# Deep Image Prior for Disentangling Mixed Pixels

by

Yuan Fang

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2022

**Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

| | |
|---|---|
| External Examiner: | Paolo Gamba<br>Professor,<br>Dept. of Electrical, Computer and Biomedical Engineering,<br>University of Pavia |
| Supervisor(s): | David A. Clausi<br>Professor,<br>Dept. of Systems Design Engineering, University of Waterloo |
| | Linlin Xu<br>Research Assistant Professor,<br>Dept. of Systems Design Engineering, University of Waterloo |
| Internal Member: | Paul Fieguth<br>Professor,<br>Dept. of Systems Design Engineering, University of Waterloo |
| Internal Member: | Andrea Scott<br>Associate Professor,<br>Dept. of Systems Design Engineering, University of Waterloo |
| Internal-External Member: | Jonathan Li<br>Professor,<br>Dept. of Geography and Environmental Management,<br>University of Waterloo |

## Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

## Statement of Contributions

Chapters 3, 4 and 5 contain material from three multi-author papers, where the manuscripts are minorly changed for format consistence. As the lead author of these three papers, I was responsible for conceptualizing the studies, writing codes, carrying out main experiments, drafting and submitting the manuscripts. My co-author, Yuxian Wang, contributed to the comparison experiments by running compared methods for paper 1 and 2. Rongming Zhuo helped to collect the datasets for paper 1. Yuhao Chen and Wei Zhou provided feedback on the loss function design for paper 3. My supervisors Prof. David A. Clausi and Prof. Linlin Xu, provided guidance throughout the process and feedback on draft manuscripts. The references for the three papers are provided below:

1 Fang, Yuan, Yuxian Wang, Linlin Xu, Rongming Zhuo, Alexander Wong, and David A. Clausi."BCUN: Bayesian Fully Convolutional Neural Network for Hyperspectral Spectral Unmixing." IEEE Transactions on Geoscience and Remote Sensing 60 (2022): 1-14.

2 Fang, Yuan, Yuxian Wang, Linlin Xu, Alexander Wong, and David A. Clausi. " Unsupervised Bayesian Subpixel Mapping Neural Network for Hyperspectral images", submitted to IEEE Transactions on Geoscience and Remote Sensing.

3 Fang, Yuan, Linlin Xu, Yuhao Chen, Wei Zhou, Alexander Wong, and David Clausi. "A Bayesian Deep Image Prior Downscaling Approach for High-resolution Soil Moisture Estimation." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2022).

I additionally co-authored several papers in related areas of remote sensing image processing and environmental monitoring during my PhD program, including

- Chen, Yujia, Linlin Xu, Yuan Fang, Junhuan Peng, Wenfu Yang, Alexander Wong, and David A. Clausi. "Unsupervised bayesian subpixel mapping of hyperspectral imagery based on band-weighted discrete spectral mixture model and Markov random field." IEEE Geoscience and Remote Sensing Letters 18, no. 1 (2020): 162-166.

- Zhuo, Rongming, Yuan Fang, Linlin Xu, Yujia Chen, Yuxian Wang, and Junhuan Peng. "A novel spectral-temporal Bayesian unmixing algorithm with spatial prior for Sentinel-2 time series." Remote Sensing Letters 13, no. 5 (2022): 522-532.

- Chen, Yuhao, Alexander Wong, Yuan Fang, Yifan Wu, and Linlin Xu."Deep Residual Transform for Multi-scale Image Decomposition." Journal of Computational Vision and Imaging Systems 6, no. 1 (2020): 1-5.

- Fang, Yuan, Linlin Xu, Alexander Wong, and David A. Clausi. "Multi-Temporal Landsat-8 Images for Retrieval and Broad Scale Mapping of Soil Copper Concentration Using Empirical Models." Remote Sensing 14, no. 10 (2022): 2311.

- Fang, Yuan, Zhongzheng Hu, Linlin Xu, Alexander Wong, and David A. Clausi. "Estimation Of Iron Concentration In Soil Of A Mining Area From Uav-Based Hyperspectral Imagery." In 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1-5. IEEE, 2019.

- Fang, Yuan, Linlin Xu, Yuhao Chen, and Alexander Wong. "Domain Adaptive Shake-shake Residual Network for Corn Disease Recognition.", AI for Earth Sciences Workshop at NeurIPS 2020.

## Abstract

A mixed pixel in remotely sensed images measures the reflectance and emission from multiple target types (e.g., tree, grass and building) from a certain area. Mixed pixels exist commonly in spaceborne hyper-/multi-spectral images due to sensor limitations, causing the signature ambiguity problem and impeding high-resolution remote sensing mapping. Disentangling mixed pixels into the underlying constituent components is a challenging ill-posed inverse problem, which requires efficient modeling of spatial prior information and other application-dependent prior knowledge concerning the mixed pixel generation process.

The recent deep image prior (DIP) approach and other application-dependent prior information are integrated into a Bayesian framework in the research, which allows comprehensive usage of different prior knowledge. The research improves mixed pixel disentangling using the Bayesian DIP in three key applications: spectral unmixing (SU), subpixel mapping (SPM) and soil moisture product downscaling (SMD). The main contributions are summarized as follows.

First, to improve the decomposition of mixed pixels into pure material spectra (i.e., endmembers) and their constituting fractions (i.e., abundances) in SU, a designed deep fully convolutional neural network (DCNN) and a new spectral mixture model (SMM) with heterogeneous noise are integrated into a Bayesian framework that is efficiently solved by a new iterative optimization algorithm.

Second, to improve the decomposition of mixed pixels into class labels of subpixels in SPM, a dedicated DCNN architecture and a new discrete SMM are integrated into the Bayesian framework to allow the use of both spatial prior and the forward model.

Third, to improve the decomposition of mixed pixels into soil moisture concentrations of subpixels in SMD, a new DIP architecture and a forward degradation model are integrated into the Bayesian framework that is solved by the stochastic gradient descent approach.

These new Bayesian approaches improve the state-of-the-art in their respective applications (i.e., SU, SPM and SMD), which can be potentially utilized for solving other ill-posed inverse problems where simultaneously modeling of the spatial prior and other prior knowledge is needed.

**Acknowledgements**

I would like to express my deepest gratitude to my supervisors, Prof. David A. Clausi and Prof. Linlin Xu, for their invaluable support and mentorship. I would also like to thank my committee members – Profs. Paolo Gamba, Paul Fieguth, Andrea Scott, and Jonathan Li for their feedback. A special thanks to Prof. Scott and Prof. Clausi for involving me in a research project with Agriculture and Agri-Food Canada on soil moisture retrieval.

I would like to thank all my paper co-authors for their contributions. A special thanks to Prof. Alex Wong for involving me in the corn disease detection project with Nutrien.

I would like to thank the kind and smart people of the Remote Sensing Group, Dr. Xinwei Chen, Mingzhe Jiang, Saeid Taleghanidoozdoozan, Yifan Wu, Peter Lee, and Javier Noa for their support. I also want to thank the incredible friends of the Vision and Image Processing Lab for their help and concern during my PhD program, including Dr. Mehrnaz Fani, Dr. Yuhao Chen, Xuefeng Ni, Armina Soleymani and so on.

I would like to thank my friends, Ze Yang, Jun Jiang, Shiyu Yan and Wen Ding, for their company and encouragement during my PhD program.

Last but not least, many thanks to my parents for their unconditional support and encouragement.

# Table of Contents

# List of Figures

# List of Tables

# List of Acronyms

| | |
|---|---|
| SU | Spectral unmixing |
| SPM | Subpixel mapping |
| SMD | Soil moisture downscaling |
| DIP | Deep image prior. |
| BDIP | Bayesian deep image prior |
| SM | Soil moisture |
| DS | Downscaling |
| IFOV | Instantaneous field-of-view. |
| HR | High resolution |
| LR | Low resolution |
| HSI | Hyperspectral image |
| MSI | Multispectral image |
| RS | Remote sensing |
| ML | Machine learning |
| DL | Deep learning |
| DCNN | Deep convolutional neural network |
| FCNN | Fully convolutional neural network |
| MRF | Markov random field |
| CRF | Conditional random field |
| NNLS | Nonnegative least squares |
| SNR | Signal-to-noise ratio |
| SAD | Spectral angle distance |
| SID | Spectral information divergence |
| AAD | Abundance angle distance |
| AID | Abundance information divergence |
| SDA | Spatial dependence assumption |
| EM | Expectation-maximization |
| MAP | Maximum a posteriori |
| OA | Overall accuracy |
| GT | Ground truth |
| MODIS | Moderate Resolution Imaging Spectrometer |
| AVIRIS | Airborne Visible/Infrared Imaging Spectrometer |
| SMAP | Soil Moisture Active/Passive |
| NDVI | Normalized difference vegetation index |
| LST | land surface temperature |
| ISMN | International soil moisture network |
| R | Correlation coefficient |
| MSE | Mean square error |
| RMSE | Root mean square error |
| nrRMSE | Normalized root mean square error |
| ubRMSE | Unbiased root mean square error |

# Chapter 1

# Introduction

## 1.1 Background

### 1.1.1 Mixed pixels in remote sensing images

A mixed pixel in remotely sensed images measures the reflectance and emission from multiple target types (e.g., tree, grass and building) from a certain area, which is illustrated in Figure 1.1. Mixed pixels exist commonly in spaceborne hyper-/multi-spectral images due to the limitation of the hardware, e.g., the limitation in storage (i.e., the data volume collected by the sensor) and bandwidth transmission (i.e., the incoming radiation Ene to the sensor), as well as the trade-off effect between spatial resolution and spectral resolution [1, 2]. To achieve high spectral resolution with many spectral channels, the spatial resolution usually has to be compromised, leading to a large instantaneous field-of-view (IFOV) of remotely sensed images. Moreover, high spectral resolution also tends to reduce the signal-to-noise ratio (SNR), because the signal magnitude is reduced due to very narrow bandwidth in high spectral resolution images. Because of these sensor limitations, spaceborne hyperspectral images (HSIs) and multispectral images (MSIs) tend to have low spatial resolution with many mixed pixels. For example, Moderate Resolution Imaging Spectrometer (MODIS) has the band-dependent resolution ranging from 250m to 1km, and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral imagery has resolution of 20m with a high flight height. The existence of mixed pixels in HSIs and MSIs causes the signature ambiguity problem, and impedes high-resolution (HR) RS mapping.

### 1.1.2 Mixed pixel disentangling problem

Remote sensing (RS) images provide essential information for a wide range of Earth system applications [3, 4, 5, 6]. However, the signature ambiguity issue caused by mixed

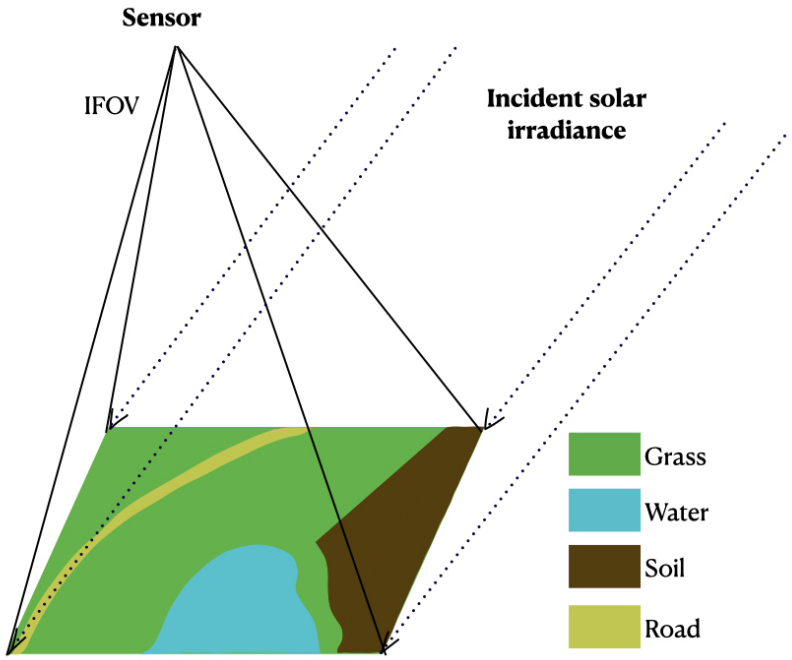



Figure 1.1: Illustration of a mixed pixel. A mixed pixel in remotely sensed images measures the reflectance and emission from multiple target types than just one type. Mixed pixels exist commonly in spaceborne hyper-/multi-spectral images due to the limitation of hardware and bandwidth transmission, as well as the trade-off effect between spatial resolution and spectral resolution. To achieve high spectral resolution with many spectral channels, the spatial resolution usually has to be compromised, leading to remotely sensed images that has large IFOV.

pixels makes it difficult to obtain precise information for RS applications [7, 8, 9]. Therefore, disentangling mixed pixels, i.e., separating the underlying constituent components within mixed pixels is critical for RS image processing to support key RS applications. In this thesis, three different mixed pixel disentangling tasks are performed, namely spectral unmixing (SU), subpixel mapping (SPM) and soil moisture product downscaling (SMD). These three tasks are illustrated in Figure 1.2.

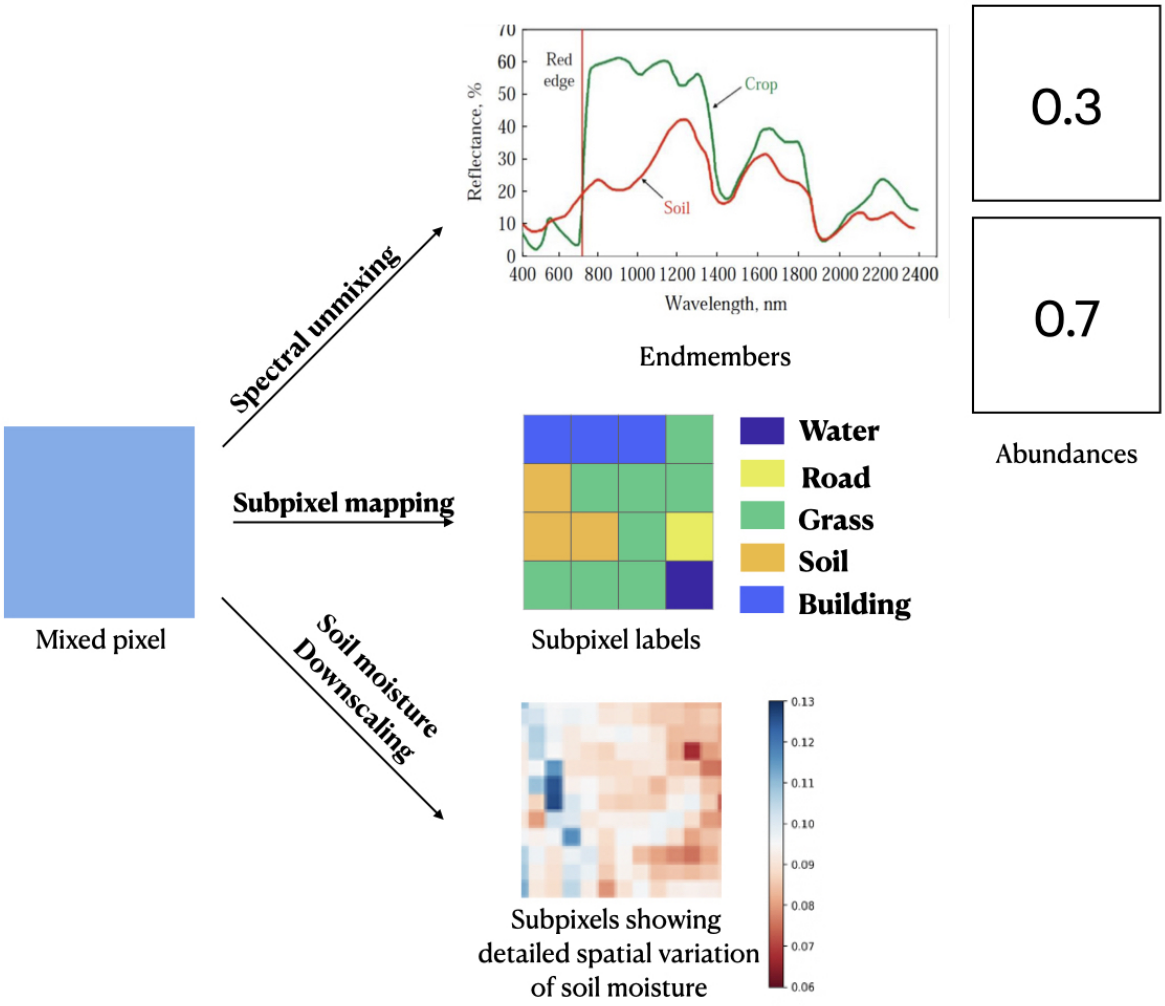



Figure 1.2: Illustration of SU, SPM and SMD. SU decomposes a mixed pixel in HSIs into pure material spectra (i.e., endmembers) and their constituent fractions (i.e., abundances). The abundance values indicate the fractions of each endmember. SPM decomposes a mixed pixel into subpixels and estimates class labels of these subpixels. SMD decomposes a mixed pixel from a soil moisture map into soil moisture content of subpixels whichs show detailed soil moisture variation patterns. Different colors of subpixels represent different soil moisture content.

SU aims to decompose a mixed pixel in HSIs or MSIs into pure material spectra (i.e., endmembers) and their constituent fractions (i.e., abundances). SPM aims to decompose a mixed pixel into class labels of subpixels. SMD aims to decompose a mixed pixel in a

soil moisture map into soil moisture content of subpixels that show detailed soil moisture variation patterns.

The generation process of the observed images with mixed pixels from the underlying quantities can be expressed as

$$\boldsymbol{X} = \Phi(\boldsymbol{Y}) + \boldsymbol{N}, \tag{1.1}$$

where $\Phi(\cdot)$ is the forward function, $\boldsymbol{X}$ is the observed matrix representing the low-resolution (LR) RS image with mixed pixels, $\boldsymbol{Y}$ is HR underlying quantities, $\boldsymbol{N}$ is the noise matrix. Different applications have different underlying quantities $\boldsymbol{Y}$.

- For SU, $\boldsymbol{Y}$ is endmembers and abundances.
- For SPM, $\boldsymbol{Y}$ is class labels of subpixels in the image.
- For SMD, $\boldsymbol{Y}$ is soil moisture content of subpixels in the image.

Disentangling mixed pixels $\boldsymbol{X}$ into $\boldsymbol{Y}$ is essentially an ill-posed inverse problem, and the prior knowledge regarding the data generation process is required to constrain the estimation. The most important prior knowledge for restoring the underlying HR quantities in mixed pixels is the spatial prior among pixels in the image $\boldsymbol{Y}$. Other priors include the forward model $\Phi(\cdot)$, noise $\boldsymbol{N}$ distribution, other priors of the underlying quantities $\boldsymbol{Y}$. It is challenging to design tailored approaches for different priors, and to simultaneously model all these priors in a coherent way for enhanced applications.

## 1.2 Motivation and related works

### 1.2.1 Motivation

Based on Eq. 1.1, this thesis aims to explore advanced machine learning (ML) approaches that can simultaneously address different priors in the forward model and achieve efficient mixed pixel disentangling in several key applications, i.e., SU, SPM and SMD. This research is motivated by the following challenges.

1. Achieving an **accurate and efficient spatial prior** for $\boldsymbol{Y}$ is difficult. Methods for modelling the spatial correlation in image processing mainly include graphical models (e.g., Markov random field (MRF) [10] and conditional random field (CRF) [11]) and non-local approaches (e.g., non-local means [12] and non-local network [13]). However, an MRF-based prior can only model the local spatial correlation and tend to smooth high-frequency spatial features. CRF-based or non-local methods are computationally intensive by calculating the similarity matrix [14]. The recently proposed deep image prior (DIP) captures the image spatial correlation by encoding hierarchical self-similarities [15] with the structure of a deep convolutional image generator network. Unlike traditional

approaches, DIP is achieved more flexibly and efficiently by training a fully convolutional neural network (FCNN) leveraging graphics processing units (GPUs). DIP is demonstrated to be able to restore high-quality images from low-quality images without requiring a large training dataset [16, 17]. Recent publications show the success of DIP on image restoration [18, 19, 20], e.g., super-resolution [21], image inpainting [17] and denoising [22, 23]. Although DIP shows potential for modelling the spatial correlation of regular RGB images, it has not been systematically explored in a Bayesian framework for disentangling mixed pixels in RS images.

2. Efficiently **integrating a forward model $\Phi(\cdot)$ with deep learning (DL) networks** is essential but difficult. In traditional image restoration approaches, the forward model is built into error functions to regulate the model optimization process, e.g., sparse coding [24], image decomposition [25], low-rank approximation [26] and Gaussian mixture model [27]. Although DL approaches can efficiently learn features for complex RS image processing tasks, most DL-based methods are heavily data-driven and neglect the guidance of the knowledge and priors in physical forward models. Although the Bayesian framework offers a coherent way for integrating priors and knowledge, integrating forward models with DL models into a Bayesian framework has not been sufficiently studied in the context of disentangling mixed pixels for RS image inversion.

3. **Noise $\boldsymbol{N}$ modeling** is important but usually tends to be ignored or downweighted. For example, the noise level of HSIs varies dramatically over bands due to different spectral absorption properties of different spectral channels and the commonly existence of corrupted noisy bands ("junk bands"). However, most of SU methods ignore the noise heterogeneity effect, leading to undesirable preservation of noise in some bands and erasing of signal in some other bands. Inaccurate noise characterization reduces the estimation accuracy of underlying components $\boldsymbol{Y}$. Although accurate noise modelling is essential, there has not been sufficient research on integrating accurate noise model with Bayesian DIP approaches for enhanced disentangling mixed pixels.

4. **Model optimization** is important but difficult. The Bayesian method is used widely to solve inverse problems by addressing a maximum a posterior (MAP) problem using iterative optimization approaches, e.g., expectation-maximization algorithm [28, 29, 30]. Traditionally, due to various explicit prior distributions in the Bayesian model, the resulting posterior distribution has a complex form and is very difficult to be solved efficiently. The rapidly developing DL technique provides new approaches for addressing priors in inverse problems e.g., DIP [16]. Although the integration of DIP with forward models into a Bayesian framework might lead to more efficient MAP optimizations, it has not been sufficiently researched for mixed pixel disentangling in key RS applications.

5. **Task-specific algorithms** are essential for supporting key RS applications. SU, SPM and SMD are three important RS applications that all rely heavily on disentangling mixed pixels. Despite their similarities, SU, SPM and SMD have different data, forward models and underlying quantities, and thereby it is required to develop task-specific algorithms by adapting the general Bayesian DIP approach to different task characteristics.

### 1.2.2 Related works

#### Deep image prior (DIP)

The structure of the DCNN is capable to capture image statistical information and to impose an effective prior to restore high-quality images from low-quality images without requiring a large training dataset, and this prior is call the DIP [17]. Recent publications show the effectiveness of DIP for image restoration [18, 19, 20], e.g., hyperspectral image unmixing [31], super-resolution [21], image inpainting [17] and denoising [22, 23].

Traditionally, these problems can be generally expressed as an energy equation [16],

$$\boldsymbol{Y} = \min_{\boldsymbol{Y}} \mathrm{Ene}\left(\boldsymbol{Y}; \boldsymbol{X}\right) + R(\boldsymbol{Y}) \tag{1.2}$$

where $\mathrm{Ene}\left(\boldsymbol{Y}; \boldsymbol{X}\right)$ is the task-dependant term, which will be discussed for different applications in Chapter 2, 3 and 4. $R(\boldsymbol{Y})$ is the regularizer. DIP, instead of modelling the regularizer $R(\boldsymbol{Y})$ with an explicit form, models $R(\boldsymbol{Y})$ implicitly with the DCNN architecture by seeking the solution in the network parameter domain, as follows,

$$\boldsymbol{\beta} = \underset{\theta}{\mathrm{argmin}} \mathrm{Ene}\left(f_{\boldsymbol{\beta}}(\boldsymbol{Z}); \boldsymbol{X}\right), \quad \boldsymbol{Y} = f_{\boldsymbol{\beta}}(\boldsymbol{Z}) \tag{1.3}$$

where $f$ is the forward propagation of the DCNN. $\boldsymbol{Z}$ is the input random noise and $\boldsymbol{\beta}$ is the set of model parameters. The energy function is reformulated into MAP problems in a Bayesian DIP framework in Chapter 2, 3 and 4.

