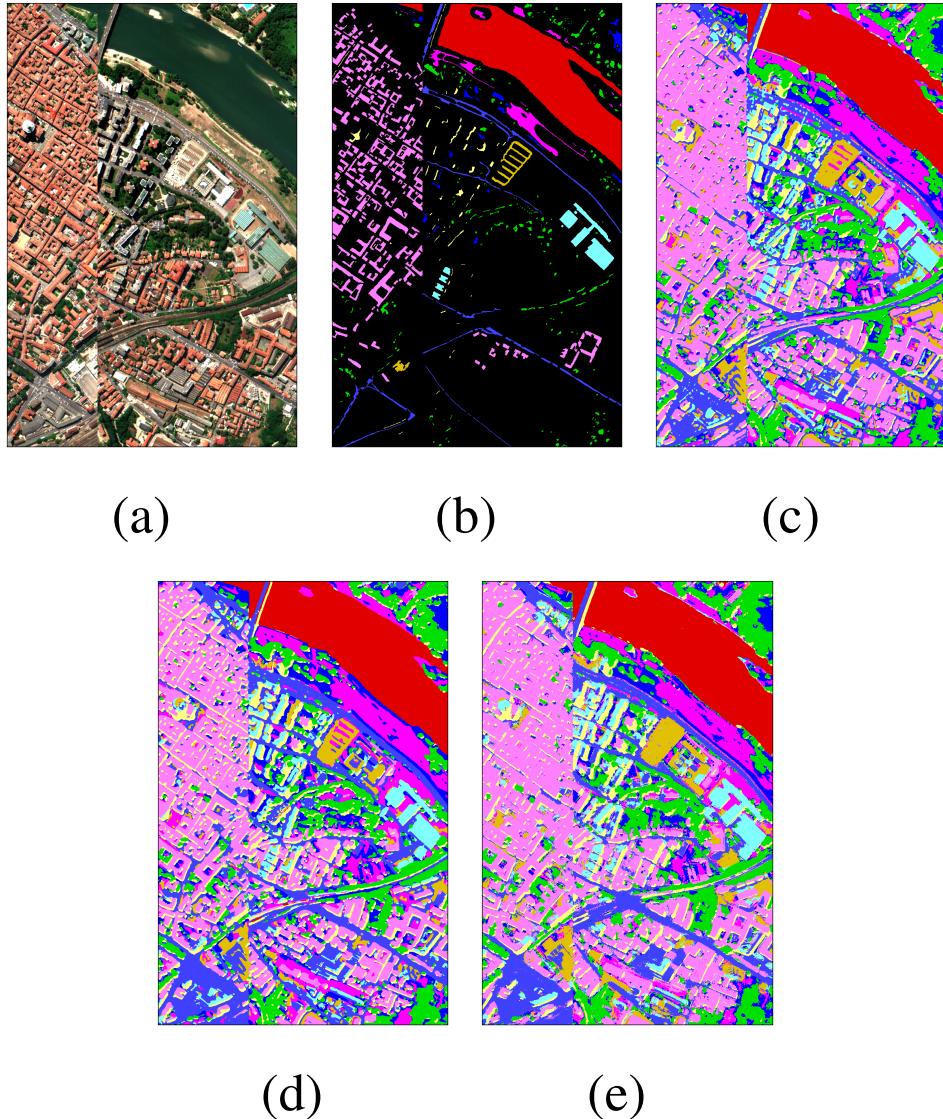


图 6.3: 三种网络在 PC 数据集上的分类图。 (a) False color image. (b) Ground truth. (c) ORI-MSHCN. (d) SS-HPN-LSTM. (e) ASSMN.