#### DCNNs with forward models for inverse problems

The strategy of combining a DCNN with a forward model to solve an inverse problem has been explored since 2018 for image restoration and 3D reconstruction [16, 32, 33]. However, the related studies are very few, especially in the context of RS image inversion. Pure data-driven DL approaches suffer from requiring large training dataset, large amount of network parameters and training time [34]. On the contrary, the integration of the DCNN with forward models makes the network lighter [34] and learn from both low-quality data (e.g., low-resolution images) and the forward model in an unsupervised way.

Compared with the traditional patch-based convolutional neural network (CNN), the FCNN can better capture the large-scale spatial correlation effect from images [35]. In this thesis, the U-Net architecture [36] with skip connections [37] is adopted for three applications given that the classic network succeeded in numerous image processing tasks [38, 39]. The importance of architecture design will be demonstrated in Chapter 4. The investigation of more architectures for future work will be discussed in Chapter 5.

## 1.3 Objectives

Based on the above challenges, the research aims to achieve the following three key objectives.

1. To improve the decomposition of mixed pixels into pure material spectra (i.e., endmembers) and their constituting fractions (i.e., abundances) in SU, the research aims to design an efficiently-optimized Bayesian framework that incorporates a deep fully convolutional neural network (DCNN) and a spectral mixture model (SMM) with heterogeneous noise (see Chapter 2).

2. To improve the decomposition of mixed pixels into class labels of subpixels in SPM, the research aims to design a Bayesian framework that integrates a DCNN and a dedicated forward model to allow the use of both spatial prior and physical knowledge (see Chapter 3).

3. To improve the decomposition of mixed pixels into soil moisture values of subpixels in SMD, the research aims to design a new DIP architecture and a forward degradation model to be integrated into the Bayesian framework (see Chapter 4).

The Bayesian approaches proposed by this thesis not only improve the state-of-the-arts in their respective applications (i.e., SU, SPM and SMD), but also provide new solutions to other ill-posed inverse problems where simultaneous modelling of the spatial prior and other prior knowledge is needed.

## 1.4 Thesis structure

Three key chapters represent the developments of SU (Chapter 2), SPM (Chapter 3), and SMD (Chapter 4). These developments share a similar auto-encoder framework that incorporates different designed forward models with DCNN on different datasets. A flowchart in Figure 1.3 shows how they are implemented in the Bayesian DIP framework. Chapter 2 proposes BCUN, a Bayesian fully convolutional neural network for hyperspectral SU. This research involves imposing constraints on endmembers, abundances and noise to regulate SU for HSIs. BCUN combines an FCNN and linear spectral mixture model in a Bayesian framework and optimizes the MAP problem with a designed EM iterative method. Chapter 3 proposes an SPM method which combines an FCNN and a discrete spectral mixture model to generate a finer-resolution classification map for HSIs. The resulting MAP problem is solved with an EM algorithm. Chapter 4 proposes an SMD method by integrating an FCNN with the forward model in a coherent manner for combining different sources of information, i.e., the knowledge in the forward model, the spatial correlation prior in FCNN architecture, and the RS data and products. Chapter 5 summarizes the thesis and discusses future work.

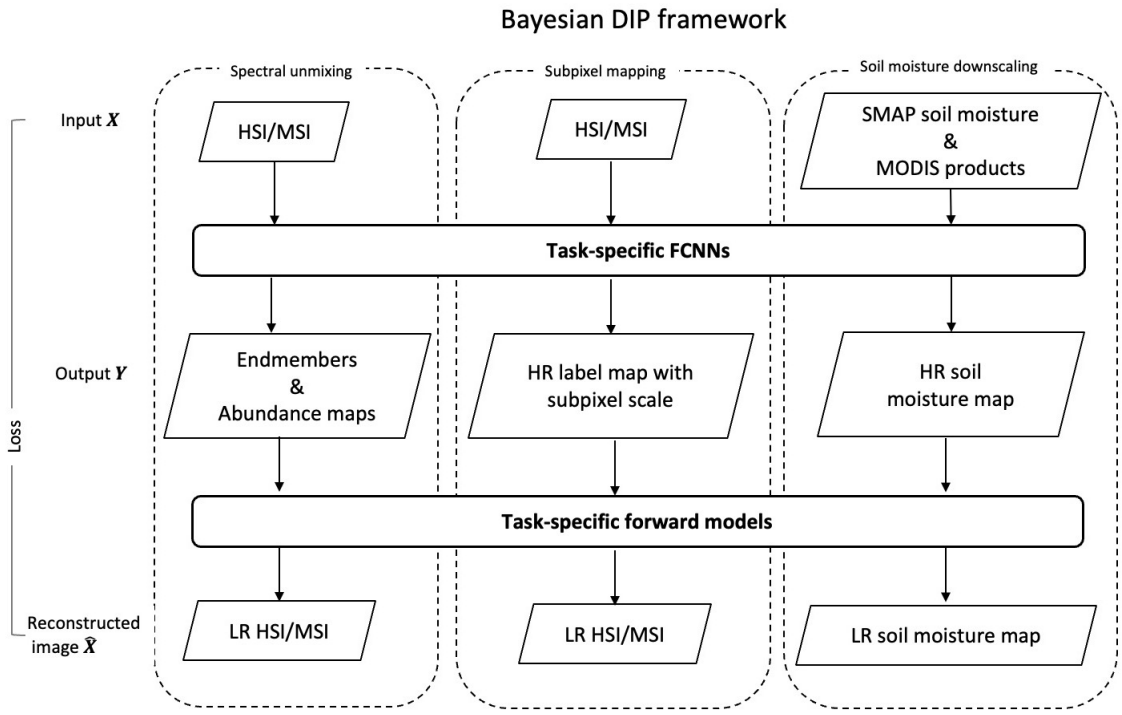



Figure 1.3: Flowchart of SU, SPM and SMD implemented in a Bayesian DIP framework.

The SU and SPM methods in Chapter 2 and 3 are tested mainly on HSIs. The SMD method in Chapter 4 is concluded on RS imagery-derived products. The validation methods for these three research are different. In SU, endmembers and abundances estimation accuracies are evaluated separately with true endmembers and abundances. The label map obtained by the SPM is evaluated compared with a true label map. The downscaled soil moisture is evaluated with both in-situ soil moisture measurements and soil moisture products.

# Chapter 2

# BCUN: Bayesian Fully Convolutional Neural Network for Hyperspectral Spectral Unmixing

Spectral unmixing (SU) plays a fundamental role in hyperspectral image (HSI) processing. Effective SU relies on the accurate and efficient characterization of the noise effect, the endmembers, and the spatial correlation effect in abundances, as well as efficient optimization techniques to estimate these effects. To address these issues, this chapter presents a Bayesian fully convolutional hyperspectral unmixing network (BCUN), with the following key characteristics. First, a fully convolutional neural network (FCNN) based deep image prior is designed for enhanced characterization and estimation of the spatial context information in abundance maps, leading to more efficient and accurate abundance modelling than the traditional non-negative least squares approaches. Second, a multivariate Gaussian distribution with anisotropic covariance matrix is designed to characterize the conditional distribution of the spectral observations, leading to a novel Mahalanobis distance-based loss for FCNN training that is better capable of addressing the noise heterogeneous effect in HSI than the Euclidean distance based mean squared error loss in traditional deep neural networks. Third, the designed conditional distribution of spectral observations also enables the incorporation of the spectral mixture model into the FCNN training process for effectively leveraging the knowledge in the forward spectral model. Fourth, the endmembers are modelled and estimated by a "purified means" approach that is capable of better characterizing endmembers. Last, the above key components are coherently integrated into a Bayesian framework, and the resulting maximum a posteriori problem is solved by a designed expectation-maximization algorithm. Experimental results on both simulated and real HSIs demonstrate that the proposed BCUN approach outperforms the other classical and state-of-the-art methods on both endmember estimation and abundance estimation.

## 2.1 Introduction

Hyperspectral imaging is a rapidly growing remote sensing technique which records the electromagnetic radiation from the earth surface via hundreds of narrow and contiguous spectral bands [40]. With rich spectral information, hyperspectral images (HSI) are critical to a wide variety of applications including ground target classification [41, 42, 29], sub-pixel mapping [43], agricultural management [3] and environmental monitoring [44, 4]. However, due to the limitation in the spatial resolution of HSI and the complexity of ground targets, spectral pixels in HSI usually contain the spectral contribution of multiple pure materials, leading to mixed pixels in HSI. Spectral unmixing (SU) aims to decompose the mixed pixels to derive both the spectral signatures of constituent components (i.e. endmembers) and their corresponding fractional proportions (i.e., abundances) from the mixed pixel in HSI [45].

SU is a challenging ill-posed inverse problem [45]. An effective SU approach relies on accurately modelling of the noise effect, the endmembers, the spatial correlation effect of abundance, as well as an effective optimization approach for estimating these effects. However, most existing SU methods ignore or fail to fully explore these key factors.

First, HSI contains not only rich spectral information but also abundant spatial information. The spatial context information in HSI is crucial for regulating and estimating endmembers and abundances in HSIs [46]. Nevertheless, most existing SU methods ignore this spatial correlation effect by treating the abundance of each pixel as independently distributed variables using nonnegative least squares (NNLS) [47], sparse unmixing by variable splitting and augmented lagrangian (Sunsal) [48] or fully connected layers [49, 50] for abundance estimation. There is still a lack of advanced modelling and estimation approaches for addressing large-scale spatial correlation effects that commonly exist in HSI. Compared with the traditional patch-based convolutional neural network (CNN), the fully convolutional neural network (FCNN) can better capture the large-scale spatial correlation effect in HSI [35]. FCNN approaches have been widely used in many tasks, e.g., semantic segmentation [51, 52], super-resolution [53] and image denoising [54]. However, they are rarely adapted to HSI for SU. Although FCNN was used to efficiently map random noise to abundances in a supervised manner [55], it was less efficient in leveraging the forward spectral mixture model for efficiently estimating both abundance and endmembers in an integrated manner. Therefore, how to integrate FCNN into SU for better modelling and estimation of spatial information in HSI is an important research issue.

Second, HSI is contaminated by noise in the data acquisition process. Due to different spectral absorption properties of different spectral channels and the commonly existence of "junk bands", the noise level of different HSI channels tend to vary dramatically, causing different noise variance in different HSI channels [56]. Since SU is very sensitive to noise, the success of SU algorithms depends on their effectiveness in accurately accounting for and resisting the noise effect [45]. However, most of the existing methods assume that different bands in HSI contain the same degree of noise by using the mean squared error (MSE) as

the reconstruction loss, leading to the undesirable preservation of noise in some bands but erasing of signal in some other bands. Consequently, the design of the model which is able to accommodate the noise variance heterogeneous effect in HSI is an important issue for SU.

Third, meaningful SU relies on effective constraints imposed on endmembers. However, existing endmember extraction methods either rely on the minimum volume simplex constraint (e.g., VCA [57] and PPI [58]) that cannot deal with highly-mixed pixels, or on prior distribution constraints in the Bayesian framework that are computationally intractable [45]. "Purified means" approach [59, 46, 60] provides simple and effective constraints for deriving endmembers that are achieved by treating the endmember as the mean vector of the "purified" pixels. Adopting this efficient constraint on endmembers boosts the accuracy and effectiveness of the estimation of endmembers and abundances.

Fourth, given the above mentioned three key factors for SU, it is critical to design a coherent framework with effective modelling capacity and an efficient optimization approach, which is capable of capturing the large-scale spatial correlation effect in abundances, the variance heterogeneity effect of noise and the proper constraint on endmembers. To solve ill-posed problems where endmembers and abundances are unknown underlying quantities, the SU process can be generalized as a maximum a posteriori (MAP) problem in a Bayesian framework and be optimized by expectation–maximization (EM) algorithm. The iterative method allows the estimated abundances and endmembers to be updated in an adaptive manner.

Therefore, in this chapter, following the linear spectral mixture model (LSMM) that is commonly used to describe the HSI generation process [61], we propose a Bayesian fully convolutional hyperspectral unmixing network (BCUN) which integrates FCNN and LSMM in a Bayesian framework for HSI unmixing.

In the proposed approach, the data likelihood is designed based on the LSMM, the conditional probability of the endmembers is derived from the "purified means" approach [59, 46, 60], and the prior of abundance is implemented by the deep image prior (DIP) [16]. The key contributions of this chapter are summarized as follows:

1. A skip-connection FCNN is designed for abundance estimation, where DIP is used to model the spatial correlation of the abundance field. Compared to NNLS or fully connected network, the FCNN is able to efficiently and accurately estimate abundances by leveraging GPUs and the large-scale spatial correlation of HSI.

2. The noise is modelled as a multivariate Gaussian distribution to account for the noise variance heterogeneity in HSI. As a result, the loss function of BCUN is designed based on the M-distance rather than MSE loss. The designed conditional distribution of spectral observations also enables the incorporation of the spectral mixture model (SMM) into the FCNN training process for effectively leveraging the knowledge in the forward spectral model.

3. The endmember is modelled and estimated by a "purified means" approach which can be seamlessly integrated into the Bayesian framework by a designed conditional distribution of the endmembers given the abundance.

4. The above key components are coherently integrated into a Bayesian framework, and the resulting maximum an MAP is solved by a designed EM algorithm.

Experiments on both simulated and real HSI demonstrate that comparing with traditional and state-of-the-art approaches, the proposed BCUN can extract endmembers and map abundances more accurately by exploring spatial correlation effect and noise heterogeneity effect in HSI. The remainder of the chapter is organized as follows. Section 2.2 formulates the unmixing problem in a Bayesian framework. The design of the network and its rationale are detailed in Section 2.3. Section 2.4 introduces the optimization scheme of BCUN. Section 2.5 conducts experiments on both simulated and real HSI.

## 2.2 Problem formulation

Assuming that there are $P$ spectral bands and $N$ pixels in a HSI, we denote the observed reflectance of the pixel at site $i$ by $\boldsymbol{x}_i$, which is $P \times 1$ vector. Then the HSI can be expressed as $\boldsymbol{X} = \{\boldsymbol{x}_i | i = 1, 2, ..., N\}$. The HSI is assumed to contain $K$ endmembers. According to the LSMM, the observed HSI $\boldsymbol{X} \in \mathbb{R}^{P \times N}$ is represented by the product of the endmember matrix $\boldsymbol{A} \in \mathbb{R}^{P \times K}$ and the abundance matrix $\boldsymbol{S} \in \mathbb{R}^{K \times N}$, plus some Gaussian noise $\boldsymbol{N} \in \mathbb{R}^{P \times N}$, as follows:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}, \tag{2.1}$$

where $\boldsymbol{S} = \{\boldsymbol{s}_i | i = 1, ..., N\}$ and $\boldsymbol{A} = \{\boldsymbol{a}_k | k = 1, ..., K\}$. $\boldsymbol{s}_i$ is a $K$-dimensional vector and $\boldsymbol{a}_k$ is a $P$-dimenstional vector. Then the pixel $\boldsymbol{x}_i$ can be formulated as a linear combination of the endmembers $\boldsymbol{A}$ weighted by their associated abundances $\boldsymbol{s}_i$ plus noise $\boldsymbol{n}_i$:

$$\boldsymbol{x}_i = \sum_{k=1}^{K} \boldsymbol{a}_k s_i^k + \boldsymbol{n}_i, \quad \sum_k s_i^k = 1, \forall s_i^k > 0. \tag{2.2}$$

Considering that current imaging systems are designed based on the assumption of additive Gaussian noise [62], and the Gaussian noise assumption simplifies the computation and the noise variance estimation [63], we assume that the noise distribution satisfy a Gaussian model as follows:

$$p(\boldsymbol{n}_i) = \frac{1}{\sqrt{(2\pi)^P |\boldsymbol{\Lambda}|}} \exp(-\frac{1}{2} \boldsymbol{n}_i{}^T \boldsymbol{\Lambda}^{-1} \boldsymbol{n}_i) \tag{2.3}$$

where $\mathbf{\Lambda}$ is commonly expressed by the most existing methods as a diagonal matrix

$$\mathbf{\Lambda} = \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \ddots & \\ & & & \sigma^2 \end{bmatrix}$$

in which $\sigma^2$ is the noise variance of the each band.

The spectral unmixing of HSI aims to infer the endmembers $\boldsymbol{A}$ and the abundance $\boldsymbol{S}$ based on the spectrum observation $\boldsymbol{X}$, which in a Bayesian framework can be achieved by maximizing the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{X})$, i.e.,

$$p(\boldsymbol{\theta}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{2.4}$$

where

$$\boldsymbol{\theta} = \{\boldsymbol{A}, \boldsymbol{S}\} \tag{2.5}$$

According to the Bayes' theorem, the posterior distribution can be rewritten as:

$$p(\boldsymbol{A}, \boldsymbol{S}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{S})p(\boldsymbol{A}|\boldsymbol{S})p(\boldsymbol{S}) \tag{2.6}$$

Given the LSMM describing the generative model of HSI in Eq. 2.1 and the posterior distribution in Eq. 2.6, several key factors for effective SU are identified as follows:

1. The accurate modelling of the abundance prior $p(\boldsymbol{S})$ is critical for regulating and estimating the abundance $\boldsymbol{S}$.
2. Properly characterizing the noise $\boldsymbol{N}$ in HSI facilitates the accurate estimation of $\boldsymbol{A}$ and $\boldsymbol{S}$.
3. Meaningful constraints on endmembers $\boldsymbol{A}$ that guide and regulate the endmember estimation process.
4. An efficient optimization scheme for solving the Bayesian inverse problem is critical.

In this chapter, $p(\boldsymbol{S})$ is achieved by Gaussian distribution where the spatial correlation effect in $\boldsymbol{S}$ is modeled by the deep image prior of FCNN, as detailed in Section 2.3.1. Under the assumption that noise $\boldsymbol{n}_i$ has heterogeneous band-dependent noise variance, the data likelihood $p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{S})$ is modeled by the anisotropic multivariate Gaussian distribution, as detailed in Section 2.3.2. $p(\boldsymbol{A}|\boldsymbol{S})$ is achieved by Gaussian distribution where the expectation of $\boldsymbol{A}$ is derived by the purified-means, as detailed in Section 2.3.3. And an efficient EM algorithm based optimization scheme is designed and implemented in Section 2.4 for solving the new Bayesian inverse problem.

## 2.3 BCUN: Bayesian fully convolutional hyperspectral unmixing network

### 2.3.1 Prior of abundances

There are three key requirements on the abundance $\boldsymbol{S}$ when designing the abundance prior $p(\boldsymbol{S})$.

1. The large-scale spatial correlation effect in abundance should be fully exploited.
2. Abundances should be subject to the non-negative and sum-to-one constraints, i.e., $\sum_k s_i^k = 1$ and $\forall s_i^k > 0$
3. Abundances prior should allow efficient optimization.

Here, the prior of S is expressed as a Gaussian distribution because (i) it is a common form of abundace distribution [64, 65, 66], and (ii) it is simple to be incorporated and solved in the final objective function.

$$p(\boldsymbol{S}) = \frac{1}{z} exp(-||\boldsymbol{S} - E(\boldsymbol{S})||^2) \tag{2.7}$$

where $E(\boldsymbol{S})$ is the expectation of $\boldsymbol{S}$, which is implemented as an FCNN. Comparing with the patch-based CNN, FCNN has a wide field of view of the input image and enables better modelling of the spatial correlation effect in HSI.

The prior spatial information of abundances $\boldsymbol{S}$ can be captured by an FCNN structure [16] which has a wide field of view of the input image and can be optimized efficiently on GPUs. Using $f(\cdot)$ to represent the FCNN forward propagation, the expected $\boldsymbol{S}$ is written as:

$$E(\boldsymbol{S}) = f(\boldsymbol{Z}, \boldsymbol{\beta}). \tag{2.8}$$

where $\boldsymbol{Z}$ is the input random noise and $\boldsymbol{\beta}$ is the set of model parameters including all convolution kernels and biases. The non-negative and the sum-to-one constraint can be achieved using "softmax" activation approach which is expressed as:

$$\text{Softmax}(I) = \frac{e^{I_i}}{\sum_j^K e^{I_j}} \tag{2.9}$$

In this work, we use a U-Net type "hourglass" architecture with skip-connection [16] to model a mapping $f(\cdot)$ from the input variable $\boldsymbol{Z}$ to abundance maps $\boldsymbol{S}$.

As shown in Figure 2.1, the FCNN network is an "hourglass" architecture with encoder and decoder parts, as well as the skip connection. The constitutional unit of the encoder

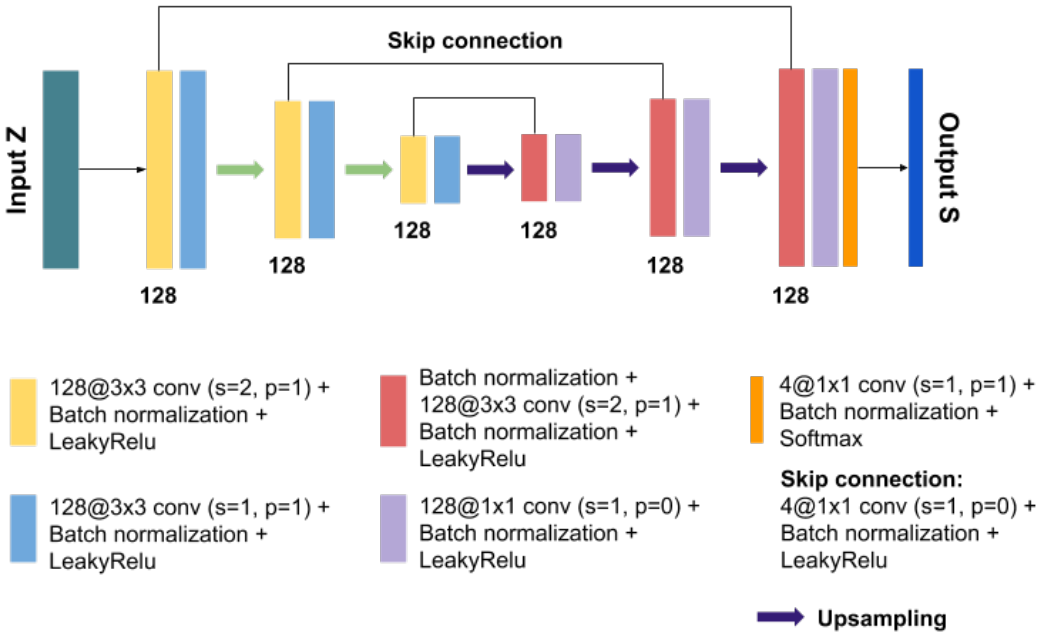



Figure 2.1: FCNN structure in BCUN framework

part involves mainly two convolution operations (the yellow and blue blocks in Figure 2.1) which can be formulated as:

$$En(I) = LR(BN(W_e^2 \circledast (LR(BN(W_e^1 \circledast I + B_e^1))) + B_e^2)) \tag{2.10}$$

where $I$ represents the input data, $LR$ represents the "leaky ReLU" activation function, and $BN$ denotes the batch normalization. $\circledast$ is the 2D convolution operation. $W$ and $B$ are convolution kernel and the bias separately.