Blocking Bricks 与 Bitumen 类别很好的区分开, 这是因为利用大尺寸补丁的网络会被这种形状地物的地面真值所误导。与 ORI-MSHCN 相比, ASSMN 也能够取得更精细的分类图, 例如 WDC 图像中右侧中间部分呈环形的植被的复杂轮廓在被 ASSMN 的分类图中很好的刻画出来, 而 ORI-MSHCN 就未能达到这一目标。除了良好的边缘保持作用, ASSMN 同样能够兼顾地物内部, 使得同种地物内部连续一致且平滑, 这得益于提取出的有效空间上下文信息, 这一现象在 SA 数据集上体现的尤为明显: 与 ORI-MSHCN 和 SS-HPN-LSTM 相比, ASSMN 具有更少的错分, 且 Grapes-untrained 类别更平滑并且更接近于原始地物真值。当然, 与其他两种方法相比, SA 数据集中地

物的边界也被 ASSMN 有效的识别，因而在三者中精度最高。

## 6.2 模型复杂度比较

表 6.4: 三种网络的模型复杂度

模型复杂度	ORI-MSHCN	SS-HPN-LSTM	ASSMN
#Params(M)	33.91	14.55	2.38
FLOPs(M)	331.41	143.11	1066.10

最后，本文通过模型参数量 (#Params) 和每秒浮点数运算次数 (floating-point operations per second, FLOPs) 对 ORI-MSHCN, SS-HPN-LSTM 及 ASSMN 的模型复杂度进行了比较，其结果见表6.4。可以看到，卷积核核数较多的 ORI-MSHCN 和 SS-HPN-LSTM 参数量较大。由于 SS-HPN-LSTM 中序列模型的输入主要为  $1 \times 1$  和  $3 \times 1$  的区域，因而 SS-HPN-LSTM 的参数也主要集中于提取非关系型信息的卷积操作中。相应的，ASSMN 具有最少的参数量，这是因为其每个卷积仅有 32 个通道。不过，与 ORI-MSHCN 及 SS-HPN-LSTM 相比，ASSMN 拥有最大的输入补丁和多时间步运算的 ConvLSTM，从而具有最高的计算复杂度。事实上，ASSMN 的计算量主要集中于 ConvLSTM，因为 ConvLSTM 中的参数不断的参与重复计算。例如，在我们的实验中，ConvLSTM 中的一个卷积操作便需要 10 MFLOPs。

## 6.3 章末总结

本章主要对从前文挑选出的三个最有代表性的模型 ORI-MSHCN, SS-HPN-LSTM 及 ASSMN 进行了进一步的比较和分析：首先是模型性能比较，为此，本章特意使用三个先前实验中未曾采用的三个数据集 WDC, SA 和 PC。接着，本章又通过参数量和计算量两个维度来衡量并分析以上三个模型的复杂度。实验结果表明：相比于单纯进行空间多尺度特征提取的 ORI-MSHCN 和利用 LSTM 进行逐方向的空间局部上下文信息探索的 SS-HPN-LSTM，ASSMN 结合并进一步拓展了 ORI-MSHCN 和 SS-HPN-LSTM 各自的优点，因而在三者中取得了最有力的分类结果。ASSMN 的优势主要归功于提取到的多尺度光谱特征和利用 ConvLSTM 挖掘到的大尺度空间上下文

信息，前者有效的保持了地物边缘，后者则保证了地物内部连续且平滑。此外，尽管 ASSMN 参数量在三者中最小，然而，由于独有的“交替分组策略”及“级联序列设置”，ConvLSTM 中的参数得到了更充分的运用，并不断参与重复计算，因而计算复杂度大大增加。

## 第 7 章 结论

本文主要介绍了三项分别基于多尺度特征和空谱上下文信息提取进行高光谱图像端到端分类的研究工作。

首先本文介绍了一种基于 CNN 的高光谱图像多尺度特征提取模块，其旨在补偿前向信息传播时细节信息的丢失同时保留语义信息。多尺度来源于特征上的不同感受野，以及卷积操作更加多样化了提取的特征。在此模块基础上，本文进一步提出了一个端到端的空间信息提取网络 MSHCN 以进行高光谱分类。

接下来，本文介绍了一个可以端到端获取多种特征用于高光谱图像空谱联合分类的网络 SS-HPN。除了不同尺寸的补丁被用来获得多尺度信息外，本文主要设计了一个新颖的基于 RNN 的模块，其能够在空间层面感受上下文信息。