The skip connection operation mainly contains a convolution operation where the kernel size is illustrated in Figure 2.1. The entire operation of skip connection is as follows:

$$\text{Skip}(I) = LR(BN(W_s \circledast I + B_s)) \tag{2.11}$$

The unit operation of decoder (the red and purple blocks in Figure 2.1) is expressed as:

$$De(I) = LR(BN(W_d^2 \circledast (LR(BN(W_d^1 \circledast BN(I) + B_d^1))) + B_d^2)) \tag{2.12}$$

The encoder part of FCNN decreases the size of the input via convolutions and the decoder part recovers the image size by the decoder unit expressed in Eq. 2.12 and the up-sampling operation. By connecting the corresponding parts which have the same scale of encoder and decoder, the spatial information in a certain scale is able to be well preserved. So, the resulting layer in a certain scale can be obtained by

$$\text{Scale}(I) = De(up(En(I)) \oplus \text{skip}(I)) \tag{2.13}$$

where $\oplus$ is the matrix concatenation operation. To preserve and utilize the spatial information of different image scale, the number of scale $L$ is usually set as a number more than 1. Then the main part of the FCNN is expressed as

$$\text{SCALE}_L(I) = \text{Scale}_L ... (\text{Scale}_2(\text{Scale}_1(I))) \tag{2.14}$$

In Figure 2.1, $L = 3$, the last layer is a convolution layer with the kernel size of $1 \times 1$ (the orange block in Figure 2.1). The entire operation of the output layer is as follows:

$$\text{Out}(I) = BN(W^o \circledast I + B^o) \tag{2.15}$$

By applying the "softmax" activation to the output layer, the whole FCNN is formulated by

$$f(I) = \text{Softmax}(\text{Out}(\text{SCALE}_L(I))). \tag{2.16}$$

### 2.3.2 Data likelihood

The formulation of data likelihood $p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{S})$ relies on the probabilistic distribution of noise $\boldsymbol{N}$. The noise distribution is assumed to satisfy a Gaussian model as Eq. 2.3. in which $\boldsymbol{\Lambda}$ is the covariance matrix of noise. Traditionally, most methods assume hogmogeneous noise variance across different HSI bands, and treat $\boldsymbol{\Lambda}$ as an isotropic diagonal matrix with the same diagonal elements, leading to Euclidean distance based loss. Instead, in this chapter, considering the variance heterogeneity of noise, $\boldsymbol{\Lambda}$ is expressed as follows,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_P^2 \end{bmatrix}$$

in which $\sigma_p^2$ is the noise variance of the $p$th band. This matrix allows different channels to have different noise variance, and thereby it can better accommodate the noise variance heterogeneity issue for enhanced noise characterization and spectral unmixing.

According to the LSMM in Eq. 2.2, the data likelihood is expressed as

$$\begin{aligned} & p(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{S}) \\ & = \frac{1}{\sqrt{(2\pi)^P |\boldsymbol{\Lambda}|}} exp(-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{AS})^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{X} - \boldsymbol{AS})) \end{aligned} \tag{2.17}$$

where $(\boldsymbol{X} - \boldsymbol{AS})^T \boldsymbol{\Lambda}^{-1} (\boldsymbol{X} - \boldsymbol{AS})$ is the reconstruction error which is essentially the square of M-distance $Dist_M(\boldsymbol{X}, \hat{\boldsymbol{A}}\hat{\boldsymbol{S}})$ between the original HSI $\boldsymbol{X}$ and the reconstructed HSI $\hat{\boldsymbol{A}}\hat{\boldsymbol{S}}$.

### 2.3.3 Conditional distribution of endmembers given abundance

We choose a conditional prior distribution for the endmembers with the form of a Gaussian distribution [67] encourages the small distance between the estimated endmember and its expectation,

$$p(\boldsymbol{a}_k|\boldsymbol{S}, \boldsymbol{a}_{j\neq k}) = \frac{1}{z} exp(-||\boldsymbol{a}_k - E(\boldsymbol{a}_k|\boldsymbol{S}, \boldsymbol{a}_{j\neq k})||^2) \tag{2.18}$$

Then, the joint density $p(\boldsymbol{A}|\boldsymbol{S})$ is as follows:

$$\begin{aligned} p(\boldsymbol{A}|\boldsymbol{S}) &= \prod_{j=1}^{K} p(\boldsymbol{a}_k|\boldsymbol{S}, \boldsymbol{a}_{j\neq k}) \\ &= \prod_{j=1}^{K} \frac{1}{z} exp(-||\boldsymbol{a}_k - E(\boldsymbol{a}_k|\boldsymbol{S}, \boldsymbol{a}_{j\neq k})||^2) \end{aligned} \tag{2.19}$$

The above implementation is based on a conditional independence assumption of endmember $\boldsymbol{a}_k$ given the rest of the endmembers $\boldsymbol{a}_{j\neq k}$.

## 2.4 Model Optimization

### 2.4.1 MAP estimation

The unmixing problem in Eq. 2.4 can be solved by the MAP approach, where the endmembers $\boldsymbol{A}$ and the abundance $\boldsymbol{S}$ are estimated by maximizing the posterior distribution of $\{\boldsymbol{A}, \boldsymbol{S}\}$ given the observed HSI $\boldsymbol{X}$, i.e.,

$$\{\hat{\boldsymbol{A}}, \hat{\boldsymbol{S}}\} = arg \max_{\boldsymbol{A},\boldsymbol{S}} \{p(\boldsymbol{A}, \boldsymbol{S}|\boldsymbol{X})\} \tag{2.20}$$

Maximizing $p(\boldsymbol{A}, \boldsymbol{S}|\boldsymbol{X})$ is equivalent to minimizing its negative logarithm likelihood, i.e.,

$$\{\hat{\boldsymbol{A}}, \hat{\boldsymbol{S}}\} = arg \min_{\boldsymbol{A},\boldsymbol{S}} \{-logp(\boldsymbol{A}, \boldsymbol{S}|\boldsymbol{X})\} \tag{2.21}$$

Then, the objective function can be written as

$$\begin{aligned} J_{A,S} &= arg \min_{\boldsymbol{A},\boldsymbol{S}} \{-logp(\boldsymbol{A}, \boldsymbol{S}|\boldsymbol{X})\} \\ &\propto arg \min_{\boldsymbol{A},\boldsymbol{S}} \{-logp(\boldsymbol{X}|\boldsymbol{A}, \boldsymbol{S}) - logp(\boldsymbol{S}) - logp(\boldsymbol{A}|\boldsymbol{S})\} \end{aligned} \tag{2.22}$$

Considering Eq. 2.7, 2.17 and 2.19, the objective function in Eq. 2.22 can be reformulated as

$$\begin{aligned} J_{A,S} =& arg\min_{\boldsymbol{A},\boldsymbol{S}}\{((\boldsymbol{X}-\boldsymbol{A}\boldsymbol{S}))^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{A}\boldsymbol{S})) \\ &+||\boldsymbol{S}-E(\boldsymbol{S})||^2+\alpha\sum_{k=1}^{K}||\boldsymbol{a}_k-E(\boldsymbol{a_k}|\boldsymbol{S},\boldsymbol{a}_{j\neq k})||^2\} \end{aligned} \tag{2.23}$$

In E-step of EM algorithm, $\boldsymbol{S}$ is estimated by $E(\boldsymbol{S})$. So, we replace $\boldsymbol{S}$ with $E(\boldsymbol{S})$ to simplify the objective function. As a result, the objective function can be reformulated as [16]

$$\begin{aligned} J_{A,S} =& arg\min_{\boldsymbol{A},\boldsymbol{S}}\{((\boldsymbol{X}-\boldsymbol{A}E(\boldsymbol{S}))^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{A}E(\boldsymbol{S})) \\ &+\alpha\sum_{k=1}^{K}||\boldsymbol{a}_k-E(\boldsymbol{a_k}|\boldsymbol{S},\boldsymbol{a}_{j\neq k})||^2\} \end{aligned} \tag{2.24}$$

where $\alpha$ is a weighting parameter. This objective function has following characteristics:

1. The EM algorithm is used to estimate all parameters by treating $\boldsymbol{S}$ as missing observations and $\boldsymbol{A}$ as model parameters, and iteratively update the estimation of $\boldsymbol{S}$ and $\boldsymbol{A}$, as illustrated in Section 2.4.2.

2. Because $\boldsymbol{\Lambda}$ is anisotropic, the resulting distance is Mahalanobis distance that can account for the noise heterogeneous effect, rather than the Euclidean distance that ignores band-dependent noise characteristics.

3. $\boldsymbol{S}$ is unknown and treated as missing observations in the EM algorithm framework. We use $E(\boldsymbol{S})$ as the estimation of $\boldsymbol{S}$. To estimate parameters in $E(\boldsymbol{S})$,s we use $\boldsymbol{X}$ as the input to FCNN and optimize FCNN parameters. Given the estimated parameters in FCNN, we achieve $\hat{\boldsymbol{S}}=E(\boldsymbol{S})$, as illustrated in Section 2.4.3.

4. When estimating parameters in FCNN for obtaining $E(\boldsymbol{S})$, we use a reconstruction loss based on $(\boldsymbol{X}-\boldsymbol{A}E(\boldsymbol{S}))^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}-\boldsymbol{A}E(\boldsymbol{S}))$, which incorporates the SMM to ensure meaningful $\boldsymbol{S}$ estimation, as illustrated in Section 2.4.3.

5. The second term in the objective function Eq. 2.24 is used to regulate the endmembers, and $\alpha$ defines the relative importance of this regulation. The purified means approach is adopted to estimate $\boldsymbol{A}$, as illustrated in Section 2.4.4.

The proposed BCUN framework is illustrated as Figure 2.2.

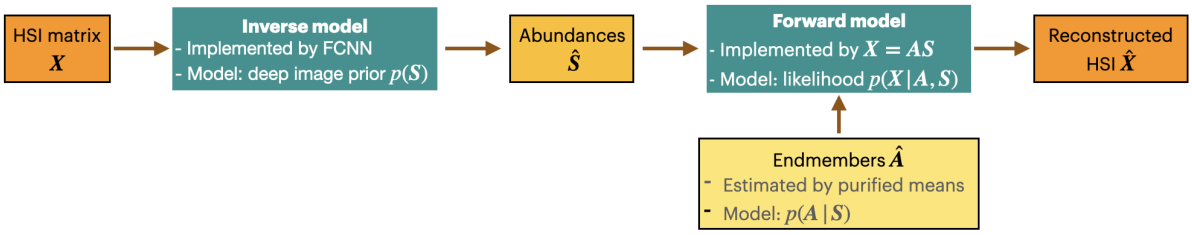



Figure 2.2: BCUN framework.The inverse unmixing model is implemented by FCNN, where DIP $p(\boldsymbol{S})$ is used to model the spatial correlation of the abundance field. The LSMM works as the forward model to reconstruct the HSI using the estimated $\boldsymbol{A}$ and $\boldsymbol{S}$. The data likelihood $p(\boldsymbol{A}|\boldsymbol{S})$ designed based on it is integrated into the Bayesian framework. The endmember $\boldsymbol{A}$ is modelled and estimated by a "purified means" approach which can be seamlessly integrated into the Bayesian framework by a designed conditional distribution of the endmembers given the abundance $p(\boldsymbol{A}|\boldsymbol{S})$.

### 2.4.2 EM iteration

The EM algorithm is an iterative approach which is widely used to optimize the incomplete data problem by iterating between the estimation of model parameters given missing observations and the estimation of missing observations given the model parameters[68]. In this chapter, we treat the abundances $\boldsymbol{S} = \{\boldsymbol{s}_i | i = 1, 2, ..., N\}$ as missing observations and treat endmembers $\boldsymbol{A}$ as model parameters to iteratively estimate both $\boldsymbol{A}$ and $\boldsymbol{S}$. The E-step is the computation of the expectation over the entire range of possible values of $\boldsymbol{S}$, i.e. $E(\boldsymbol{S})$, and the M-step updates $\boldsymbol{A}$ by minimizing Eq. 2.26 given $E(\boldsymbol{S})$. The main steps in EM algorithm to estimate $\boldsymbol{S}$ and $\boldsymbol{A}$ are summarized as follows.

- Initialization: Set the initial value for $\boldsymbol{A}$. The vertex component analysis (VCA) algorithm [57], a fast and popular unsupervised endmember extraction approach, is used to estimate the initial value of $\boldsymbol{A}$.
- E-step: Given endmembers $\boldsymbol{A}$ and the noise variance $\boldsymbol{\Lambda}$, estimate abundances $\boldsymbol{S}$ by optimizing a FCNN, as introduced in Section 2.4.3.
- M-step: Given $\boldsymbol{S}$ , estimate endmembers $\boldsymbol{A}$ and the noise variance $\boldsymbol{\Lambda}$. Endmembers $\boldsymbol{A}$ are estimated using purified means approach [59], which is discussed in Section 2.4.4. The noise variance $\boldsymbol{\Lambda}$ is estimated by the reconstructed residual, which is introduced in Section 2.4.5.

### 2.4.3 FCNN training

The objective of E-step of the EM algorithm introduced in Section 2.4.2 is to obtain the estimated abundance $\hat{\boldsymbol{S}}$ which is achieved by a FCNN model in Eq. 2.8.

To estimate $E(\boldsymbol{S})$, we first need to estimate the parameters in FCNN, i.e., $\boldsymbol{\beta}$. Here, we construct the following objective function to estimate $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = arg\min_{\boldsymbol{\beta}}\{((\boldsymbol{X} - \boldsymbol{A}f(\boldsymbol{X},\boldsymbol{\beta}))^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{X} - \boldsymbol{A}f(\boldsymbol{X},\boldsymbol{\beta})))\tag{2.25}$$

where $(\boldsymbol{X} - \boldsymbol{A}f(\boldsymbol{X},\boldsymbol{\beta}))^T\boldsymbol{\Lambda}^{-1}(\boldsymbol{X} - \boldsymbol{A}f(\boldsymbol{X},\boldsymbol{\beta}))$ is the Mahalanobis distance to account for the noise heterogeneous effect. Adam stochastic optimizer [69] is adopted in this work to estimate $\boldsymbol{\beta}$.

### 2.4.4 Purified means

For the M-step of the iterative EM algorithm introduced in Section 3.4.2, the update of endmember $\boldsymbol{A}$ is achieved by minimizing the following objective function, which is the second term in Eq. 2.24 constrained by the first term in Eq. 2.24.

$$\begin{aligned} J_A = arg\min_{\boldsymbol{a}_k}\sum_{j=1}^{K}||(\boldsymbol{a}_k - E(\boldsymbol{a}_k|\boldsymbol{S},\boldsymbol{a}_{j\neq k}))||^2, \\ \text{s. t. } \boldsymbol{x}_i - \sum_{j=1}^{K}\boldsymbol{a}_j s_i^j = 0 \end{aligned}\tag{2.26}$$

where $E(\boldsymbol{a}_k|\boldsymbol{S},\boldsymbol{a}_{j\neq k})$ is the conditional expectation of the $k$th endmember given $\boldsymbol{S}$ and all the other endmembers. Here, $E(\boldsymbol{a}_k|\boldsymbol{S},\boldsymbol{a}_{j\neq k})$ is estimated as the mean value of all purified pixels of the $k$th endmember $\boldsymbol{y}_i^k$, i.e., the pixels that are purified by removing the contribution of all the endmembers other than the $k$th endmember.

$$E(\boldsymbol{a}_k|\boldsymbol{S},\boldsymbol{a}_{j\neq k}) = \frac{1}{N}\sum_{i}^{N}\boldsymbol{y}_i^k.\tag{2.27}$$

where $N$ is the total number of pixels in the image. The constraint in Eq. 2.26 is implemented in the process of obtaining the purified pixel $\boldsymbol{y}_i^k$, which is formulated as follows [59],

$$\boldsymbol{y}_i^k = (\boldsymbol{x}_i^k - \sum_{j\neq k}^{K}s_i^j\boldsymbol{a}_j)/s_i^k, \quad s_i^k > 0.\tag{2.28}$$

### 2.4.5 Noise variance $\boldsymbol{\Lambda}$ update

Inspired by related researches [46], the noise variance $\boldsymbol{\Lambda}$ is estimated by calculating the variance of the reconstructed residual, which is formulated as follows.

$$\boldsymbol{r}_i = \boldsymbol{x}_i - \mathbf{A}\boldsymbol{s}_i \text{ for } i = 1, 2, \ldots, N\tag{2.29}$$

$$\mathbf{\Lambda} = \mathrm{VAR}\left(\{\boldsymbol{r}_i\}\right) \tag{2.30}$$

where VAR($\mathbf{t}$) is the function of calculating the variance of $\mathbf{t}$.

### 2.4.6 Summary of Complete Algorithm

Based on the EM steps described in Section 2.4.2, the complete algorithm used for solving BCUN can be achieved which is summarized in Algorithm 1.

**Algorithm 1** BCUN

**Input**: HSI $\boldsymbol{X}$, numbers of endmembers $K$, and iteration numbers $\tau$
**Output**: endmember matrix $\hat{\boldsymbol{A}}$, abundance matrix $\hat{\boldsymbol{S}}$
**Initialization**: $t := 1$, $\boldsymbol{A}^{(0)} = \mathrm{VCA}(\boldsymbol{X})$, $\mathbf{\Lambda}^{(0)} = \mathrm{VAR}(\boldsymbol{X})$

```
While t ⩽ τ do
  E-step:
    estimate β, given {X, A, Λ} according to Eq. 2.25.
    estimate S by f(X, β) in Eq. 2.8
  M-step:
    for k=1,2,...,K
      estimate {y_i^k} according to Eq. 2.28
      estimate a_k using {y_i^k} according to Eq. 2.27
    end for
    update Λ according to Eq. 2.29 and 2.30
end while
```

**Note**: $\boldsymbol{A} = \{\boldsymbol{a}_k | k = 1 : K\}$ and $\boldsymbol{S} = \{\boldsymbol{s}_i | i = 1 : N\}$

## 2.5 Experiments

### 2.5.1 Datasets

#### Simulated HSI

In this experiment, we simulate a $104 \times 104$ sized HSI with four endmembers with 200 bands. Each pixel in the simulated HSI is a mixture of the four endmembers. Mixed pixels are created using the four endmembers multiplied by four abundance maps following LSMM. Abundance maps are generated by first being divided into $8 \times 8$ sized homogeneous blocks of one of the four endmembers, then degrading the blocks by applying a spatial low-pass filter of $9 \times 9$. The resulting HSI is further degraded by zero-mean Gaussian noise with different noise variances in different bands. The band-dependent SNR values used

for simulation are estimated from the benchmark Indian Pines image. Suppose that the estimated SNR vector $\boldsymbol{q}$ has been centralized and normalized, the simulated SNR vector $\boldsymbol{r}$ can be obtained following the rule of $\boldsymbol{r} = \rho\boldsymbol{q} + \boldsymbol{c}$ [46], where $\rho$ is the amplitude that determines the magnitude of fluctuation of band-dependent SNR values, and $c$ is the center value that defines the overall SNR of all bands. Three HSIs with different noise levels (SNR = 10, 20, 30dB) are simulated by fixing $\gamma = 5$ and varying $c$.

#### Jasper Ridge HSI dataset

Jasper Ridge is a popular HSI with $512 \times 614$ pixels. Each pixel is recorded at 224 channels ranging from 0.38 $\mu m$ to 2.5 $\mu m$. Due to the difficulties of ground truth (GT) acquisition, a subimage with $100 \times 100$ was selected and its true endmember and abundance were collected. After removing the channels 1-3, 108-112, 154-166 and 220-224 (due to dense water vapor and atmospheric effects), 198 channels remained. There are four endmembers in this data, i.e., "road", "soil", "water" and "tree". The GT for the dataset is generated by Zhu (2014) [70], which has been widely used [71, 72, 73]. The RGB color composition of the HSI is shown in Figure 2.3 (a).

#### Saint Andre HSI dataset

Saint Andre HSI dataset is used in a unmixing paper [74] and published on the website https://zenodo.org/record/2142185#.YWbDyC0RoUt. The HSI contains $50 \times 50$ pixels, composed of 415 spectral bands ranging from 0.40 to 2.40 $\mu m$. The spectral bands with strong noise in the spectral ranges 1.34–1.55 and 1.80–1.98 $\mu m$ have been removed. The RGB color composition of the HSI is shown in Figure 2.3 (b). Endmember spectra of six distinct materials (i.e., "tree", "grass", "soil", "road", "building 1", and "building 2") have been manually extracted based on prior knowledge of the scene. This HSI dataset has only true endmembers without true abundances. Methods are evaluated mainly based on their performances on the endmember extraction. Abundances are also evaluated visually by comparing with the RGB image.

### 2.5.2 Experimental Setup

#### Methods Compared

The proposed method BCUN is compared with several traditional endmember extraction method including N-FINDR [75], PPI [58], VCA [57], K-P-Means [59] and several state-of-the-art methods including the spatial group sparsity regularized nonnegative matrix factorization (SGSNMF) approach [76], uDAs [49, 77] and a typical linear plug-and-play priors framework, PnP [78, 79, 80], which are proposed recently. PnP and uDAs are deep learning-based methods. It should be noted that K-P-Means and uDAs provide

(a) Jasper (b) Saint

Figure 2.3: RGB images of real HSIs.

estimations of endmembers and abundances, simultaneously, while for N-FINDR, PPI and VCA, only endmembers are estimated and the abundances are obtained in a post-processing stage by using the Sunsal method. PnP is an abundance mapping method which require knowing true endmembers. We used VCA-estimated endmembers as the input of PnP. K-P-Means is also solved by the EM algorithm where the E-step estimates abundances using Sunsal and M-step achieves endmembers update via purified means. Therefore, by comparing the performances between K-P-Means, the advantage of proposed method in terms of further improvement over K-P-Means can be demonstrated. Moreover, since both uDAs, PnP and the proposed method are deep neural network-based approaches but with vastly different implementations, the comparison between the two can justify the proposed BCUN approach and other contributions in the context of the proposed Bayesian MAP optimization framework. To demonstrate the advantage of addressing the heterogeneous noise effect, we design the BCUN0 method by replacing the Mahalanobis-distance loss in BCUN with Euclidean-distance loss, and compare it with BCUN.

### Numerical Measures

The spectral angle distance (SAD) defined as $\text{SAD} = \cos^{-1}\left(\left(\boldsymbol{a}^T\widehat{\boldsymbol{a}}\right)/(\|\widehat{\boldsymbol{a}}\|\|\boldsymbol{a}\|)\right)$ and the spectral information divergence (SID) defined as $\text{SID} = D(\boldsymbol{a}/\widehat{\boldsymbol{a}}) + D(\widehat{\boldsymbol{a}}/\boldsymbol{a})$ are used to measure the precision of the endmember extraction, where $D(\boldsymbol{x}/\boldsymbol{y})$ measures the relative entropy between $\boldsymbol{x}$ and $\boldsymbol{y}$. The accuracy of abundance estimation is measured using the abundance angle distance (AAD) and the abundance information divergence (AID) by replacing $\boldsymbol{a}$ with $\boldsymbol{s}$ in the above SAD and SID equations, as well as the mean squared error (MSE).

**Implement settings**

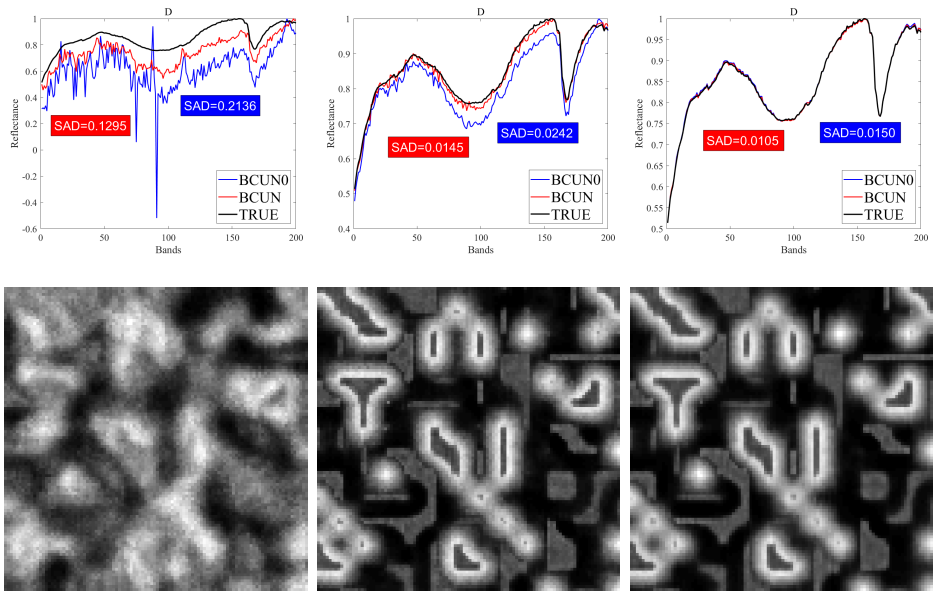
For the test on the simulated HSI, the learning rate of BCUN is empirically set as 0.00025. The number of EM iteration is 50. For each EM iteration, we use 20 epochs to train the neural network. We adopt three skip-connections in the FCNN as shown in Figure 2.1. For the test on the real Jasper HSI, the learning rate of BCUN is set as 0.03. The number of EM iteration is 60. For each EM iteration, the network is trained 200 times. We adopt one skip-connection in the FCNN for Jasper HSI dataset. For the test on the real Saint HSI, the learning rate of BCUN is set as 0.001. The number of EM iteration is 10. For each EM iteration, the network is trained 150 times. We adopt two skip-connections in the FCNN for Saint HSI dataset.

Model parameters including the number of layers, the learning rate and iteration times are determined empirically. Hyperparameters are required to be adjust for different HSIs since HSIs vary in the sensor, data size and complexity. The real HSIs contain fewer mixed pixels and spatial texture features, which can be unmixed with a simpler network. So, we reduce the layers of the network to accelerate the model training. The learning rate needs to be adjusted to suit the network architecture. All data processing is conducted using the Python language under the Pytorch framework.

The benchmark PnP, uDAs and SGSNMF methods were implemented in the MATLAB 2018 under an Intel Xeon Silver 4110 CPU @2.10GHz. The VCA, PPI, N-FINDR, K-P-means and proposed BCUN methods were implemented in the Pytorch toolbox with an NVIDIA GeForce RTX 2080 Ti GPU.

### 2.5.3 Simulated Study

To compare the performances between traditional MSE loss that is built upon the Euclidean distance in BCUN0 and the proposed Mahalanobis distance loss in BCUN, we first apply BCUN and BCUN0 on the three simulated HSIs with SNR = 10, 20, 30 dB. Figure 2.4 shows the endmember (the first row) and abundance maps of one endmember achieved by BCUN and BCUN0 (the second and the third row separately) with different SNR values, i.e., 10, 20, 30 dB. On all endmember figures (in the first row), the endmember extracted by BCUN (red lines) is closer to the true endmember (black lines) than the endmember estimated by BCUN0 (blue lines). In addition, endmembers achieved by BCUN are smoother than those obtained by BCUN0, which is specifically obvious when the noise level is high (e.g., on Figure 2.4 (left) where SNR = 10dB). This demonstrates that the proposed M-distance loss in FCNN can better deal with the heterogeneous noise than the traditional MSE loss for enhanced endmember and abundance estimation, especially when the noise level of HSI is high. Figure 2.4 shows negative values in endmembers at SNR=10 because BCUN0 does not consider the noise heterogeneous effect and leads to large bias in endmember estimation at some spectral bands. The endmembers estimated by BCUN are all positive at any noise level. It also shows the importance of addressing the noise heterogeneity, especially when SNR values are low.

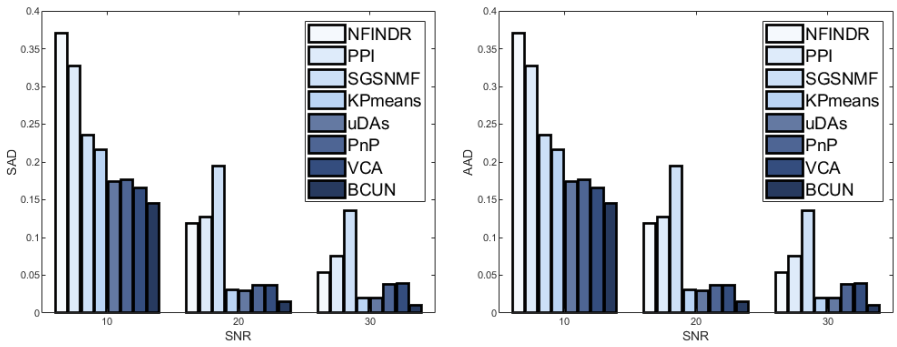



(a) AAD=0.5678 (b) AAD=0.1732 (c) AAD=0.1636

(d) AAD=0.4573 (e) AAD=0.1023 (f) AAD=0.0735

Figure 2.4: The endmember (the first row) and abundance maps of one endmember achieved by of BCUN0 and BCUN (the second and the third row separately) with different SNR values, i.e., 10, 20, 30dB respectively from left column to right column.

To evaluate the performance of the proposed BCUN, we compare it with the other methods introduced in Section 2.5.2 by testing all methods with 10 independent runs on three simulated HSIs with different SNR levels. The mean values and the standard deviation (std) values of SAD, AAD, SID, AID, and MSE are summarized in Table. 2.1. The running time of different SU methods are summarized in Table 2.2. The methods comparison in terms of SAD and AAD are also illustrated in Figure 2.5. Given that the smallest SAD and AAD mean values were all achieved by BCUN, BCUN outperformed all the other methods on endmember extraction and abundance estimation at all noise levels, indicating that the proposed BCUN is able to more accurately model the abundance and endmember information in mixed pixels by accounting for the heterogeneous noise and the spatial correlation effect in a Bayesian fully convolutional neural network framework. In particular, the observation that the mean AAD values achieved by BCUN are smaller than those achieved by other values clearly reflects advantages of the designed BCUN approach over the traditional Sunsal approach for abundance mapping. Some abrupt changes and spatial detail information was well delineated by BCUN, which justified the use of DIP to capture the nonstationary spatial correlation information in HSI. Convolutional layers tend to smooth features, but the FCNN structure is able to capture the spatial textural information which is DIP. The methods comparison in terms of SAD and AAD is also illustrated in Figure 2.5.




(a) SAD (b) AAD

Figure 2.5: AAD, AAD bar graphs achieved by different methods with different SNR values, i.e., 10, 20, 30dB.

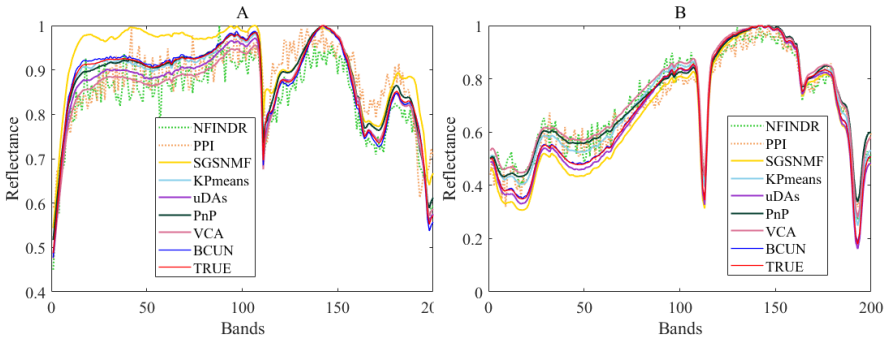
Figure 2.6 shows the four endmembers estimated by different SU methods as well as the true endmember (the red line) at the noise level of SNR=30dB. The proposed method BCUN appears to achieve the closest endmember spectrum (dark blue lines) to the true endmember (red lines) for all the 4 endmembers. Endmembers estimated by deep learning-based uDAs and PnP algorithms (purple and dark green lines) are closer to the true endmembers than the other traditional methods but also introduces an obvious bias away from the true endmember 1. Endmembers achieved by K-P-Means (light blue lines) are smooth and very close to the true endmember 1 and 4, but have a larger bias away from true endmember 2 and 3. VCA (pink lines) generates endmembers relatively smooth

Table 2.1: Average SAD, AAD, SID, AID and MSE of abundances, obtained from 10 independent runs by different methods using the simulated data over SNR from 10 to 30dB. The best results are in bold.

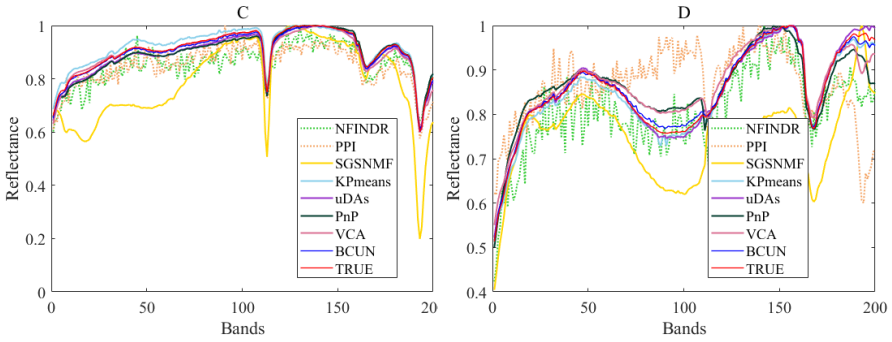
| SNR=10, 4 endmembers, 10816 pixels | | | | | |
|---|---|---|---|---|---|
| Method | SAD | AAD | SID | AID | MSE |
| NFINDR | 0.3704±3.09% | 0.7009±2.38% | 0.1537±0.0308% | 6.0106±0.3139% | 0.0469 |
| PPI | 0.3271±5.56% | 0.7284±**1.47%** | 0.1791±0.0071% | 6.3262±**0.1043%** | 0.0559 |
| SGSNMF | 0.2360±7.51% | 0.5704±2.37% | 0.3374±0.0869% | 4.6836±0.2111% | 0.0351 |
| KPmeans | 0.2162±4.93% | 0.7208±4.50% | 0.0268±0.0183% | 6.5955±0.6952% | 0.0475 |
| uDAs | 0.1736±2.67% | 0.7043±6.17% | 0.0185±0.0111% | 6.3495±1.2081% | 0.0432 |
| PnP | 0.1766±**1.98%** | 0.7422±7.66% | 0.0201±**0.0063%** | 7.8469±2.1994% | 0.0580 |
| VCA | 0.1659±3.37% | 0.7745±7.55% | **0.0156**±0.0098% | 8.9710±2.1885% | 0.0560 |
| BCUN | **0.1449**±9.28% | **0.4573**±12.68% | 0.0394±0.0545% | **3.7158**±0.8387% | **0.0161** |
| SNR=20, 4 endmembers, 10816 pixels | | | | | |
| Method | SAD | AAD | SID | AID | MSE |
| N-FINDR | 0.1184±**0.21%** | 0.4280±**1.03%** | 0.0168±0.0011% | 3.4733±0.1174% | 0.0217 |
| PPI | 0.1330±1.97% | 0.6112±7.20% | 0.0210±0.0037% | 5.1706±0.3289% | 0.0354 |
| SGSNMF | 0.1950±13.55% | 0.4832±17.36% | 0.1609±0.1651% | 4.0203±2.0509% | 0.0154 |
| KPmeans | 0.0308±0.90% | 0.2979±4.92% | 0.0012±0.0008% | 1.9187±0.5029% | 0.0100 |
| uDAs | 0.0294±1.51% | 0.2451±4.07% | 0.0014±0.0014% | 1.4097±0.4577% | 0.0060 |
| PnP | 0.0364±0.96% | 0.2788±6.98% | 0.0020±0.0008% | 1.6709±1.0786% | 0.0064 |
| VCA | 0.0362±1.21% | 0.3055±6.18% | 0.0020±0.0012% | 1.8566±0.7442% | 0.0081 |
| BCUN | **0.0145**±0.59% | **0.1023**±1.70% | **0.0004±0.0002%** | **0.5288±0.1152%** | **0.0012** |
| SNR=30, 4 endmembers, 10816 pixels | | | | | |
| Method | SAD | AAD | SID | AID | MSE |
| N-FINDR | 0.0537±0.75% | 0.2215±5.88% | 0.0038±0.0009% | 1.3271±0.5318% | 0.0039 |
| PPI | 0.0752±1.35% | 0.5262±2.98% | 0.0235±**0.0000%** | 32.1737±**0.0006%** | 0.0334 |
| SGSNMF | 0.1359±13.19% | 0.3427±9.88% | 0.1660±0.3383% | 2.2419±0.9875% | 0.0201 |
| KPmeans | 0.0195±**0.48%** | 0.1126±2.68% | 0.0010±0.0003% | 0.4808±0.2388% | 0.0014 |
| uDAs | 0.0195±1.10% | 0.1377±4.95% | 0.0007±0.0010% | 0.6136±0.3938% | 0.0034 |
| PnP | 0.0383±0.81% | 0.2657±5.12% | 0.0025±0.0007% | 1.4322±0.5323% | 0.0060 |
| VCA | 0.0396±1.03% | 0.2321±7.14% | 0.0026±0.0009% | 1.0945±0.6274% | 0.0041 |
| BCUN | **0.0105**±0.56% | **0.0735±1.14%** | **0.0002**±0.0002% | **0.2965**±0.1047% | **0.0010** |

Table 2.2: The running times of different methods on the simulated data with SNR=30dB.

| **Methods** | **VCA** | **PPI** | **BCUN** | **N-FINDR** | **KPmeans** | **PnP** | **uDAs** | **SGSNMF** |
|---|---|---|---|---|---|---|---|---|
| **Time(s)** | **0.07** | 0.16 | 18.61 | 0.35 | 43.76 | 12.18 | 24.51 | 91.01 |




(a) Endmember 1 (b) Endmember 2

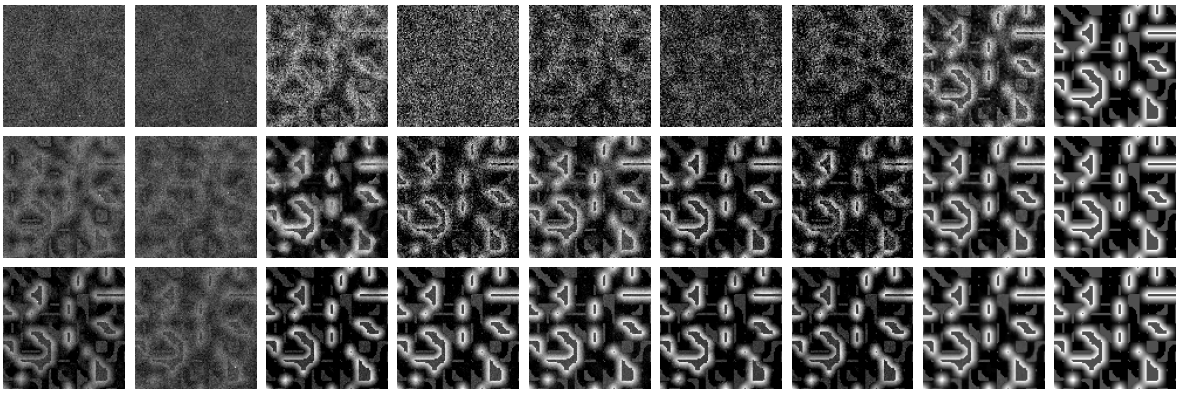



(c) Endmember 3 (d) Endmember 4

Figure 2.6: The endmembers achieved by different methods when SNR equals 30dB.

by with a very large bias from the true endmembers. Endmembers obtained by PPI and N-FINDR (orange and green dotted lines) are very noisy indicating that these two methods are sensitive to resist noise. These visual interpretations align with the quantitative results in Table 2.1.

Figure 2.7 displays the abundance maps of one endmember generated by PPI, N-FINDER, VCA, K-P-Means, uDAs, PnP and the proposed BCUN. As we can see, BCUN achieved abundance maps that are very close to the GT maps. K-P-Means tends to preserve some noise due to the failure to account for the spatial correlation effect in HSI, which is especially true when the noise level is high (i.e. SNR = 10, 20dB) where the abundance maps of these traditional methods are very noisy. The state-of-the-art methods uDAs and PnP achieve smoother and clearer abundance maps than other methods but are still noisy than BCUN at all noise levels.



(a) N-FINDR (b) PPI (c) SGSNMF (d) KPmeans (e) uDAs (f) PnP (g) VCA (h) BCUN (i) GT

Figure 2.7: The abundance maps achieved by different methods on one endmember with different SNR values, i.e., 10, 20, 30dB from the top row to bottom row. Brighter pixels indicate high abundance while darker pixels indicate low abundance.

### 2.5.4 Test on real HSIs

#### Jasper Ridge Scene

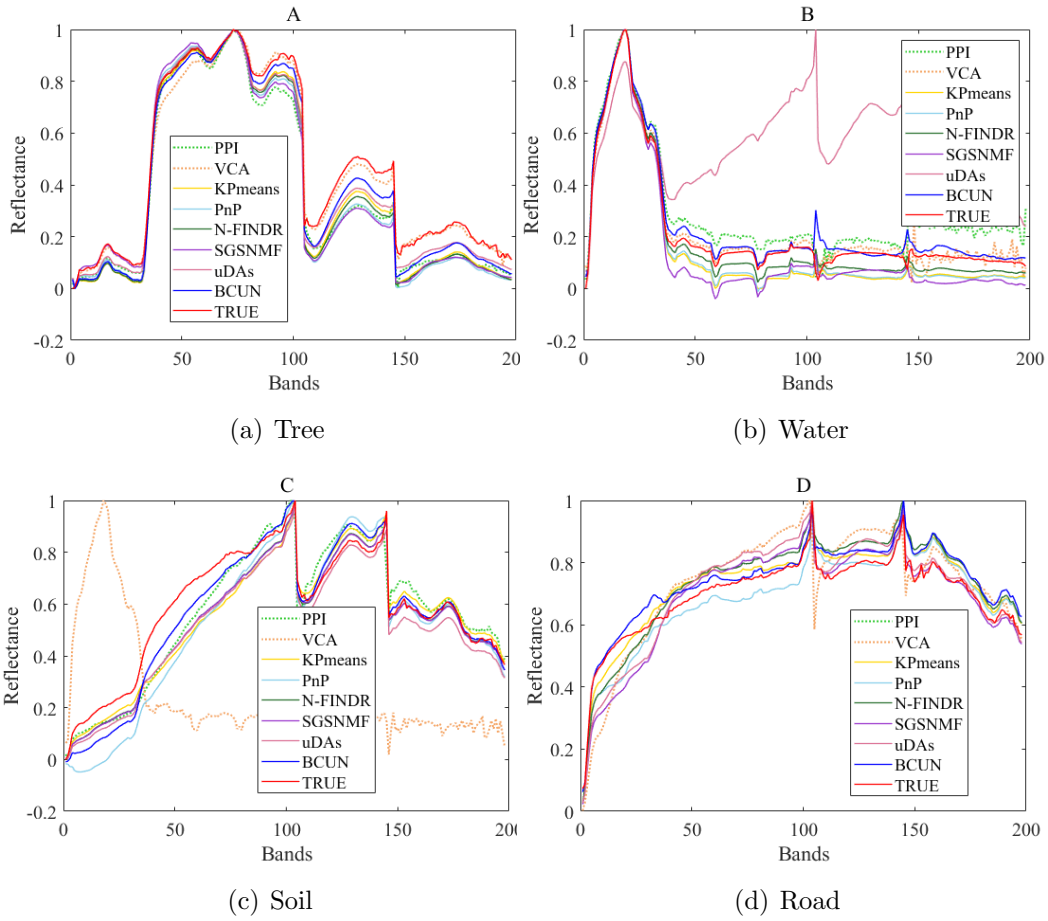
All methods are conducted on the Jasper Ridge HSI. The estimated endmembers and abundances are compared in Figure 2.8 and Figure 2.9 separately. Numerical measurements achieved by all methods are summarised in Table 2.3. Table 2.3 indicates that BCUN achieves the lowest mean values of SAD, SID, AAD and AID among all methods tested, indicating that BCUN extracts endmembers and abundances more accurately than other methods.

Figure 2.9 shows that overall BCUN is more capable of generating abundance maps that are close to the GT maps in terms of both the intense brightness and the structural characteristics. For example, the bright water area in the second row of Figure 2.9 generated by BCUN is closer to the intensity brightness of the GT, and also BCUN generates the highlighted red box area in Figure 2.9 that is closest to the GT, whereas the red box areas of the other methods wrongly highlight the road as water in the water abundance maps.

Figure 2.8 shows the four endmembers achieved by different methods as well as the true endmembers (red line). On average, endmembers obtained by the proposed BCUN method are the closest to the true endmembers among all methods, which is consistent with the SAD statistics in Table 2.3.

#### Saint Andre Scene

Table 2.4 demonstrates that the proposed BCUN achieves the best endmember estimation performance with the lowest SAD and SID values, compared with other methods. Figure



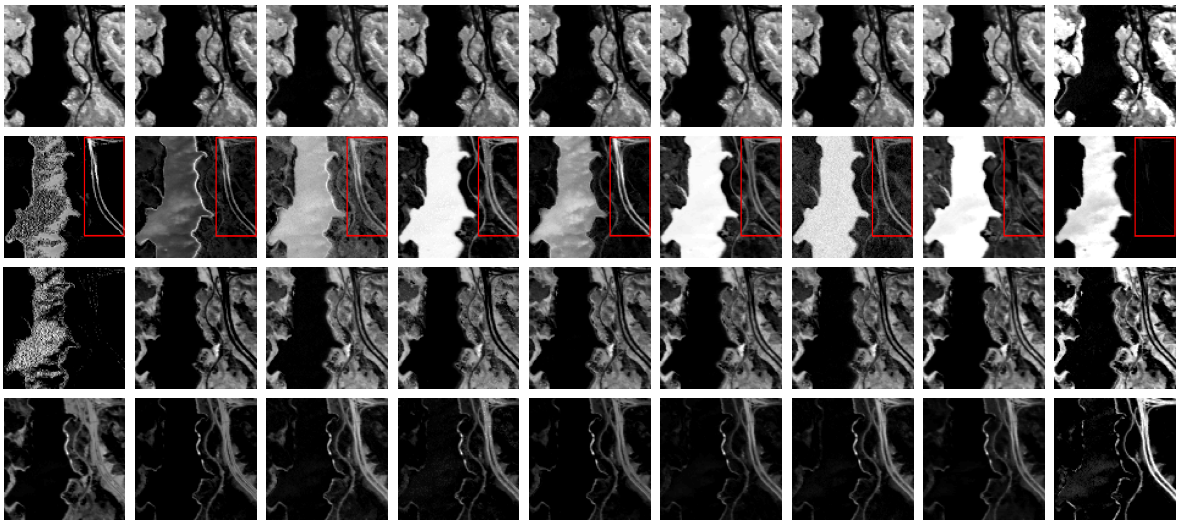

(a) Tree (b) Water

(c) Soil (d) Road

Figure 2.8: The estimated endmembers achieved by different methods for Jasper Ridge HSI data, along with the true endmember.