在前面两项研究的基础上，本文介绍了一种端到端用来提取多尺度上下文信息进行高光谱图像分类的自适应分类网络 ASSMN，它包含光谱和空间特性提取两个部分。与先前获得单尺度光谱信息的方法相比，借助于波段分组策略，在光谱部分中开发出两种分别基于波长和感受野的多尺度特征提取方案。在空间部分，采用扩张卷积和激活层替代最大池化层。这种安排同时增大了感受野并保持了特征图的分辨率。在已有尺度结构的基础上，凭借 ConvLSTM 获得了空间上下文信息，并且提出了使得不同时间步的输入交流更充分的多维交替分组策略。此外，还设计出了能够平衡计算效率和可以同时在行列方向进行上下文信息感知的级联式输入序列。在获得光谱和空间特征后，提出了一种考虑了两部分特征不同重要性的自适应特征融合模块。

实验结果表明，无论是信息补偿模块带来的多尺度特征还是空间上下文感知模块提供的关系型信息，都有助于进一步提高分类精度，这显示了高光谱图像分类中信息多元化的作用。此外，相比于单纯进行空间多尺度特征提取的 ORI-MSHCN 和利用 LSTM 进行逐方向空间局部上下文信息探索的 SS-HPN-LSTM，ASSMN 结合并进一步拓展了 ORI-MSHCN 和 SS-HPN-LSTM 各自的优点，因而在这三者中取得了最有力的分类结果，其能够获得更高精度和在地物之间和内部同时具有区分度边缘以及连续表面的分类图，因而对土地覆盖识别、城市规划、环境监测等领域具有重要意义。

## 参考文献

- [1] Zhang X, Sun Y, Shang K, et al. Crop Classification Based on Feature Band Set Construction and Object-Oriented Approach Using Hyperspectral Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(9): 4117–4128.
- [2] Yang X, Yu Y. Estimating Soil Salinity Under Various Moisture Conditions: An Experimental Study[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(5): 2525–2533.
- [3] Lu X, Zhang W, Li X. A Hybrid Sparsity and Distance-Based Discrimination Detector for Hyperspectral Images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(3): 1704–1717.
- [4] Tu B, Wang J, Kang X, et al. KNN-Based Representation of Superpixels for Hyperspectral Image Classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(11): 4032–4047.
- [5] Foody G M, Mathur A. A relative evaluation of multiclass image classification by support vector machines[J]. IEEE Transactions on Geoscience and Remote Sensing, 2004, 42(6): 1335–1343.
- [6] Chein-I Chang, Qian Du, Tzu-Lung Sun, 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] Licciardi G, Marpu P R, Chanussot J, et al. Linear Versus Nonlinear PCA for the Classification of Hyperspectral Data Based on the Extended Morphological Profiles[J]. IEEE Geoscience and Remote Sensing Letters, 2012, 9(3): 447–451.
- [8] Villa A, Benediktsson J A, Chanussot J, et al. Hyperspectral Image Classification With Independent Component Discriminant Analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2011, 49(12): 4865–4876.
- [9] Benediktsson J A, Palmason J A, Sveinsson J R. Classification of hyperspectral data from urban areas based on extended morphological profiles[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(3): 480–491.
- [10] Fauvel M, Benediktsson J A, Chanussot J, et al. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles[J]. IEEE Transactions on Geoscience

- and Remote Sensing, 2008, 46(11) : 3804–3814.
- [11] Yu H, Gao L, Liao W, et al. Multiscale Superpixel-Level Subspace-Based Support Vector Machines for Hyperspectral Image Classification[J]. IEEE Geoscience and Remote Sensing Letters, 2017, 14(11) : 2142–2146.
- [12] Cui B, Xie X, Ma X, et al. Superpixel-Based Extended Random Walker for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(6) : 3233–3243.
- [13] Fang L, Li S, Kang X, et al. Spectral–Spatial Classification of Hyperspectral Images With a Superpixel-Based Discriminative Sparse Model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(8) : 4186–4201.
- [14] Zou J, Li W, Du Q. Sparse Representation-Based Nearest Neighbor Classifiers for Hyperspectral Imagery[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(12) : 2418–2422.
- [15] Fang L, Wang C, Li S, et al. Hyperspectral Image Classification via Multiple-Feature-Based Adaptive Sparse Representation[J]. IEEE Transactions on Instrumentation and Measurement, 2017, 66(7) : 1646–1657.
- [16] Li W, Du Q. Gabor-Filtering-Based Nearest Regularized Subspace for Hyperspectral Image Classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(4) : 1012–1022.
- [17] Kang X, Li S, Benediktsson J A. Spectral–Spatial Hyperspectral Image Classification With Edge-Preserving Filtering[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(5) : 2666–2677.
- [18] Guo Y, Cao H, Bai J, et al. High Efficient Deep Feature Extraction and Classification of Spectral-Spatial Hyperspectral Image Using Cross Domain Convolutional Neural Networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(1) : 345–356.
- [19] Xia J, Bombrun L, Adali T, et al. Spectral–Spatial Classification of Hyperspectral Images Using ICA and Edge-Preserving Filter via an Ensemble Strategy[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8) : 4971–4982.
- [20] Jackson Q, Landgrebe D A. Adaptive Bayesian contextual classification based on Markov random fields[J]. IEEE Transactions on Geoscience and Remote Sensing, 2002, 40(11) : 2454–2463.
- [21] Tarabalka Y, Fauvel M, Chanussot J, et al. SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images[J]. IEEE Geoscience and Remote Sensing Letters,

- 2010, 7(4) : 736–740.
- [22] Li J, Bioucas-Dias J M, Plaza A. Spectral – Spatial Classification of Hyperspectral Data Using Loopy Belief Propagation and Active Learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(2) : 844–856.
- [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] Zhang L, Zhang L, Du B. Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art[J]. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(2) : 22–40.
- [25] Chen Y, Lin Z, Zhao X, et al. Deep Learning-Based Classification of Hyperspectral Data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(6) : 2094–2107.
- [26] Li T, Zhang J, Zhang Y. Classification of hyperspectral image based on deep belief networks[C] // Proceedings of the IEEE International Conference on Image Processing (ICIP). 2014 : 5132–5136.
- [27] HU W, HUANG Y, WEI L, et al. Deep convolutional neural networks for hyperspectral image classification[J]. Journal of Sensors, 2015, 2015.
- [28] Li W, Wu G, Zhang F, et al. Hyperspectral Image Classification Using Deep Pixel-Pair Features[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(2) : 844–853.
- [29] Zhao W, Du S. Spectral-Spatial Feature Extraction for Hyperspectral Image Classification: A Dimension Reduction and Deep Learning Approach[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8) : 4544–4554.
- [30] XU Y, DU B, ZHANG F, et al. Hyperspectral image classification via a random patches network[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 142 : 344–357.
- [31] Chen Y, Jiang H, Li C, et al. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(10) : 6232–6251.
- [32] Zhong Z, Li J, Luo Z, et al. Spectral – Spatial Residual Network for Hyperspectral Image Classification: A 3-D Deep Learning Framework[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(2) : 847–858.
- [33] Mao J, Xu W, Yang Y, et al. Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN)[J]. arXiv e-prints, 2014 : arXiv:1412.6632.
- [34] Cho K, van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations us-

- ing RNN Encoder-Decoder for Statistical Machine Translation[J]. arXiv e-prints, 2014: arXiv:1406.1078.
- [35] Sundermeyer M, Ney H, Schlüter R. From Feedforward to Recurrent LSTM Neural Networks for Language Modeling[J]. IEEE/ACM Transactions on Audio Speech and Language Processing, 2015, 23(3) : 517–529.
- [36] Mou L, Ghamisi P, Zhu X X. Deep Recurrent Neural Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(7) : 3639–3655.
- [37] FENG Z, HANG R, LIU Q, et al. Hyperspectral Image Classification Using Spectral-Spatial LSTMs[C] // Proceedings of the CCF Chinese Conference on Computer Vision (CCCV). 2017.
- [38] ZHOU F, HANG R, LIU Q, et al. Integrating Convolutional Neural Network and Gated Recurrent Unit for Hyperspectral Image Spectral-Spatial Classification[C] // Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV). 2018.
- [39] XU Y, ZHANG L, DU B, et al. Spectral-Spatial Unified Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(10) : 5893–5909.
- [40] Hang R, Liu Q, Hong D, et al. Cascaded Recurrent Neural Networks for Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019 : 1–11.
- [41] Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions[J]. arXiv e-prints, 2014: arXiv:1409.4842.
- [42] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision[J]. arXiv e-prints, 2015 : arXiv:1512.00567.
- [43] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning[J]. arXiv e-prints, 2016 : arXiv:1602.07261.
- [44] Chen L-C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J]. arXiv e-prints, 2017: arXiv:1706.05587.
- [45] Chen L-C, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[J]. arXiv e-prints, 2018 : arXiv:1802.02611.
- [46] Niu Z, Liu W, Zhao J, et al. DeepLab-Based Spatial Feature Extraction for Hyperspectral Image Classification[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(2) : 251–255.
- [47] Zhang X, Sun Y, Jiang K, et al. Spatial Sequential Recurrent Neural Network for Hyperspectral Image Classification[J]. IEEE Journal of Selected Topics in Applied Earth Observations