Table 2.3: Average SAD, SID, AAD, AID and MSE of abundances, obtained by different methods for Jasper Ridge HSI data. The best results are in bold.

| Jasper Ridge HSI, 4 endmembers, 10000 pixels | | | | | |
|---|---|---|---|---|---|
| Methods | SAD | SID | AAD | AID | MSE |
| PPI | 0.3085 | 0.3710 | 0.6809 | 8.5008 | 0.0898 |
| VCA | 0.2537 | 0.1589 | 0.4288 | 4.1735 | 0.0530 |
| KPmeans | 0.1680 | 0.1190 | 0.3938 | 4.1484 | 0.0555 |
| PnP | 0.1622 | 0.1322 | 0.3116 | 2.6532 | 0.0202 |
| N-FINDR | 0.1604 | 0.0565 | 0.3537 | 3.0730 | 0.0278 |
| SGSNMF | 0.1392 | 0.1571 | 0.2897 | 2.8536 | **0.0201** |
| uDAs | 0.1181 | 0.0624 | 0.3044 | 2.8851 | 0.0242 |
| BCUN | **0.1065** | **0.0383** | **0.2892** | **2.4576** | 0.0207 |

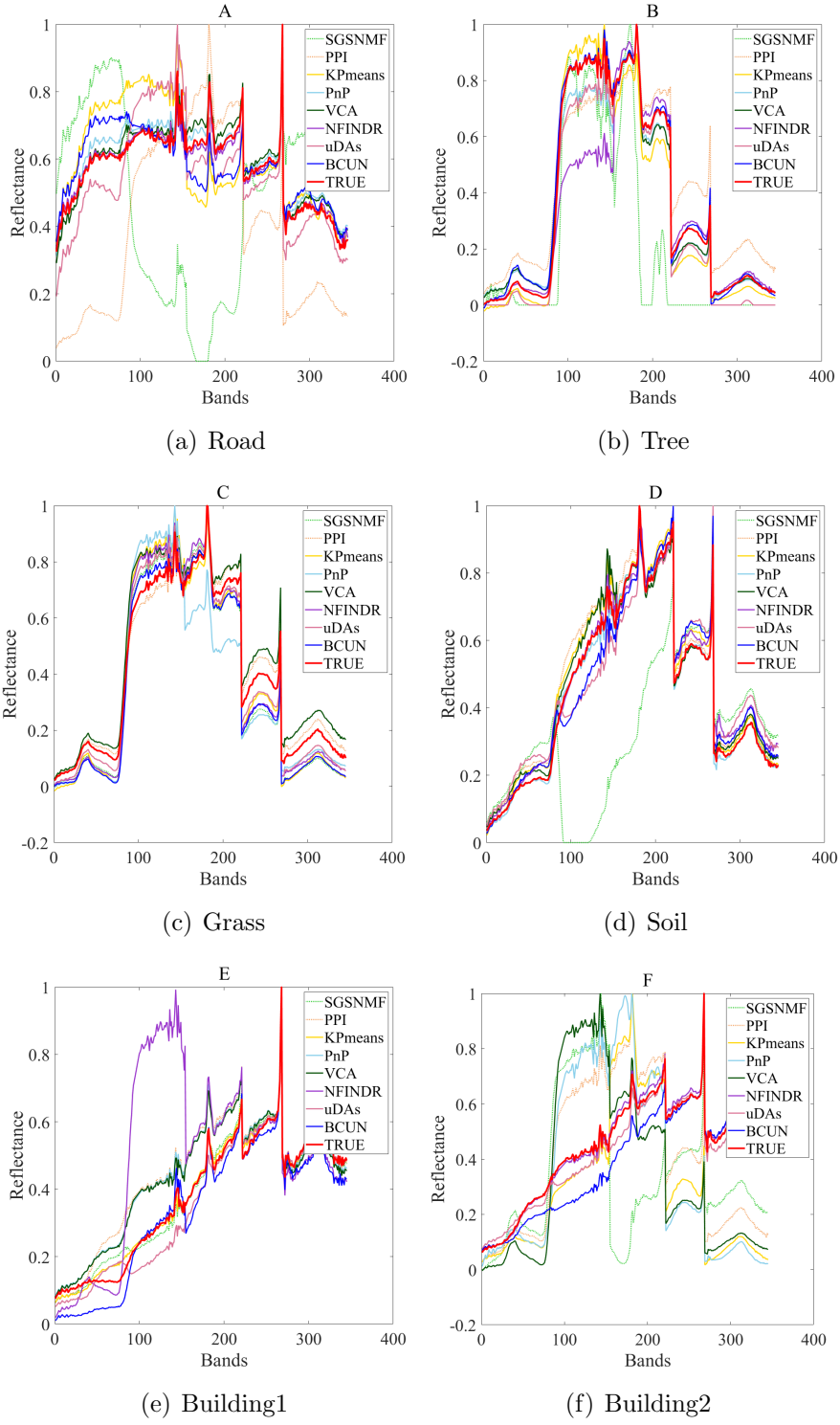


(a) PPI (b) VCA (c) KPmeans (d) PnP (e) N-FINDR (f) SGSNMF (g) uDAs (h) BCUN (i) GT

Figure 2.9: The abundance maps achieved by different methods on four endmembers (tree, water, soil, road) respectively from the top row to bottom row for Jasper Ridge HSI data.
Brighter pixels indicate high abundance while darker pixels indicate low abundance. It shows that overall BCUN is more capable of generating abundance maps that are close to the GT maps in terms of both the intense brightness and the structural characteristics. For example, the bright water area in the second row generated by BCUN is closer to the intensity brightness of the GT, and also BCUN generates the highlighted red box area that is closest to the GT, whereas the red box areas of the other methods wrongly highlight the road as water in the water abundance maps.

Table 2.4: Average SAD and SID obtained by different methods for Saint HSI data. The best results are in bold. This HSI provide only true endmembers manually extracted based on prior knowledge of the scene without true abundances.

| Saint HSI, 6 endmembers, 2500 pixels | | |
|---|---|---|
| Methods | SAD | SID |
| SGSNMF | 0.4462 | 2.8081 |
| PPI | 0.2327 | 0.1614 |
| KPmeans | 0.1992 | 0.2286 |
| PnP | 0.1930 | 0.1872 |
| VCA | 0.1813 | 0.1732 |
| N-FINDR | 0.1343 | 0.0503 |
| uDAs | 0.1292 | 0.4733 |
| BCUN | **0.1137** | **0.0408** |

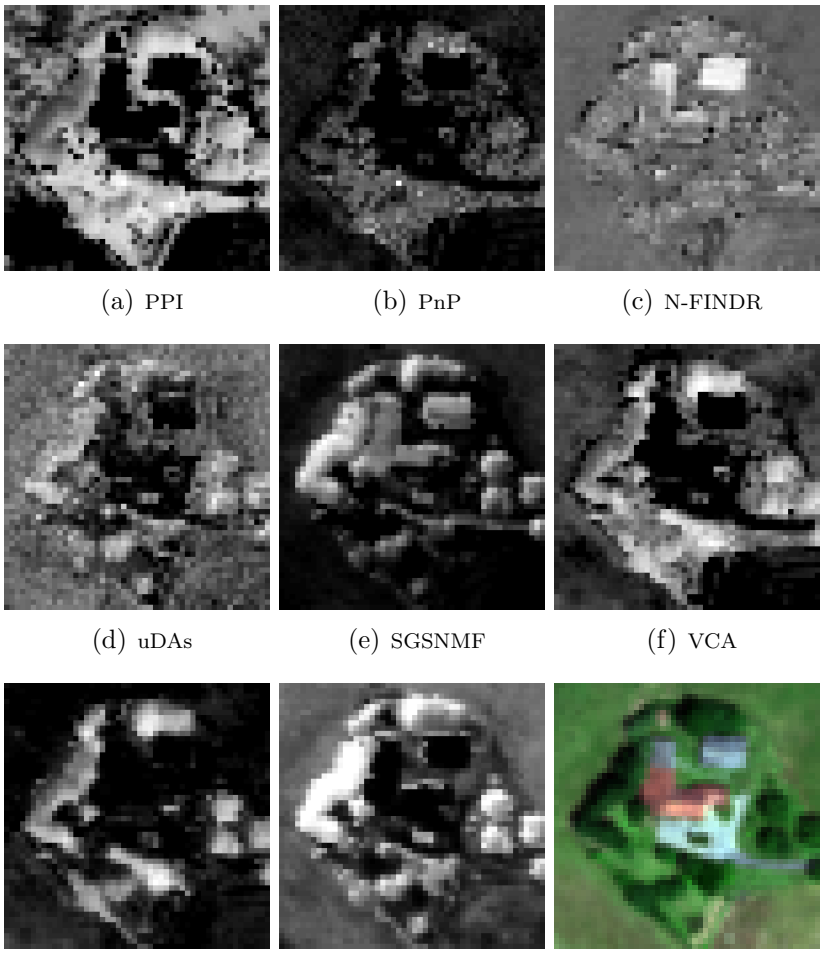



(a) Road (b) Tree

(c) Grass (d) Soil

(e) Building1 (f) Building2

Figure 2.10: The estimated endmembers achieved by different methods for Saint HSI data, along with the true endmember. Endmembers extracted by BCUN method (blue line) are the closest to the true endmembers (red line) over six categories. Although some methods perform very well on specific endmembers (e.g., uDAs on soil, and N-FINDR on building 2), they generate big bias on some other endmembers (e.g., uDAs on road, and N-FINDR on building 1). BCUN is demonstrated to be able to distinguish and extract different endmembers well.

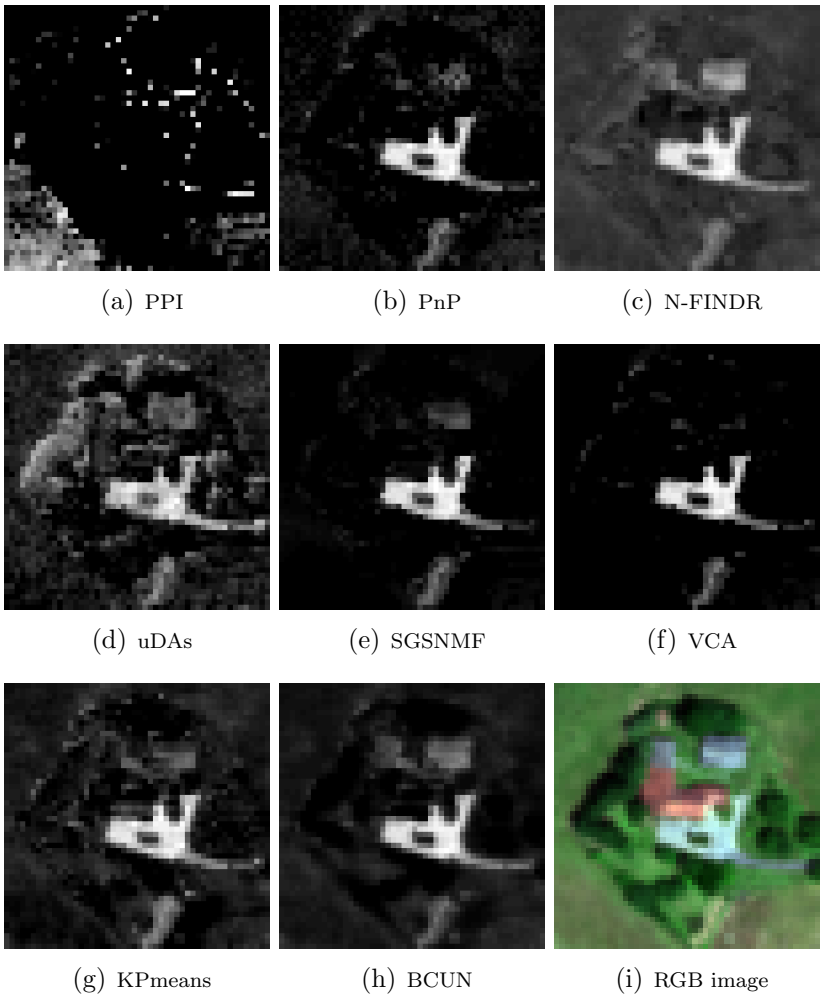


(a) PPI (b) PnP (c) N-FINDR

(d) uDAs (e) SGSNMF (f) VCA

(g) KPmeans (h) BCUN (i) RGB image

Figure 2.11: The abundance maps of tree achieved by different methods on Saint HSI data.
Brighter pixels indicate high abundance while darker pixels indicate low abundance. BCUN extracts the tree endmember very well (see Figure 2.10 (b)). Correspondingly, the tree abundance map generated by BCUN (h) shows very clear tree positions and edges, which is highly consistent with the RGB image. PPI, N-FINDR and uDAs cannot distinguish tree and grass very well. PnP, uDAs and VCA tend to preserve more noise than BCUN. K-P-means shows less spatial texture information than BCUN.



(a) PPI (b) PnP (c) N-FINDR

(d) uDAs (e) SGSNMF (f) VCA

(g) KPmeans (h) BCUN (i) RGB image

Figure 2.12: The abundance maps of road achieved by different methods on Saint HSI data.
Brighter pixels indicate high abundance while darker pixels indicate low abundance. BCUN outperforms other methods in spatial information preservation by delineating the road outline clearly and smoothly.

2.10 displays six endmembers extracted by different methods. Endmembers extracted by BCUN method (blue line) are the closet to the true endmembers (red line) over six categories. Although some methods perform very well on specific endmembers (e.g., uDAs on soil, and N-FINDR on building 2), they generate big bias on some other endmembers (e.g., uDAs on road, and N-FINDR on building 1). BCUN is demonstrated to be able to distinguish and extract different endmembers well.

Since the true abundances were not provided in the HSI dataset, we visually evaluate abundances by comparing abundance maps with the RGB image. Abundances of road and tree are presented in Figure 2.12 and Figure 2.11 separately. BCUN extracts the tree endmember very well by achieving the endmember spectra (blue line) very close to the true endmember (red line) (see Figure 2.10 (b)). Correspondingly, the tree abundance map generated by BCUN (Figure 2.11 (h)) shows very clear tree positions and edges, which is highly consistent with the RGB image. Some of other methods (e.g. PPI, N-FINDR and uDAs) cannot distinguish tree and grass very well. PnP, uDAs and VCA tend to preserve more noise than BCUN. KPmeans shows less spatial texture information than BCUN.

BCUN outperforms other methods in spatial information preservation by delineating the road outline clearly and smoothly. Although the road abundance achieved by BCUN (Figure 2.12 (g)) preserves some other objects in the background, the road object is still the brightest area which is easily distinguished from other objects. It is reasonable that the abundance of buildings is not zero on the road abundance map because building and road may contain similar physical materials.

## 2.6 Conclusion

In this chapter, we presented a Bayesian framework for hyperspectral unmixing in which the EM algorithm was applied to solve the MAP problem. Unlike the most deep learning-based SU techniques, BCUN adopted FCNN rather than fully connected layers or Sunsal to better exploit the spatial correlation effect in HSI for enhanced abundance modelling and estimation. In addition, the noise heterogeneity effect in HSI was addressed by the M-distance loss. A conditional distribution of the endmember is designed, leading to an efficient purified means approach for endmember estimation. The above key components are seamlessly integrated into a Bayesian MAP framework, which is solved by the proposed EM approach. Therefore, the proposed BCUN approach constitutes a complete Bayesian approach with advanced modelling and optimization approaches for enhanced spectral unmixing. The proposed approach was tested on both real and simulated HSIs, in comparison with several other popular SU methods, and results demonstrated that the proposed BCUN method was more capable of accurately and efficiently estimating both the endmember and abundance in HSIs.

# Chapter 3

# Bayesian Subpixel Mapping Autoencoder Network for Hyperspectral Images

Although subpixel mapping (SPM) methods for hyperspectral images (HSIs) via deep learning technologies have the advantage of learning the complex non-linear relationship between HSIs and subpixel labels, it is challenging to integrate the fully convolutional neural network (FCNN) and forward models in a coherent framework. This chapter proposes an unsupervised Bayesian subpixel mapping network for HSIs with the following characteristics. First, the deep image prior achieved by an FCNN is used to efficiently and adaptively model the spatial correlation in the subpixel label domain. Second, a discrete spectral mixture model is integrated with the FCNN, and as such the forward model information is leveraged to enhance SPM. Third, the combination of FCNN and forward models in an auto-encoder architecture allows the model learn from both data and knowledge without requiring the groundtruth data. Fourth, a designed expectation-maximization approach is applied to solve the resulting maximum a posteriori problem, where a purified means approach extracts endmembers and the FCNN estimates subpixel labels iteratively. Comparative experiments on both real and simulated HSIs demonstrate that the proposed method outperforms other methods from the perspectives of numerical accuracies and visual subpixel mapping results.

## 3.1 Introduction

Hyperspectral imaging is a rapidly growing remote sensing technique and has been used widely used for applications, such as ground target classification [41, 42, 29], agricultural management [3] and environmental monitoring [44, 4]. However, due to the trade-off between the spectral resolution and spatial resolution in HSIs, pixels in hyperspectral

images (HSIs) usually contain spectral contributions from multiple materials. Subpixel mapping (SPM) aims to generate a label map with a finer spatial resolution by dividing the original mixed pixel into several subpixels [81]. SPM relies on the spatial dependence assumption between and within pixels to improve the spatial resolution of a HSI label map [81, 82]. However, most SPM methods suffer from insufficient spatial correlation modelling or heavily relying on training samples. Therefore, it is valuable to develop an efficient unsupervised SPM algorithm with the prior information that can adaptively characterize the global heterogeneity spatial correlation in real HSIs in a unified framework. To achieve this, three key issues are identified to be addressed.

First, although the use of deep learning (DL)-based prior is essential for accurate SPM, it is not sufficiently researched. Spatial dependence assumption (SDA)-based methods assume that close pixels have higher correlation than distant ones [83], such as the spatial attraction SPM (SASM) [84, 85, 86], pixel swapping algorithm (PSSM) [87] and genetic algorithm SPM [88, 89]. However, SDA prior is limited by the sampling scale on HSI pixels [90], and near pixels are not always more correlated than distant ones. Although Markov random field (MRF) prior is more precise than SDA priors, [91, 43], the fixed priors cannot sufficiently characterize a HSI with the geographically realistic distribution [92]. DL-based methods make the use of neural networks to achieve a learnable prior from training samples. Both fully-connected networks [93, 94] and convolutional neural networks (CNNs) [95] have been adopted to solve the SPM problem. Compared with traditional patch-based CNN, the fully CNN (FCNN) can better capture the spatial correlation effect in HSIs [35, 16]. The structure of FCNN is capable of capturing statistical image information and imposing an effective DIP to restore high-quality images from low-quality images without seeing a large training dataset [17]. Recent publications show the effectiveness of DIP for image restoration [18, 19, 20], e.g., HSI unmixing [31], super-resolution [21], image inpainting [16] and denoising [22, 23]. However, FCNN-based approaches leveraging the DIP have not been adapted to HSI for enhanced SPM. Therefore, how to integrate the DIP into SPM for better modelling the spatial correlation in HSIs is an important research issue.

Second, although the use of the prior knowledge is essential for enhanced SPM, most DL-based SPM methods heavily rely on the groundtruth data and ignore the prior knowledge. For example, He et al.'s method [83] creates a network for SPM that requires training sample pairs. However, the groundtruth is barely available for HSI datasets. To overcome the training data limitation, the knowledge information (e.g., a forward model mapping the finer label map to a HSI) and the observed HSI need to be effectively leveraged. The encoder-decoder architecture has the potential to address this issue. It takes a learnable inverse model (i.e., a deep neural network) as the encoder and a fixed forward model as the decoder, which is trained with a data reconstruction loss without requiring any groundtruth data. Although the strategy succeeds in several image restoration tasks [16], as well as in the HSI unmixing [31], it has been rarely adopted to SPM. Therefore, it is critical to build a coherent auto-encoder that integrates a designed forward model for unsupervised SPM.

Third, it is critical to design an optimization framework to estimate the underlying subpixel labels in a coherent way for enhanced SPM, which is capable of capturing spa-

tial correlation effect in the subpixel label field and incorporating the forward generation model. The Bayesian method is used widely to solve the image inverse problem by integrating prior knowledge of the desired variable with the posterior distributions given the image observation [28, 30]. For example, the MRF has been incorporated as the prior for HSIs in the Bayesian framework [29]. The DL technique solves inverse problems in many applications from the new perspective by various network architectures [96]. However, studies using DL for SPM in a Bayesian framework are few. The SPM problem can be derived as a maximum a posteriori (MAP) problem in a Bayesian framework and solved with the expectation-maximization (EM) approach. Therefore, it is important to design an EM algorithm tailored for the SPM problem in the Bayesian framework.

In this chapter, a Bayesian SPM autoencoder network (BSMAN) for HSI which integrates the forward model with an FCNN in a Bayesian framework is designed and implemented. The BSMAN with an encoder-decoder architecture has the following three characteristics.

1. A skip-connection FCNN works as the encoder to generate subpixel labels, where the deep image prior (DIP) is used to model the spatial correlation in the subpixel label field.

2. A dedicated discrete linear spectral mixture model (DSMM) is integrated with the FCNN. This forward model maps subpixel labels to the discrete abundance of original coarse pixel, and then reconstructs the HSI with the discrete abundance and extracted endmembers.

3. An SPM autoencoder network is designed in a Bayesian framework without requiring label samples for training. The resulting MAP problem is solved by a designed EM algorithm. The latent variable (i.e., subpixel labels) and the model parameters (i.e., endmembers) are updated iteratively.

Comparative experiments on both simulated and real HSI demonstrate that the proposed approach can generate subpixel labels with higher accuracy than other traditional and state-of-the-art approaches by exploring the spatial correlation effect in the HSI. The remainder of the paper is organized as follows. Section 3.2 formulates the SPM problem in a Bayesian framework. The design of the network and its rationale are detailed in Section 3.3. Section 3.4 introduces the optimization scheme of the proposed BSMAN. Section 3.5 conducts experiments on both simulated and real HSIs.

## 3.2 Problem formulation

Following the notation in Chapter 2, we assume that an observed HSI data cube $\boldsymbol{X}$ has $P$ spectral bands, $N$ pixels containing $m$ rows and $n$ columns, and the term $I$ represents the

set of coarse pixel sites in HSI, we denote the observed reflectance of the pixel at site $i$ by $\boldsymbol{x}_i$, which is a $P \times 1$ vector. Then the HSI can be expressed as $\boldsymbol{X} = \{\boldsymbol{x}_i | i = 1, 2, ..., m \times n\}$. The term $J$ represents the set of subpixel positions within each coarse pixel, which contains a total of $c^2$ positions. For example, when $c = 2$, a coarse pixel is divided into $2 \times 2$ subpixels, and the coarse HSI containing $m \times n$ pixels corresponds to a fine HSI with the size of $2m \times 2n$ subpixels.

Assuming that the HSI covers $K$ classes, the SPM aims to infer subpixel labels $\boldsymbol{L} = \{\boldsymbol{l}_{i,j} | i \in I, j \in J\}$ in the HSI, where $\boldsymbol{l}_{i,j}$ is a one-hot $K \times 1$ vector where the non-zero element defines the class of the subpixel. The discrete abundance $\boldsymbol{s}_i$ of a coarse pixel is determined by the proportions of subpixels labels, which is expressed as follows,

$$\boldsymbol{s}_i = \frac{\sum_{j=1}^{c^2} \gamma\left(\max(\boldsymbol{l}_{i,j}), k\right)}{c^2} \tag{3.1}$$

where $\max(\boldsymbol{l}_{i,j})$ returns the index of the non-zero element (i.e., the hard class label), and $\gamma(u, v)$ is the Kronecker delta function where $\gamma(u, v) = 1$ for $u = v$ and $\gamma(u, v) = 0$ otherwise. For example, as illustrated in Figure 3.1, when $c = 2$, there is one "tree" subpixel and three "flower" subpixels. Then the discrete abundance $\boldsymbol{s}_i$ corresponding to the coarse pixel is a $2 \times 1$ vector written as $[1/4; 3/4]$.

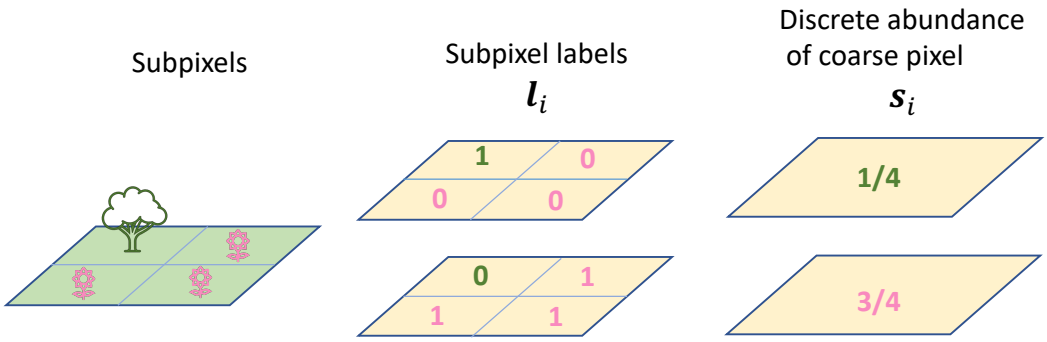



Figure 3.1: Illustration of the relationship between the subpixel labels and discrete abundances coarse pixel.

Then, a coarse pixel $\boldsymbol{x}_i$ can be formulated as a linear combination of $K$ endmembers $\boldsymbol{A} = \{\boldsymbol{a}_k | k = 1, 2, ..., K\}$ weighted by discrete abundances plus noise $\boldsymbol{n}_i$, which is a DSMM:

$$\boldsymbol{x}_i = \sum_{k=1}^{K} \boldsymbol{a}_k \frac{\sum_{j=1}^{c^2} \gamma\left(\max(\boldsymbol{l}_{i,j}), k\right)}{c^2} + \boldsymbol{n}_i, \tag{3.2}$$

The noise distribution is assumed to satisfy a Gaussian model [62] as follows:

$$p(\boldsymbol{n}_i) = \frac{1}{\sqrt{(2\pi)^P}} \exp(-\frac{1}{2} \boldsymbol{n}_i{}^T \boldsymbol{n}_i) \tag{3.3}$$

SPM in a Bayesian framework can be achieved by maximizing the posterior distribution of $\{\boldsymbol{l}_{i,j}\}$ given $\{\boldsymbol{x}_i\}$, i.e.,

$$p(\{\boldsymbol{l}_{i,j}\}|\{\boldsymbol{x}_i\}) \propto p(\{\boldsymbol{x}_i\}|\{\boldsymbol{l}_{i,j}\})p(\{\boldsymbol{l}_{i,j}\}) \tag{3.4}$$

Based on the above formulation, main characteristics of proposed method are summarized below.

1. Instead of using traditional methods using MRFs or conditional random fields, DIP is adopted to exploit the spatial correlation in the subpixel label field and to model the $p(\{\boldsymbol{l}_{i,j}\})$, as detailed in Section 3.3.1.
2. Discrete abundances $\{\boldsymbol{s}_i\}$ are obtained from subpixel labels $\{\boldsymbol{l}_{i,j}\}$ and used to reconstruct the coarse pixels $\{\boldsymbol{x}_i\}$.
3. Accurately modeling endmembers is important since the error in endmembers estimation would propagate to the subpixel mapping results. Endmembers $\{\boldsymbol{a}_k\}$ in this chapter are modelled by a purified means approach, as detailed in Section 3.4.4.
4. Characterizing the noise is critical for the modelling of the data likelihood $p(\{\boldsymbol{x}_i\}|\{\boldsymbol{l}_{i,j}\})$, which is modeled by a Gaussian distribution, as detailed in Section 3.3.2.
5. An efficient EM algorithm based optimization scheme is designed and implemented in Section 3.4 for solving the new Bayesian inverse problem.

## 3.3 Bayesian Subpixel Mapping Autoencoder Network

The proposed BSMAN has an encoder-decoder structure. The encoder is a skip-connection FCNN designed for the estimation of subpixel labels $\{\boldsymbol{l}_{i,j}\}$ , where DIP is used to model the spatial correlation of the label field. The encoder maps the input (i.e, coarse HSI $\{\boldsymbol{x}_i\}$) to the subpixel labels $\{\boldsymbol{l}_{i,j}\}$. The decoder has two parts. One part is the forward model which maps the subpixel labels $\{\boldsymbol{l}_{i,j}\}$ to the discrete abundances. The other part of the encoder reconstructs the HSI with class proportions and endmembers extracted from the coarse HSI.

### 3.3.1 Prior of subpixel labels

Following the formulation in Chapter 2, the prior of $\{\boldsymbol{l}_{i,j}\}$ is expressed as,

$$p\left(\boldsymbol{l}_{i,j}\right) = \frac{1}{w}\exp\left(-\left\|\boldsymbol{l}_{i,j} - E\left(\boldsymbol{l}_{i,j}\right)\right\|^2\right) \tag{3.5}$$

where $E(\boldsymbol{l}_{i,j})$ is the expectation of $\boldsymbol{l}_{i,j}$. The prior encourages pixels that are spatially correlated belonging to the same class. By implementing $E(\boldsymbol{l}_{i,j})$ with an FCNN, the DIP

in the HSI can be exploited [16]. Comparing with the patch-based CNN, FCNN has a wider field of view of the input image and enables better modeling of the spatial correlation effect in HSI. Using $f(\cdot)$ to represent the FCNN forward propagation, the expected $\boldsymbol{l}_{i,j}$ is written as:

$$E(\boldsymbol{l}_{i,j}) = f(\boldsymbol{z}_i, \boldsymbol{\beta}). \tag{3.6}$$

where $\boldsymbol{z}_i$ is the input random noise and $\boldsymbol{\beta}$ is the set of model parameters including all convolution kernels and biases. The non-negative and the sum-to-one constraint can be achieved using "softmax" activation approach. In Eq. 3.5, $\boldsymbol{l}_{i,j}$, as the output of the last softmax layer, refers to the soft labels of subpixels instead of the hard labels.

### 3.3.2 Data likelihood

Based on Eq. 3.2 and Eq. 3.3, the data likelihood is formulated as

$$\begin{aligned} & p\left(\boldsymbol{x}_i \mid \boldsymbol{l}_{i,j}\right) \\ & = \frac{1}{z} \exp\left\{ -\frac{1}{2} \left\| \boldsymbol{x}_i - \left( \sum_{k=1}^{K} \boldsymbol{a}_k \frac{\sum_{t=1}^{c^2} \gamma\left(\max(\boldsymbol{l}_{i,t}), k\right)}{c^2} \right) \right\|^2 \right\} \end{aligned} \tag{3.7}$$

where the exponential part is the reconstruction error between the original HSI pixel $\boldsymbol{x}_i$ and the reconstructed one.

## 3.4 Model Optimization

### 3.4.1 MAP estimation

The SPM problem in Eq. 3.4 can be solved by the MAP approach by maximizing the posterior distribution of $\boldsymbol{L}$ given the observed HSI $\boldsymbol{X}$ and the model parameters (i.e., endmembers) $\boldsymbol{A}$ as follows,

$$\hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{L}} \{ p(\boldsymbol{L} \mid \boldsymbol{X}, \boldsymbol{A}) \} \tag{3.8}$$

Maximizing $p(\boldsymbol{L} \mid \boldsymbol{X}, \boldsymbol{A})$ is equivalent to minimizing its negative logarithm likelihood. Then, the objective function can be written as

$$\begin{aligned} J = \arg\min_{\mathbf{L}} & \sum_{i=1}^{N} \sum_{j=1}^{c^2} \left\{ \left\| \boldsymbol{l}_{i,j} - E\left(\boldsymbol{l}_{i,j}\right) \right\|^2 \right\} + \\ \alpha & \sum_{i=1}^{N} \left\{ \left\| \boldsymbol{x}_i - \left( \sum_{k=1}^{K} \boldsymbol{a}_k \frac{\sum_{j=1}^{c^2} \gamma\left(\max(\boldsymbol{l}_{i,j}), k\right)}{c^2} \right) \right\|^2 \right\} \end{aligned} \tag{3.9}$$

where $\alpha$ is a weighting parameter. Both endmembers $\{\boldsymbol{a}_k\}$ and subpixel labels $\{\boldsymbol{l}_{i,j}\}$ are unknown variables, making the SPM problem ill-posed. In the EM algorithm, $\boldsymbol{l}_{i,j}$ is estimated by $E(\boldsymbol{l}_{i,j})$. Replacing $\boldsymbol{l}_{i,j}$ with $E(\boldsymbol{l}_{i,j})$, the objective function can be reformulated as follows [16],

$$J = \arg\min_{\mathbf{L}} \sum_{i=1}^{N} \left\{ \left\| \boldsymbol{x}_i - \left( \sum_{k=1}^{K} \boldsymbol{a}_k \frac{\sum_{j=1}^{c^2} \gamma\left(\max(E(\boldsymbol{l}_{i,j})), k\right)}{c^2} \right) \right\|^2 \right\} \tag{3.10}$$

This objective function has following characteristics:

- The EM algorithm estimates parameters by treating $\{\boldsymbol{l}_{i,j}\}$ as missing observations and $\{\boldsymbol{a}_k\}$ as model parameters, and iteratively updates the estimation of $\{\boldsymbol{l}_{i,j}\}$ and $\{\boldsymbol{a}_k\}$, as illustrated in Section 3.4.2.
- We use $E(\boldsymbol{l}_{i,j})$ as the estimation of $\boldsymbol{l}_{i,j}$, where $E(\boldsymbol{l}_{i,j})$ is modelled by an FCNN. $\{\boldsymbol{x}_i\}$ is used as the input to the FCNN. Once the FCNN is trained, we obtain $\hat{\boldsymbol{l}}_{i,j} = f(\boldsymbol{l}_{i,j})$, as illustrated in Section 3.4.3.
- When estimating parameters in the FCNN for obtaining $E(\boldsymbol{l}_{i,j})$, we use a reconstruction loss based on $||\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i||^2$, which incorporates a forward model to constrain meaningful $\boldsymbol{l}_{i,j}$ estimation, as illustrated in Section 3.4.3.

### 3.4.2 EM Iteration

The EM algorithm optimizes the incomplete data problem by iterating between E-step (i.e., estimation of model parameters given missing observations) and M-step (i.e., estimation of missing observations given the model parameters) [68]. The main steps in EM algorithm to estimate $\{\boldsymbol{l}_{i,j}\}$ and $\{\boldsymbol{a}_k\}$ are summarized as follows.

- Initialization: Set the initial value for $\{\boldsymbol{a}_k\}$. The endmember of each class is manually selected from the coarse HSI.
- E-step: Given endmembers $\{\boldsymbol{a}_k\}$, estimate subpixel labels $\{\boldsymbol{l}_{i,j}\}$ by optimizing an FCNN, as introduced in Section 3.4.3.
- M-step: Given $\{\boldsymbol{l}_{i,j}\}$, estimate endmembers $\{\boldsymbol{a}_k\}$. Endmembers $\{\boldsymbol{a}_k\}$ are estimated with the purified means approach [59], which is presented in Section 3.4.4.

### 3.4.3 Subpixel labels estimation by FCNN training

The objective of E-step of the EM algorithm introduced in Section 3.4.2 is to obtain the estimated subpixel labels $\{\hat{\boldsymbol{l}}_{i,j}\}$ which is achieved by an FCNN model in Eq. 3.6.

To estimate $E(\boldsymbol{l}_{i,j})$, we first need to estimate the parameters in FCNN, i.e., $\boldsymbol{\beta}$. Here, we construct the following objective function to estimate $\boldsymbol{\beta}$:

$$\arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N} \left\{ \left\| \boldsymbol{x}_i - \left( \sum_{k=1}^{K} \boldsymbol{a}_k \frac{\sum_{j=1}^{c^2} \gamma\left(\max(f(\boldsymbol{x}_i, \boldsymbol{\beta})), k\right)}{c^2} \right) \right\|^2 \right\} \tag{3.11}$$

Adam stochastic optimizer [69] is adopted to estimate $\boldsymbol{\beta}$.

### 3.4.4 Endmembers update by purified means

The endmember update in the M-step of the iterative EM algorithm introduced in Section 3.4.2 is achieved by the purified means approach, where the purified pixel $\boldsymbol{y}_i^k$ is represented as [59]:

$$\boldsymbol{y}_i^k = \boldsymbol{x}_i^k - \sum_{q \neq k}^{K} \frac{\sum_{j=1}^{c^2} \gamma\left(\max(\boldsymbol{l}_{i,j}), q\right)}{c^2} \boldsymbol{a}_q \tag{3.12}$$

The endmember matrix $\{\hat{\boldsymbol{a}}_k\}$ can be iteratively updated using their conditional expectation, which was introduced in Chapter 2, Eq. 2.27

### 3.4.5 Summary of Complete Algorithm

Based on the EM steps described in Section 3.4.2, the complete algorithm used for solving BSMAN can be achieved, which is summarized in Algorithm 1.

## 3.5 Experiments

### 3.5.1 BSMAN implementation

We use the coarse HSI as the input of the FCNN, and the output is the soft subpixel labels. The forward model is implemented with the average pooling to map soft subpixel labels to discrete abundance of coarse pixels, and a linear combination to reconstruct the HSI using discrete abundances and endmembers. The implementation of BSMAN is illustrated in Figure 3.2. The FCNN is implemented with a U-Net type "hourglass" architecture with skip-connection [16] to model a mapping from the input to soft labels of subpixels.

**Algorithm 2** BSMAN

**Input**: HSI $\boldsymbol{X}$, numbers of classes $K$, and iteration numbers $\tau$
**Output**: Class labels of subpixels $\{\boldsymbol{l}_{i,j}\}$
**Initialization**: $t := 1$, $\{\boldsymbol{a}_k^{(0)}\}$

**While** $t \leqslant \tau$ **do**
  E-step:
    estimate $\boldsymbol{\beta}$ in FCNN,
    estimate $\{\boldsymbol{l}_{i,j}\} \leftarrow fcnn(\{\boldsymbol{x}_i\}, \{\boldsymbol{a}_k\})$ in Eq. 3.6
  M-step:
    **for** k=1,2,...,K
      estimate $\{\boldsymbol{y}_i^k\}$ according to Eq. 3.12
      estimate $\boldsymbol{a}_k$ using $\{\boldsymbol{y}_i^k\}$
    **end for**
**end while**

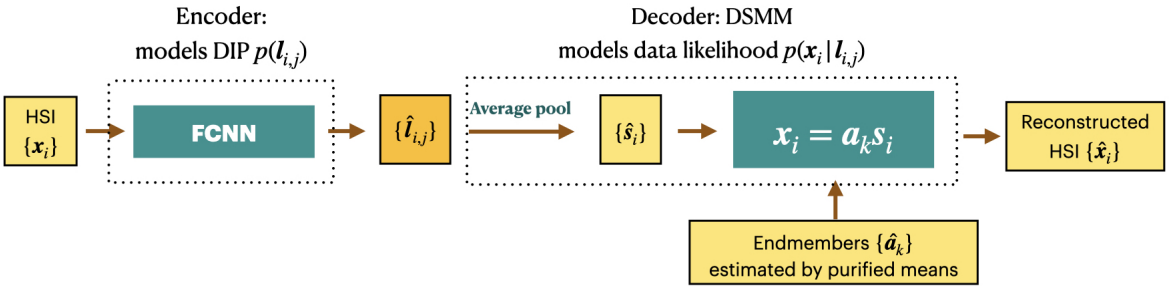



Figure 3.2: Subpixel mapping framework. The encoder is a skip-connection FCNN designed for the estimation of soft labels of subpixels $\{\boldsymbol{l}_{i,j}\}$, where DIP is used to model the spatial correlation of the label field. The decoder is the forward DSMM model which contains two parts. One part is the forward downsampling which maps soft labels of subpixels to the class proportions $\boldsymbol{S}$. The other part reconstructs the HSI $\hat{\boldsymbol{X}}$ with $\boldsymbol{s}_i$ and endmembers $\boldsymbol{A}$ extracted from the HSI $\boldsymbol{X}$.

### 3.5.2 Datasets and pre-processing

Six images are used to evaluated the proposed BSMAN, including three simulated HSIs, two real HSIs and one time-series multispectral images.

#### Simulated HSIs

In this experiment, a $78 \times 78$ sized fine-resolution HSI is simulated with four endmembers with 200 spectral bands. Each pixel in the simulated HSI is a mixture of four endmembers shown in Figure 3.4. Pixels are created using the four endmembers multiplied by four abundance maps following a linear spectral mixture model. Abundance maps are generated by first being divided into $8 \times 8$ sized homogeneous blocks of one of the four endmembers, then degrading the blocks by applying a spatial low-pass filter with the size of of $9 \times 9$ pixels. The resulting HSI is further degraded by zero-mean Gaussian noise (SNR = 40dB). The label of each pixel is determined by the dominant endmember in the coarse pixel. The coarse HSIs at three degradation scales (i.e., 2, 3 and 4) are obtained by applying mean spatial filters with the kernal sizes of 2×2, 3×3 and 4×4 pixels separately to down-sample the fine HSI. As a result, three HSIs with the sizes of 39×39, 26×26, and 19×19 are simulated.

#### Jasper Ridge HSI dataset

Jasper Ridge is a HSI dataset with $100 \times 100$ pixels, recorded at 198 channels ranging from 0.38 $\mu m$ to 2.5 $\mu m$ as introduced in Chapter 2. There are four endmembers in this data, i.e., "road", "soil", "water" and "tree". The label of each subpixel is determined by its dominant endmember. The RGB color composition of the HSI has been shown in Chapter 2, Figure 2.3 (a). The input coarse HSI with the size of 50×50 is obtained by applying a 2×2 mean filter to the original HSI.

#### Saint Andre HSI dataset

The Saint Andre HSI contains $50 \times 50$ pixels, composed of 415 spectral bands ranging from 0.40 to 2.40 $\mu m$ as introduced in Chapter 2. Endmember spectra of six distinct materials (i.e., "tree", "grass", "soil", "road", "building 1", and "building 2") have been manually extracted based on prior knowledge of the scene [74, 31]. The label of each subpixel is determined by its dominant endmember. The RGB color composition of the HSI has been shown in Chapter 2, Figure 2.3 (b). The input coarse HSI with the size of 25×25 is obtained by applying a 2×2 mean filter to the original HSI.

#### Time-series Landsat imagery

A HSI-like image was obtained by combining 11 time-series Landsat-8 images over Gulin City, Sichuan, China from 2013 to 2017 with cloud cover lower than 10%. From the image, we identify a subarea, where the land cover types stayed constant during the time series. By removing the panchromatic bands and two damaged bands, the size of the subarea dataset is 40 × 40 × 108. The dataset was used in a previous study on the HSIs classification [29]. The input coarse image with the size of 20×20×108 is obtained by applying a 2×2 mean filter to the original one. There are five land covers types in this area, i.e., road, residential area, forestland, terrace, and farmland. The groundtruth label map is obtained by reference the high-resolution UAV image shown in Figure 3.5 (b).

### 3.5.3 Experimental Setup

#### Methods Compared

The proposed method is compared with several traditional subpixel mapping methods including GAAI[89], SPM_LM[97], SASM[84], PPSM[98], RBF[86] and SPMSS[99].

#### Numerical Measures

Three kinds of accuracy indices are used to evaluate the quantitative performance of SPM algorithms on real HSIs: the accuracy of each class, the overall accuracy (OA), and the Kappa coefficient (Kappa). Only the last two indices (i.e., OA and Kappa) are used in the simulated study.

#### Parameter settings

Hyperparameters including the learning rate, iteration times and the number of skip connection layers are determined empirically, which are recorded in Table 3.1. Hyperparameters are required to be adjust for different HSIs since HSIs vary in the sensor, data size and complexity. All data processing is conducted using the Python language under the Pytorch framework.

### 3.5.4 Simulated Study

Figure 3.6, Figure 3.7 and Figure 3.8 show visual comparison between the proposed method with other methods as three downsampling-scale factors. The proposed BSMAN method always generate subpixel label maps with the richest spatial texture details and the least noise. Besides, BSMAN, GAAI and SPM_LM perform better than the rest methods (i.e.,

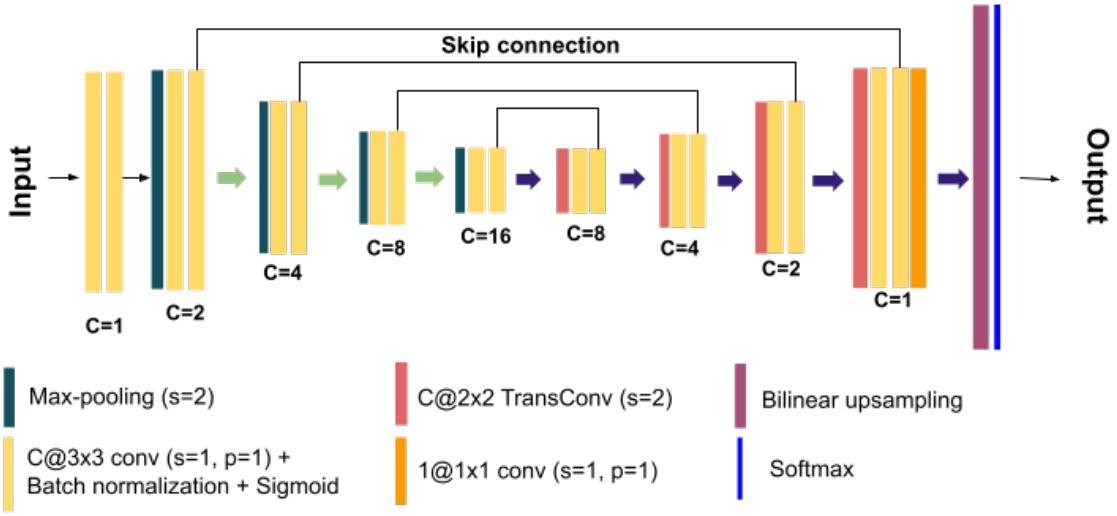
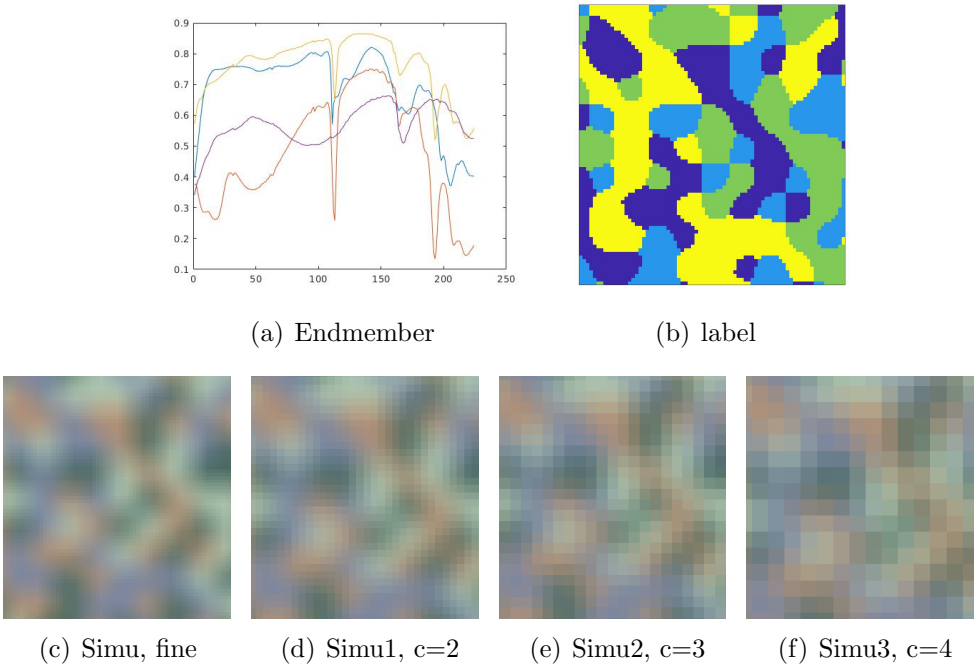



Figure 3.3: FCNN structure in BSMAN framework. $C$ represents the number of convolution channels.

Table 3.1: Hyperparameters Setting.

| HSI | HSI Size | Scale | LR | Epochs | EM | Skips |
|---|---|---|---|---|---|---|
| Simu 1 | $39 \times 39 \times 200$ | 2 | 0.0001 | 20 | 70 | 4 |
| Simu 2 | $26 \times 26 \times 200$ | 3 | 0.0001 | 20 | 100 | 4 |
| Simu 3 | $19 \times 19 \times 200$ | 4 | 0.0001 | 10 | 100 | 4 |
| Jasper | $50 \times 50 \times 224$ | 2 | 0.0002 | 200 | 50 | 4 |
| Saint | $25 \times 25 \times 415$ | 2 | 0.001 | 300 | 20 | 3 |
| Landsat | $20 \times 20 \times 108$ | 2 | 0.0001 | 100 | 10 | 3 |

Simu 1, 2 and 3 refer to simulated HSIs by degrading the original HSI with $c = 2, 3, 4$.
The fourth column is the learning rate for FCNN training in E-step.
The fifth column is the number of epochs for FCNN training in E-step.
The sixth column is the number of EM iteration.
The seventh column is the number of the skip-connections in the FCNN, determined by the network depth.