- and Remote Sensing, 2018, 11(11) : 4141–4155.
- [48] SHI X, CHEN Z, HAO W, et al. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
- [49] Hu W-S, Li H-C, Pan L, et al. Feature Extraction and Classification Based on Spatial-Spectral ConvLSTM Neural Network for Hyperspectral Images[J]. arXiv e-prints, 2019: arXiv:1905.03577.
- [50] LIU Q, FENG Z, HANG R, et al. Bidirectional-Convolutional LSTM Based Spectral-Spatial Feature Learning for Hyperspectral Image Classification[J]. Remote Sensing, 2017, 9(12) : 1330–.
- [51] HAO W, PRASAD S. Convolutional Recurrent Neural Networks for Hyperspectral Data Classification[J]. Remote Sensing, 2017, 9(3) : 298.
- [52] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. Cambridge, MA : MIT Press, 2016.
- [53] LECUN Y, BENGIO Y, HINTON G. Deep learning[J/OL]. Nature, 2015, 521(7553) : 436–444. <https://doi.org/10.1038/nature14539>.
- [54] Lipton Z C, Berkowitz J, Elkan C. A Critical Review of Recurrent Neural Networks for Sequence Learning[J]. arXiv e-prints, 2015 : arXiv:1506.00019.
- [55] GRAVES A. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8) : 1735–1780.
- [56] LIN M, CHEN Q, YAN S. Network In Network[J]. Computer Science, 2013.
- [57] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C] // Proceedings of the International Conference on Machine Learning (ICML). 2015.
- [58] 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) : págs. 212–223.
- [59] Li J, Zhao X, Li Y, et al. Classification of Hyperspectral Imagery Using a New Fully Convolutional Neural Network[J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(2) : 292–296.
- [60] Paoletti M E, Haut J M, Fernandez-Beltran R, et al. Deep Pyramidal Residual Networks for Spectral-Spatial Hyperspectral Image Classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2) : 740–754.
- [61] 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 (CVPR). 2016 : 770–778.
- [62] Werbos P J. Backpropagation through time: what it does and how to do it[J]. Proceedings of the IEEE, 1990, 78(10) : 1550–1560.
- [63] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[J]. arXiv e-prints, 2015 : arXiv:1511.07122.
- [64] Chen L-C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. arXiv e-prints, 2014 : arXiv:1412.7062.
- [65] Chen L, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4) : 834–848.
- [66] Xu Y, Du B, Zhang L. Beyond the Patchwise Classification: Spectral-Spatial Fully Convolutional Networks for Hyperpsectral Image Classificaiton[J]. IEEE Transactions on Big Data, 2019 : 1–1.
- [67] Shirish Keskar N, Socher R. Improving Generalization Performance by Switching from Adam to SGD[J]. arXiv e-prints, 2017 : arXiv:1712.07628.
- [68] CHANG C-C, LIN C-J. LIBSVM: A Library for Support Vector Machines[J/OL]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3). <https://doi.org/10.1145/1961189.1961199>.