(a) Endmember (b) label

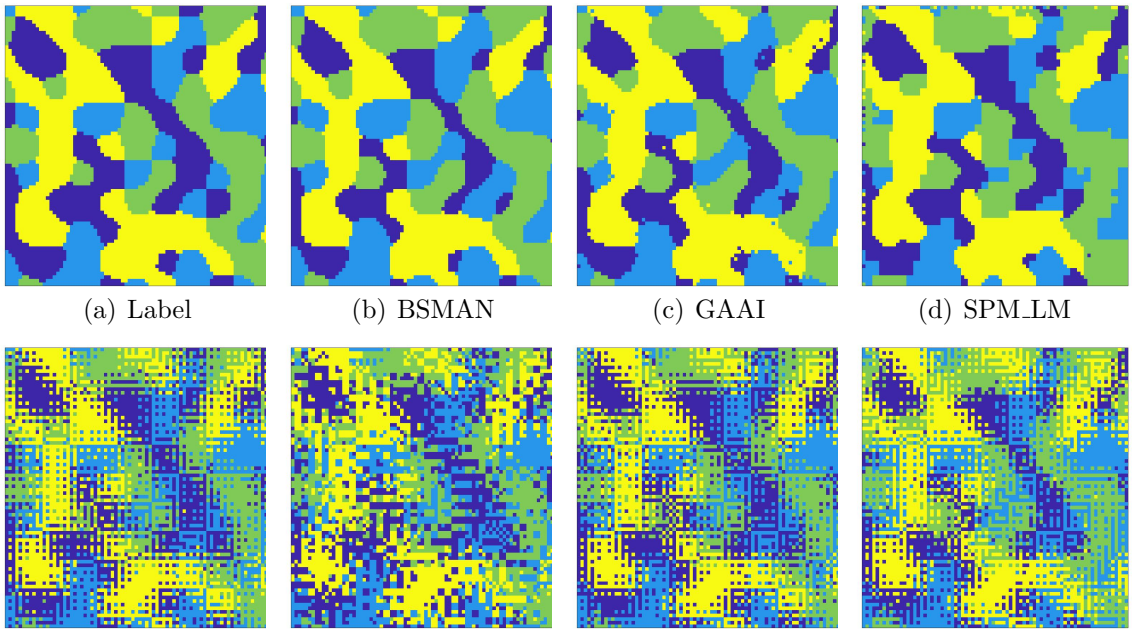
(c) Simu, fine (d) Simu1, c=2 (e) Simu2, c=3 (f) Simu3, c=4

Figure 3.4: False color images of simulated HSIs.

SASM, PSSM, RBF and SPMSS) by delineating the class boundary more accurately and with less noisy. SASM, PPSM, RBF are based on the concept of spatial dependence, assuming that pixels spatially adjacent are more likely to be in the same class, leading to the blocking effect in the label maps. GAAI shows less blocking effect because it corrects the potential errors in the estimated abundance maps and achieved a global optimization. SPM_LM minimizes an objective function using the spectral term to regulate the spatial term by combining a linear unmixing model and a maximum spatial dependence model. Although SPMSS considers both spatial and spectral correlation, it highly relies on the bilinear and bicubic interpolation results, leading to the blocking effect as well. The numerical indices summarized in Table 3.2 show consistent results with the visual demonstration. The OA and Kappa of the classification results obtained by the proposed BSMAN are the highest among all tested methods at scale factors are two and three. When $c = 4$, although BSMAN gives the richest spatial detail information in Figure 3.8 (b), the performance of BSMAN in terms of OA and Kappa rank the second after SPM_LM due to inaccurate label prediction near the edge of the scene. This is caused by the convolution operations on small-size images.

(a) True color Landsat-8 image (b) The UAV image

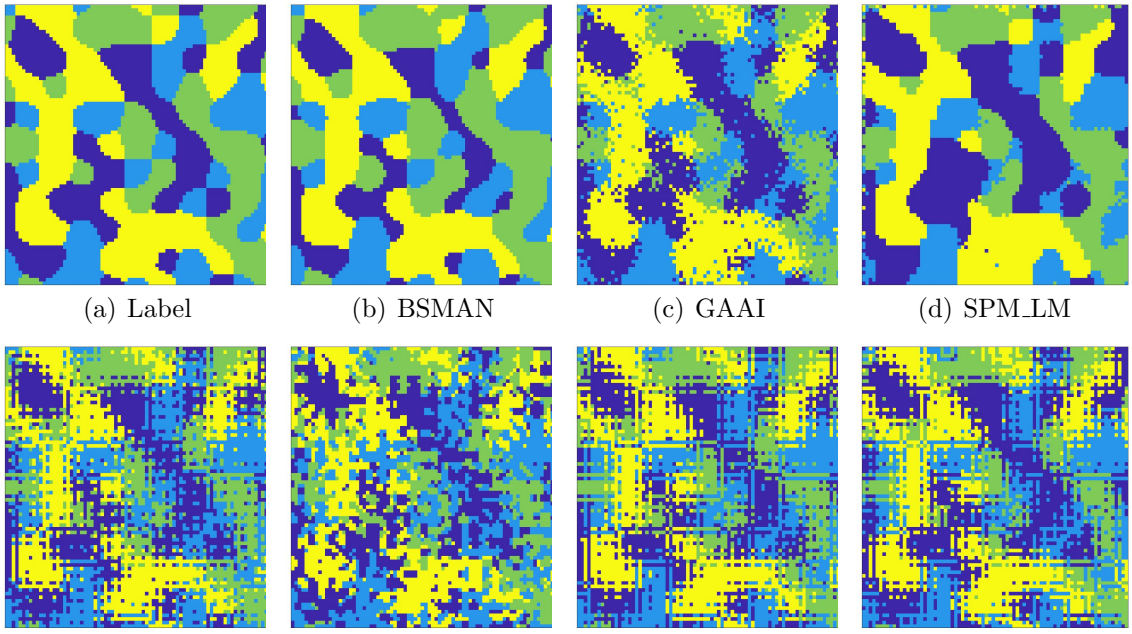
Figure 3.5: True color images of the time-series Landsat scene. (a) Low-resolution (30m) true color Landsat-8 imagery; (b) the high resolution (0.2m) UAV imagery that helps the acquisition of the groundtruth map.



(a) Label (b) BSMAN (c) GAAI (d) SPM_LM

(e) SASM (f) PSSM (g) RBF (h) SPSMM

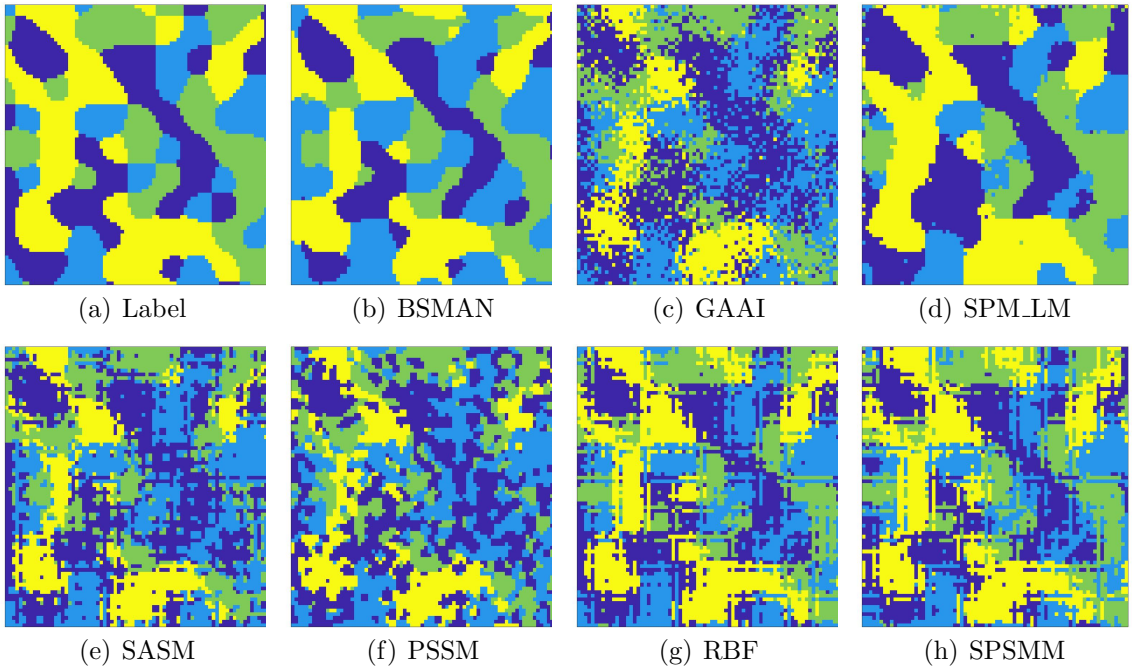
Figure 3.6: Subpixel mapping results (subpixel label maps and their OA) of different methods on the simulate dataset with $c = 2$. (a) Label. (b) BSMAN. (c)GAAI. (d) SPM_LM. (e) SASM. (f) PSSM. (g) RBF. (h) SPSMM.



(a) Label (b) BSMAN (c) GAAI (d) SPM_LM

(e) SASM (f) PSSM (g) RBF (h) SPSMM

Figure 3.7: Subpixel mapping results (subpixel label maps and their OA) of different methods on the simulate dataset with $c = 3$. (a) Label. (b) BSMAN. (c)GAAI. (d) SPM_LM. (e) SASM. (f) PSSM. (g) RBF. (h) SPSMM.



(a) Label (b) BSMAN (c) GAAI (d) SPM_LM

(e) SASM (f) PSSM (g) RBF (h) SPSMM

Figure 3.8: Subpixel mapping results (subpixel label maps and their OA) of different methods on the simulate dataset with $c = 4$. (a) Label. (b) BSMAN. (c)GAAI. (d) SPM_LM. (e) SASM. (f) PSSM. (g) RBF. (h) SPSMM.

Table 3.2: Simulated HSI data: Kappa Coefficient and Overall Accuracy.

| Methods | c = 2 | | c = 3 | | c = 4 | |
|---|---|---|---|---|---|---|
| | Kappa | OA(%) | Kappa | OA(%) | Kappa | OA(%) |
| **SPMSS** | 0.529 | 64.7 | 0.586 | 69.0 | 0.628 | 72.2 |
| **RBF** | 0.552 | 66.5 | 0.586 | 69.0 | 0.650 | 73.8 |
| **PSSM** | 0.477 | 60.9 | 0.505 | 62.9 | 0.419 | 56.2 |
| **SASM** | 0.552 | 66.5 | 0.592 | 69.4 | 0.493 | 61.7 |
| **SPM_LM** | 0.811 | 85.9 | 0.798 | 84.9 | **0.782** | **83.8** |
| **GAAI** | 0.856 | 89.2 | 0.750 | 81.3 | 0.445 | 58.1 |
| **DIP** | **0.870** | **90.2** | **0.868** | **90.1** | 0.707 | 77.9 |

### 3.5.5 Test on real HSIs

#### Test on Jasper Ridge HSI

Figure 3.9 shows the SPM results for the Jasper HSI scene obtained by different SPM methods. The subpixel label map obtained by the BSMAN methods shows the most detailed spatial textural information and achieves the highest numerical accuracy shown in Table 3.3. Table 3.4 shows the individual classification accuracy achieved by different methods on the HSI. The proposed BSMAN performs the best on the class "water", "tree", and "soil". Although the "road" class accuracy is relatively low, the outline of the road is still clearly visible in Figure 3.9. We attribute the false negative pixels for road class to the smoothness property of convolutions.

Table 3.3: Jasper Ridge HSI data: Kappa Coefficient and Overall Accuracy.

| Methods | Kappa | OA(%) |
|---|---|---|
| **SPMSS** | 0.315 | 50.2 |
| **RBF** | 0.547 | 66.2 |
| **PSSM** | 0.550 | 67.9 |
| **SASM** | 0.581 | 70.2 |
| **SPM_LM** | 0.794 | 85.5 |
| **GAAI** | 0.796 | 85.8 |
| **BSMAN** | **0.808** | **86.8** |

Table 3.4: Jasper Ridge HSI data: Individual class accuracies (%).

| **Methods** | **Water** | **Road** | **Tree** | **Soil** |
|---|---|---|---|---|
| **SPMSS** | 64.9 | 41.4 | 45.7 | 39.13 |
| **RBF** | 88.4 | **71.7** | 55.6 | 48.7 |
| **PSSM** | 91.2 | 53.2 | 58.6 | 54.0 |
| **SASM** | 92.0 | 59.3 | 60.3 | 57.8 |
| **SPM_LM** | 99.9 | 65.7 | 84.8 | 73.1 |
| **GAAI** | 99.9 | 64.2 | 87.1 | 71.2 |
| **BSMAN** | **100.0** | 23.2 | **90.5** | **83.0** |

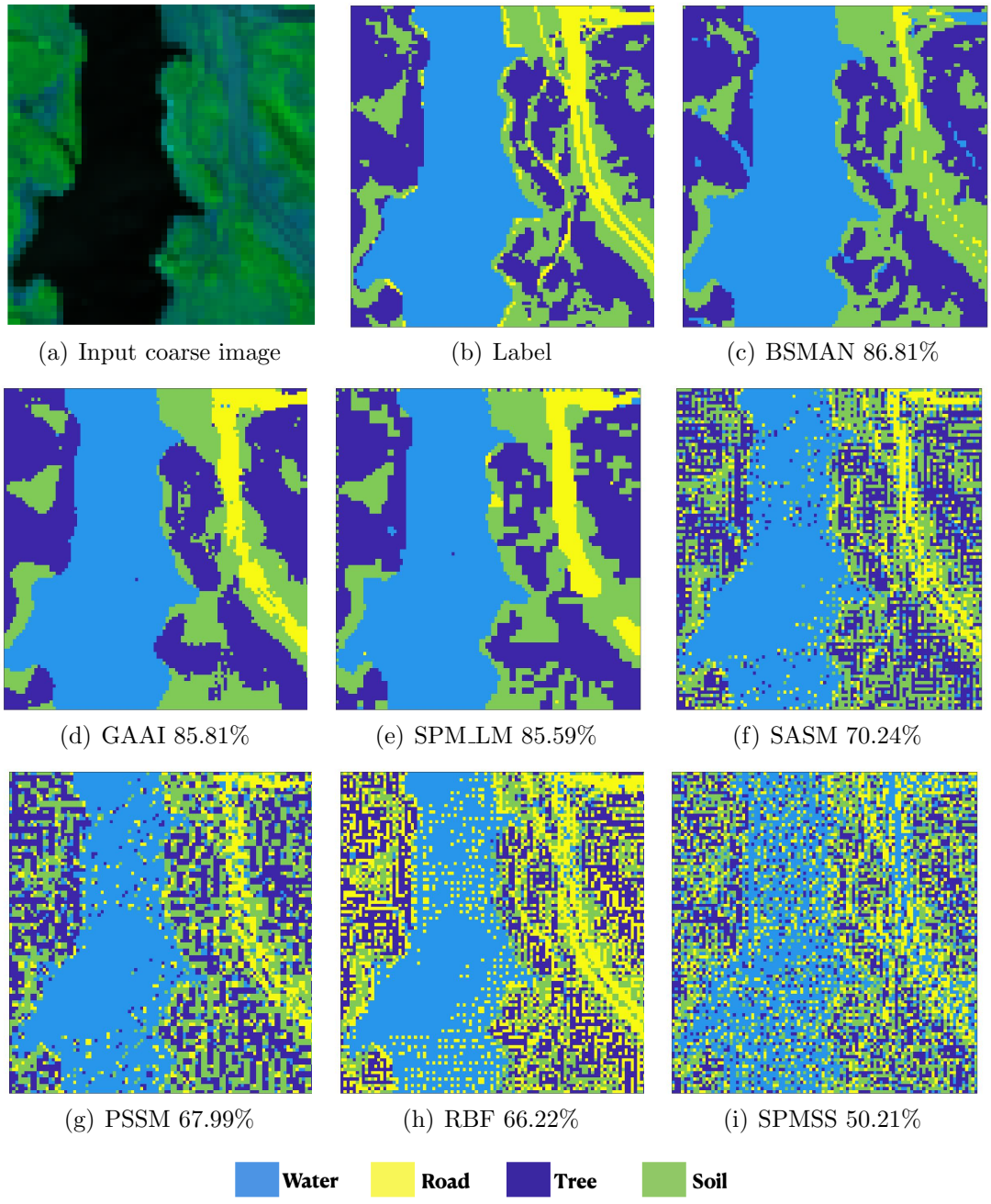



Figure 3.9: Subpixel mapping results (subpixel label maps and their OAs) of different methods on Japser dataset.

### Test on Saint Andre HSI

Figure 3.10 shows the SPM results for the Saint HSI scene obtained by different SPM methods. The subpixel label map obtained by the BSMAN methods preserve the richest the spatial textural information and achieved the highest OA and Kappa shown in Table

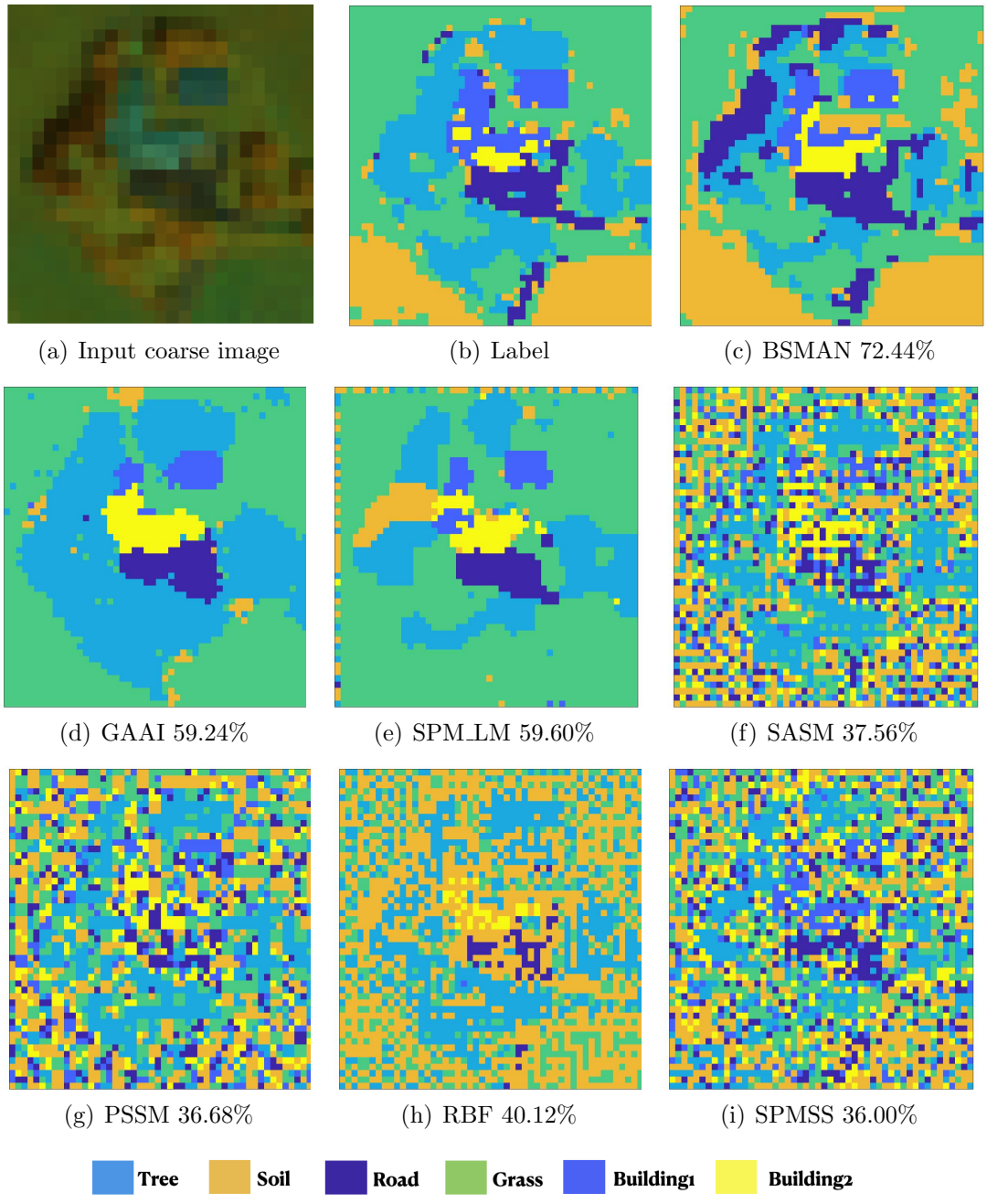
3.5. The Saint scene turns out to be more challenging than the Jasper scene given the all methods on this scene obtains relatively lower SPM accuracy and more noisy label maps. We attribute it to the similar land cover types and higher mixed pixels in the HSI. The class of soil, grass and tree are commonly mixed in real scenarios, especially for soil and grass at the bottom of this HSI scene. BSMAN successfully identified the dominant class in the HSI and restored the label map illustrated in Figure 3.10 (b), indicating that BSMAN is more capable to deal with highly-mixed HSI pixels. For individual classes, BSMAN achieves the highest individual class accuracies on classes of soil, road and building2 displayed in Table. 3.6. However, the accuracy for tree is relatively low because BSMAN treats the shadow part of trees as building1 (see Figure 3.10 (b)).

Table 3.5: Saint HSI data: Kappa Coefficient and Overall Accuracy.

| **Methods** | **Kappa** | **OA(%)** |
|---|---|---|
| **SPMSS** | 0.195 | 36.0 |
| **RBF** | 0.207 | 40.1 |
| **PSSM** | 0.191 | 36.6 |
| **SASM** | 0.204 | 37.5 |
| **SPM_LM** | 0.381 | 59.6 |
| **GAAI** | 0.407 | 59.2 |
| **BSMAN** | **0.623** | **72.4** |

Table 3.6: Saint HSI data: Individual class accuracies (%) (the highest accuracy in each row is in bold format.

| **Methods** | **Tree** | **Soil** | **Road** | **Grass** | **Building1** | **Building2** |
|---|---|---|---|---|---|---|
| **SPMSS** | 41.0 | 45.7 | 48.4 | 27.6 | 40.2 | 11.1 |
| **RBF** | 64.9 | 59.4 | 27.3 | 25.4 | 0.0 | 55.5 |
| **PSSM** | 58.4 | 35.4 | 27.9 | 28.1 | 27.0 | 44.4 |
| **SASM** | 59.7 | 36.0 | 34.1 | 27.9 | 24.3 | 63.8 |
| **SPM_LM** | 61.3 | 1.0 | 54.6 | **85.0** | 51.3 | 75.0 |
| **GAAI** | **96.3** | 0.4 | 54.0 | 66.7 | 43.0 | **94.4** |
| **BSMAN** | 53.5 | **84.8** | **84.4** | 74.7 | **75.0** | 77.7 |




(a) Input coarse image (b) Label (c) BSMAN 72.44%

(d) GAAI 59.24% (e) SPM_LM 59.60% (f) SASM 37.56%

(g) PSSM 36.68% (h) RBF 40.12% (i) SPMSS 36.00%

Figure 3.10: Subpixel mapping results (subpixel label maps and their OAs) of different methods on Saint dataset.