## 攻读测绘工程专业硕士学位期间学术成果

- D. Wang, B. Du, L. Zhang, and S. Chu, “Hyperspectral Image Classification Based on Multi-scale Information Compensation,” *Remote Sensing Letters*. (*Accept*).
- D. Wang, B. Du and L. Zhang. “Hyperspectral Image Classification Based on Spatial Contextual Information Perception,” *IEEE Geoscience and Remote Sensing Letters*. (*Under Review*).
- D. Wang, B. Du, L. Zhang, and Y. Xu, “Adaptive Spectral-Spatial Multi-Scale Contextual Feature Extraction for Hyperspectral Image Classification,” *IEEE Transactions on Geoscience and Remote Sensing*. (*Accept*)

## 致 谢

或许是前四年夜夜笙歌、寻欢作乐，而今自觉乏味平淡、身心俱变，只好埋头苦干，于万千文献中找寻一丝慰藉。不过既来之则安之，过程尚且不论，结果稳中向好，至少能做出些东西聊以自慰，所以流程该走还是要走，该致谢还是要致谢。

首先要感谢三位导师：感谢史老师提供名额在我求学道路上提供帮助；感谢杜老师的培养，特别要感谢杜老师给予我最大限度的自由，让我的科研之路畅通无阻，能够全神贯注钻研学术，而不至于像旁人一样常常被一些琐事绊住脚步；感谢张老师容忍我刚开学新生讲课上的“飞扬跋扈”，还要感谢您教我如何待人接物并亲自推荐让我不至于走投无路。

接下来要感谢各位师兄师姐，感谢永浩师兄一字一句帮我修改初稿，提供帮助，尤其是各种科研招数。那段每晚在操场散布交谈、剖析问题本质、探究网络玄妙的时光，我至今难忘；感谢李雪师姐给予我生活和学习上无微不至的关怀和呵护；感谢静文师姐前期给予我的科研帮助，虽然后来常常是我文化反哺，也让我明白之前学的有些东西并非一无是处。感谢世祯师姐每天早晨签到能够多等十分钟容忍我的作息无度。

然后是同门、同学和小伙伴们：感谢朱博士前期对我的各种指引和帮助；感谢一楚和承熙不仅给予我帮助，还让我有地方能够吹牛倾诉；感谢褚佬领我走上比赛之路，愿你在他乡异国大展宏图；感谢许神，真正的大佬，难得又很幸运的在这种环境下遇见志趣相投的你，愿你在西安年入百万，两年是不够的，咱们有缘再见！感谢季佬经常帮我拿耳机，更感谢你的热水壶让我有机会品尝本科四年未曾享受的泡面！感谢张总、叶总、薛总、杨总还有遗憾未至的勇士，魔都之旅与你们相聚一扫之前的阴霾，彻底改变了我的心态；还有感谢一直以来的小伙伴们，不管是召唤师峡谷还是绝地海岛的求生之路，有你们相伴才不会孤独！

感谢一直以来支持、鼓励与帮助我的人，感谢你们这两年以来的陪伴。除此以外，我还要感谢食堂的饭锻炼了我的胃，让我知道原来之前二十年吃的饭全是玉盘珍馐山珍海味。感谢实验室每晚 11 点赶人、宿管阿姨每晚 11 点拿铁链子锁门强行打断我一天的科研节奏提醒我要注意休息劳逸结合。更要感谢这里的酷暑与严寒、蛛丝与蚁患、宿舍窗外的垃圾堆与噪声污染，磨炼了我的意志，让我在北三环一晚两百块钱总共几平米的经济间也能睡得无比舒坦。

最后，要感谢我的父母，感谢你们一直都在，哪怕在千里之外。

## 武汉大学学位论文使用授权协议书

本学位论文作者愿意遵守武汉大学关于保存、使用学位论文的管理办法及规定，即：学校有权保存学位论文的印刷本和电子版，并提供文献检索与阅览服务；学校可以采用影印、缩印、数字化或其它复制手段保存论文；在以教学与科研服务为前提下，学校可以在校园网内公布部分及全部内容。

- 1、 在本论文提交当年，同意在校园网内以及中国高等教育文献保障系统（CALIS）高校学位论文系统提供查询及前十六页浏览服务。
- 2、 在本论文提交□当年/□一年/□两年/□三年以后，同意在校园网内允许读者在线浏览并下载全文，学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。（保密论文解密后遵守此规定）

论文作者（签名）： 齐迪

学 号： 2018286190128

学 院： 测绘遥感信息工程国家重点实验室

日期： 2020 年 5 月 22 日