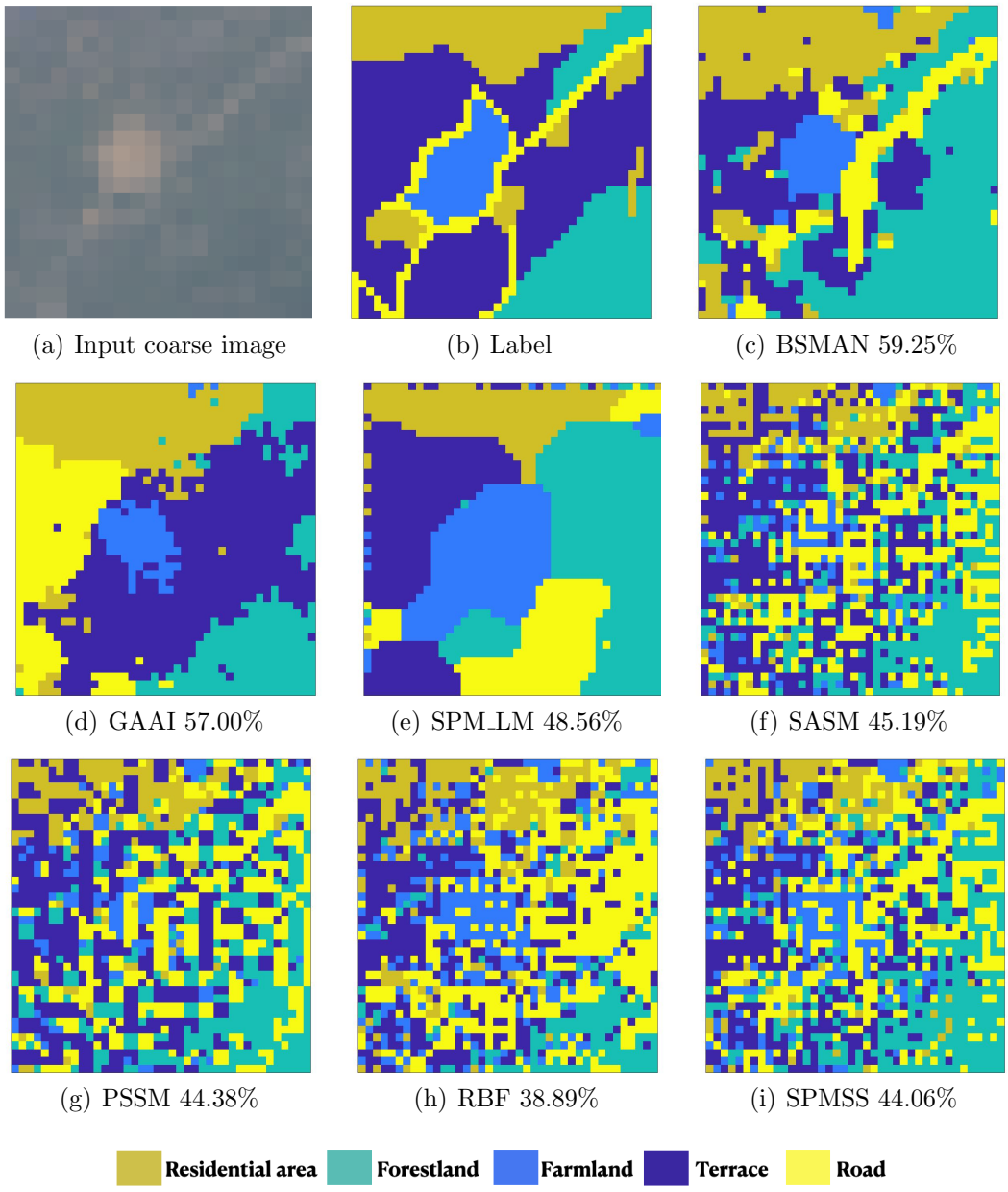



Figure 3.11: Subpixel mapping results (subpixel label maps and their OAs) of different methods on Landsat Time Series dataset.

### Test on Time-series Landsat imagery

Figure 3.11 shows the SPM results for Landsat dataset obtained by different SPM methods. The subpixel label map obtained by the BSMAN methods preserve the richest the spatial textural information and achieved the highest OA and Kappa coefficient. The Landsat dataset is more challenging to do SPM than the two HSIs because of its lower spatial

resolution, more mixed pixels, fewer spectral bands and coarser spectral resolution. As a result, the OA accuarcies are much lower than the other two datasets. However, the proposed BSMAN still shows big potential on mapping multispectral satellite imagery by achieving the highest SPM accuracies (see Table 3.7) than all other methods. BSMAN achieved the highest accuracy for the individual classes of farmland and terrace. Although the accuracy for class road of BSMAN is not the highest, its advantage of recognizing linear objects from low resolution images (see Figure 3.11(d)). Although SPMSS also identifies the road very well, it gives fragmentary classification generally on all classes. Unlike GAAI and SPM_LM under-segmenting classes or SASM, PSSM and SPSMM over-segmenting classes, BSMAN strikes a better balance between clustering large land regions smoothly and preserving the boundary features.

Table 3.7: Landsat Time Series data: Kappa Coefficient and Overall Accuracy.

| Methods | Kappa | OA (%) |
|---|---|---|
| **SPMSS** | 0.285 | 44.0 |
| **RBF** | 0.237 | 38.8 |
| **PSSM** | 0.263 | 44.3 |
| **SASM** | 0.275 | 45.1 |
| **SPM_LM** | 0.343 | 48.5 |
| **GAAI** | 0.404 | 57.0 |
| **BSMN** | **0.466** | **59.2** |

Table 3.8: Landsat Time Series data: Individual class accuracies (%) (the highest accuracy in each row is in bold format).

| Methods | Residential area | Forestland | Farmland | Terrace | Road |
|---|---|---|---|---|---|
| **SPMSS** | 36.3 | 41.3 | 65.7 | 38.2 | **54.8** |
| **RBF** | 32.3 | 48.2 | 43.0 | 45.8 | 31.5 |
| **PSSM** | 43.8 | 24.1 | 58.6 | 40.1 | 43.6 |
| **SASM** | 44.4 | 25.0 | 59.3 | 38.8 | 50.3 |
| **SPM_LM** | 38.9 | **95.6** | 58.9 | 60.5 | 9.7 |
| **GAAI** | **52.7** | 47.4 | 73.2 | **72.6** | 16.5 |
| **BSMAN** | 40.5 | 61.2 | **96.2** | **72.6** | 48.1 |

## 3.6 Conclusion

In this chapter, we presented a Bayesian subpixel mapping autoencoder network for HSIs. An encoder-decoder architecture was designed to incorporate the FCNN with DIP prior

and the forward models to effectively estimate the subpixel labels by learning from both data and prior knowledge. BSMAN adopted an FCNN to exploit the spatial correlation effect in the subpixel label field. An efficient purified means approach was adopted to the SPM framework for the endmember estimation. The resulting Bayesian MAP framework is solved by the proposed EM approach. The proposed approach was tested on both real and simulated HSIs, in comparison with several other SPM methods. The proposed BSMAN method was demonstrated effective for SPM with more accurate SPM results than other methods.

# Chapter 4

# Unsupervised Bayesian Deep Image Prior Downscaling for High-resolution Soil Moisture Estimation

Soil moisture (SM) estimation is a critical part of environmental and agricultural monitoring, with satellite-based microwave remote sensing being the main SM source. However, the limited spatial resolution of most current remote sensing SM products reduces their utility for many applications such as evapotranspiration modeling and agriculture management. To address this issue, we propose a Bayesian deep image prior (BDIP) downscaling approach to estimate the high-resolution SM from satellite products. More specifically, the high-resolution soil moisture estimation problem is formulated as a maximum a posteriori (MAP) problem, and solved via a neural network comprising of a deep fully convolutional neural network (FCNN) for modeling the prior spatial correlation distribution of the underlying high-resolution SM variables, and a forward model characterizing the SM map degeneration process for modeling the data likelihood. As such, the proposed BDIP approach provides a statistical framework that integrates deep learning with forward modelling in a coherent manner for combining different sources of information, i.e., the knowledge in the forward model, the spatial correlation prior in FCNN architecture, and the remote sensing data and products. Experiments on the downscaling of Soil Moisture Active Passive SM products using the Moderate Resolution Imaging Spectroradiometer products show that SM maps estimated using the proposed method provide greater spatial detail information than other downscaling methods, with the SM estimates very close to in-situ measurements.

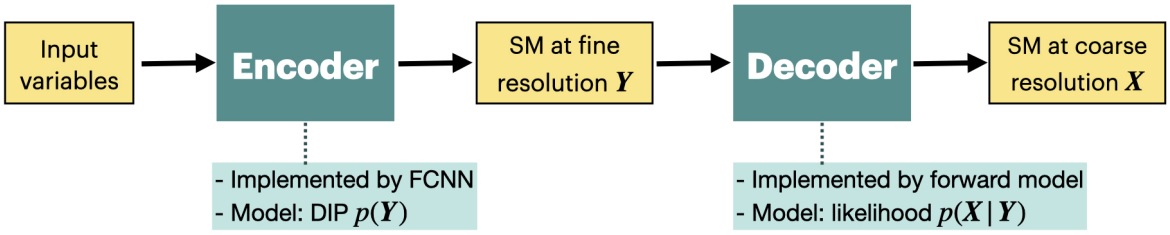



Figure 4.1: Bayesian deep image prior model for SM downscaling. The encoder is implemented as a FCNN accounting for the spatial correlation prior of the high-resolution SM, and the decoder part works as the forward model.

## 4.1 Introduction

Soil moisture (SM) highly influences hydrologic and atmospheric processes for environmental and agricultural monitoring. Microwave remote sensing (RS), with the high sensitivity to the SM variation and robustness to atmosphere conditions, is the most commonly used approach to monitor SM [100, 101, 102]. The Soil Moisture Active/Passive (SMAP) mission has been providing soil moisture at two spatial resolutions of 36 and 9km since April 2015 [103]. However, these two spatial resolutions do not meet the requirements for application to evapotranspiration modeling and agriculture management [7, 103]. Therefore, improving the spatial resolution of the SMAP SM product to 1km spatial resolution is essential. Downscaling is an inverse problem that reconstructs images at higher resolution from coarse observations. Since SM has high variation over spatial scales smaller than the SMAP resolutions, spatial heterogeneity must be properly addressed when downscaling [102].

The SM downscaling can be achieved by different strategies, e.g., data fusion or assimilation [104, 105, 106, 107], geostatistical [102, 108], traditional regression [109] and machine learning (ML) [110]. Data fusion and assimilation for the downscaling is achieved by combing multi-sources data and extracting more accurate spatial SM information. Geostatistical methods interpolate the SM product with geographical models based on the certain spatial assumption, e.g., the geographically weighted regression [102]. The traditional regression method uses a simple regression model, e.g., a linear regression model, to analysis the correlation between the SM and other RS products [109], which could not sufficiently explore the complex relationship between them. ML methods (e.g., the decision tree regression [100]) show stronger potential in SM downscaling by building the nonlinear relationships between the SM and other indices. Recently, deep learning using multi-layer perceptron (MLP) has been adopted to SM downscaling due to its capability in learning complex relationships between inputs (i.e., the coarse-resolution SM and fine-resolution ancillary products) and the target data (i.e., the fine-resolution SM), and its short inference time after training [103]. This downscaling model is trained using SM products with different resolutions, and then can be used for improving the spatial resolution of SM products by the same scale as training. However, the scale of SM products to reduce is limited by

the scale difference between the two SM products used for training. In addition, MLP cannot effectively model the spatial correlation of the SM. A convolutional neural network (CNN)-based downscaling method [111] is proposed recentlywhich can better exploit the spatial information within adjacent pixels. However, most of the ML-, MLP- and CNN-based methods are supervised, requiring the groundtruth data which could be in-situ measurements or the high-resolution SM products, and as such the performance is greatly dependent on the training dataset. [112].

Fully CNNs (FCNNs) have been widely used in various tasks including semantic segmentation [51, 52], super-resolution [113] and image denoising [54], as examples. Unlike the classic CNN, the FCNN contains no fully-connected layers and it can take input of arbitrary size. Ulyanov et al. [16] demonstrate that the structure of a FCNN is sufficient to capture low-level image statistics [16], which is called "deep image prior (DIP)" . The FCNN can capture appropriate global spatial features [35] with a wide image field-of-view compared to a CNN layer [17]. Also, compared to traditional methods [114, 102, 109], FCNN is much more computationally efficient by leveraging GPUs. In addition, many of downscaling methods assume a linear scaling relationship between optical-derived input variables and SM, which is not always satisfied [115]. Accounting for nonlinearities between SM and the input variables, FCNN, as an empirical method, has more potential compared to traditional models [115].

We integrate the DIP captured by FCNN into a Bayesian framework to address the SM downscaling inverse problem. Then the resulting downscaling model becomes a Bayesian deep image prior (BDIP) downscaling network, where the inverse model is implemented by a FCNN accounting for DIP, the forward model is modelled by a downsampler describing the relationship between low- to high-resolution SM map.

Contributions of this chapter are summarized as follows:

1. We adopt a BDIP scheme to SM downscaling to account for the spatial heterogeneity in higher resolution SM maps.

2. The forward model describing the spatial resolution decreasing process from high- to low-resolution SM map is integrated into the Bayesian framework to solve the inverse problem.

3. The resulting maximum a priori (MAP) problem is solved by the back-propagation instead of using the typical expectation-maximization iterative method, which makes the model optimization simple and effective.

4. The proposed method reconstructs the SM in high spatial resolution only by extracting information from high-resolution RS products and the low-resolution SM using DIP, without requiring any ground-truth data for model training.

The proposed method is designed to effectively downscale SMAP SM products at 9km spatial resolution to 1km resolution, which can facilitate the generation of 1km SM maps

using coarse SM product and some ancillary data, and thereby can enhance the hydrological monitoring in the study area by offering more spatially-detailed hydrological information of the study area. The method is evaluated qualitatively and quantitatively, and results demonstrate that the proposed approach achieves new state-of-the-art results compared to other unsupervised methods.

## 4.2 Problem formulation

We assume that the RS product with low spatial resolution is $\boldsymbol{X} = \{\boldsymbol{x}_i | i = 1, 2, ..., m \times n\}$, and the RS product with high spatial resolution is $\boldsymbol{Y} = \{\boldsymbol{y}_i | i = 1, 2, ..., \alpha^2 \times m \times n\}$, where $\alpha$ is the ratio between the low- and high-spatial resolutions. Given the forward mapping $\Phi(\cdot)$ from $\boldsymbol{Y}$ to $\boldsymbol{X}$, the low resolution image $\boldsymbol{X}$ can be represented as follows,

$$\boldsymbol{X} = \Phi(\boldsymbol{Y}) + \boldsymbol{N} \tag{4.1}$$

where $\boldsymbol{N} \in \mathbb{R}^{m \times n}$ is the noise matrix.

The RS product downscaling aims to infer the high-resolution image $\boldsymbol{Y}$ based on the observed low-resolution image $\boldsymbol{X}$, which in a Bayesian framework can be achieved by maximizing the posterior distribution $p(\boldsymbol{Y}|\boldsymbol{X})$, i.e.,

$$p(\boldsymbol{Y}|\boldsymbol{X}) \propto p(\boldsymbol{X}|\boldsymbol{Y})p(\boldsymbol{Y}) \tag{4.2}$$

Given the generative model $g(\cdot)$ of $\boldsymbol{X}$ in Eq. 4.1 and the posterior distribution in Eq. 4.2, several key factors for effective downscaling are identified as follows:

1) The effective modelling of the high-resolution image prior $p(\boldsymbol{Y})$ is critical for regulating and estimating the high-resolution image $\boldsymbol{Y}$.

2) Meaningful modelling the data likelihood $p(\boldsymbol{X}|\boldsymbol{Y})$ is essential for guiding and regulating the downscaling process.

3) An efficient optimization scheme for solving the Bayesian inverse problem is necessary.

In this chapter, $p(\boldsymbol{Y})$ is achieved by the DIP approach using FCNN, as detailed in Section 4.2.1. The data likelihood $p(\boldsymbol{X}|\boldsymbol{Y})$ is modelled by a distribution incorporating the forward model, as detailed in Section 4.2.2. An efficient optimization scheme is designed and implemented in Section 4.3. The designed Bayesian DIP downscaling model is illustrated in Figure 4.1.

### 4.2.1 Prior of the high-resolution SM map

There are three key requirements on the high-resolution SM $\boldsymbol{Y}$ when designing the prior $p(\boldsymbol{Y})$.

1) The large-scale heterogeneous spatial correlation effect in SM map should be fully exploited.

2) SM should be in the meaningful value range of $[0, 1]$

3) High-resolution SM prior should allow efficient optimization.

Here, we represent the prior over the high-resolution SM $\boldsymbol{Y}$ by a distribution expressed as,

$$p(\boldsymbol{Y}) = \frac{1}{z} exp(-\delta(\boldsymbol{Y}, E(\boldsymbol{Y}))) \tag{4.3}$$

where $E(\boldsymbol{Y})$ is the expectation of $\boldsymbol{Y}$, which is implemented as a FCNN, and $\delta(\boldsymbol{u}, \boldsymbol{v})$ is the distance function measuring the distance between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$.

The prior spatial information of $\boldsymbol{Y}$ can be captured by an FCNN structure [16] which has a wide field of view of the input image compared to patch-based CNN and can be optimized efficiently on GPUs. Using $f(\cdot)$ to represent the FCNN forward propagation, the expected $\boldsymbol{Y}$ is written as:

$$E(\boldsymbol{Y}) = f(\boldsymbol{Z}, \boldsymbol{\beta}). \tag{4.4}$$

where $\boldsymbol{Z}$ is the input random noise and $\boldsymbol{\beta}$ is the set of model parameters including all weights of convolution kernels and biases. We use a "hourglass" architecture with the skip-connection [16] to model a mapping $f(\cdot)$ from the input variable $\boldsymbol{Z}$ to high-resolution SM map $\boldsymbol{Y}$ due to its excellent feature extraction and noise-resistant capability.

We change the "ReLU" activation function in the original U-Net architecture to the "sigmoid" activation because the value of SM is in the range from 0 to 1. In addition, the 1x1 convolution final layer for segmentation is changed to map the extracted feature to the SM output with one layer. The output layer is activated by "sigmoid" which normalizes the value of the input into [0, 1]. We reduced the feature number for each layer from [64, 128, 256, 512] to [2,4,8,16]. The "hourglass" architecture used is shown in Figure 4.2.

### 4.2.2 Data likelihood

The data likelihood is expressed as

$$p(\boldsymbol{X}|\boldsymbol{Y}) = \frac{1}{Z} exp(-\delta(\boldsymbol{X}, \Phi(\boldsymbol{Y})) \tag{4.5}$$

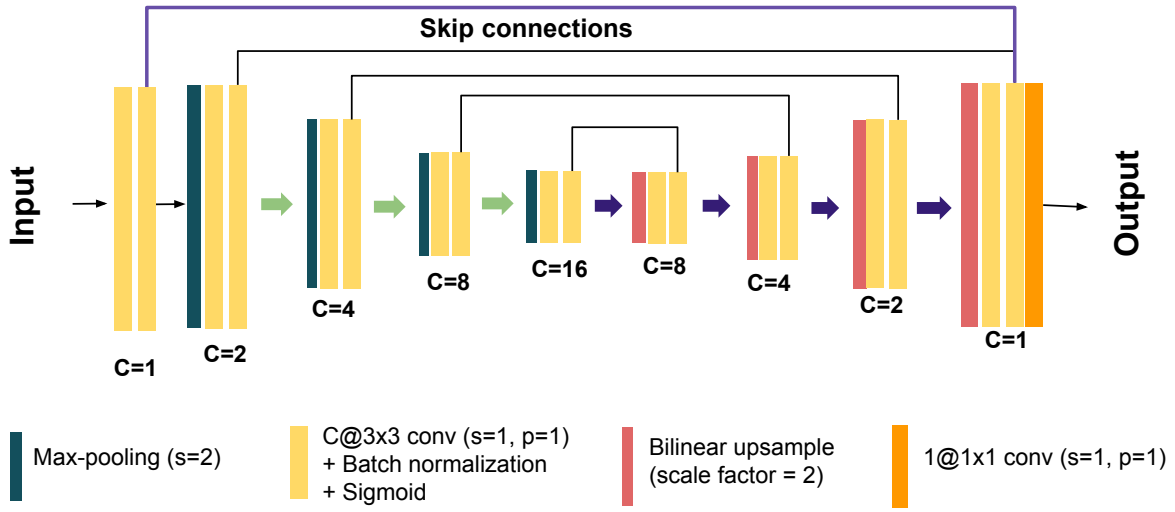



Figure 4.2: "Hourglass" architecture with skip-connections of the FCNN part in Figure 4.1 accounting for DIP.
Blocks in the figure represent for operations rather than features. The U-Net type "hourglass" architecture [36] encodes an input image to a feature tensor with a smaller size and more channels at the bottleneck, and decodes the feature tensor to the output image of the same size as the input image. The downsampling reduces the feature size which is essentially achieved by the max-pooling operation (green blocks). The up-sampling recovers the feature size step-by-step by bilinear upsampling operations (red blocks). Double regular convolution operations (yellow blocks) are conducted after each max-pooling or TransConv operation, which does not change the feature size but increases or decreases the channel number (i.e., $C$) of features. The skip connection is implemented by copying and concatenating features.

where $\delta(\boldsymbol{X}, \Phi(\boldsymbol{Y}))$ is the distance between the low-resolution SM map $\boldsymbol{X}$ and the reconstructed low-resolution SM map $\Phi(\boldsymbol{Y})$. The distance function could be implemented with different specific functions based on the real data characteristic. For example, it can be implemented as a L2-norm function when the image noise satisfies the Gaussian distribution, or as a L1-norm for Laplace distribution.

## 4.3 BDIP model optimization

The downscaling problem in Eq. 4.2 can be solved by the MAP approach, where the high-resolution SM map $\boldsymbol{Y}$ is estimated by maximizing the posterior distribution of $\boldsymbol{Y}$ given the observed low-resolution SM map $\boldsymbol{X}$, i.e.,

$$\hat{\boldsymbol{Y}} = \arg\max_{\boldsymbol{Y}} \{p(\boldsymbol{Y}|\boldsymbol{X})\} \tag{4.6}$$

Maximizing $p(\boldsymbol{Y}|\boldsymbol{X})$ is equivalent to minimizing its negative logarithm likelihood, i.e.,

$$\hat{\boldsymbol{Y}} = arg\min_{\boldsymbol{Y}} \{-logp(\boldsymbol{Y}|\boldsymbol{X})\} \tag{4.7}$$

Then, the objective function can be written as

$$\begin{aligned} J_Y &= arg\min_{\boldsymbol{Y}}\{-logp(\boldsymbol{Y}|\boldsymbol{X})\} \\ &\propto \arg\min_{\boldsymbol{Y}}\{-logp(\boldsymbol{X}|\boldsymbol{Y}) - logp(\boldsymbol{Y})\} \end{aligned} \tag{4.8}$$

Considering Eq. 4.3, 4.5, the objective function can be reformulated as

$$J_Y = arg\min_{\boldsymbol{Y}}\{\delta(\boldsymbol{X}, \Phi(E(\boldsymbol{Y}|\boldsymbol{M})))\} \tag{4.9}$$

where $E(\boldsymbol{Y}|\boldsymbol{M})$ is the posterior expectation of $\boldsymbol{Y}$ if given ancillary RS data with the high spatial resolution $\boldsymbol{M}$. We use $E(\boldsymbol{Y}|\boldsymbol{M})$ as the expectation of $\boldsymbol{Y}$. To estimate parameters in $E(\boldsymbol{Y}|\boldsymbol{M})$, we use $\boldsymbol{M}$ as input to FCNN and optimize FCNN parameters. Given the estimated parameters in FCNN, we achieve $\hat{\boldsymbol{Y}} = E(\boldsymbol{Y}|\boldsymbol{M})$, as illustrated in Section 4.2.1. When estimating parameters in FCNN for obtaining $E(\boldsymbol{Y}|\boldsymbol{M})$, we use a reconstruction distance based on $\delta(\boldsymbol{X}, \Phi(E(\boldsymbol{Y}|\boldsymbol{M})))$, which incorporates the forward model to constrain the meaningful $\boldsymbol{Y}$ estimation, as illustrated in Section 4.4.2.

To estimate $E(\boldsymbol{Y}|\boldsymbol{M})$, we first need to estimate the model parameters in FCNN, i.e., $\boldsymbol{\beta}$. Here, we construct the following objective function to estimate $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\{\delta(\boldsymbol{X}, \Phi(f(\boldsymbol{M}, \boldsymbol{\beta})))\} \tag{4.10}$$

Backpropagation with the Adam stochastic optimizer [69] is adopted in this work to estimate $\boldsymbol{\beta}$.

## 4.4 Method

### 4.4.1 Study area and datasets

We select a rectangular study area (i.e., the area inside the green box shown in Figure 4.3) where both SMAP SM products and the Moderate Resolution Imaging Spectroradiometer (MODIS) products cover the area on all eight dates in 2020. The area is across United States and Mexico ranging from 27°N to 33°N and 100°W to 108°W. The distribution of stations and the land cover map are illustrated in Figure 4.4. The distribution of stations and the land cover map are illustrated in Fig.4. The study area is mainly covered by different vegetation species including the shrublands, savannas, cropland, and the sparsely vegetated region. The open shrublands (in the middle in Figure 4.4) are normally drier than the grassland (on the right in Figure 4.4) and the woody savannas (on the bottom-left in Figure 4.4). Therefore, the SM value is lower in the middle part of the study area than side parts. So, the soil moisture can be largely spatially variated and suitable for the soil moisture study.

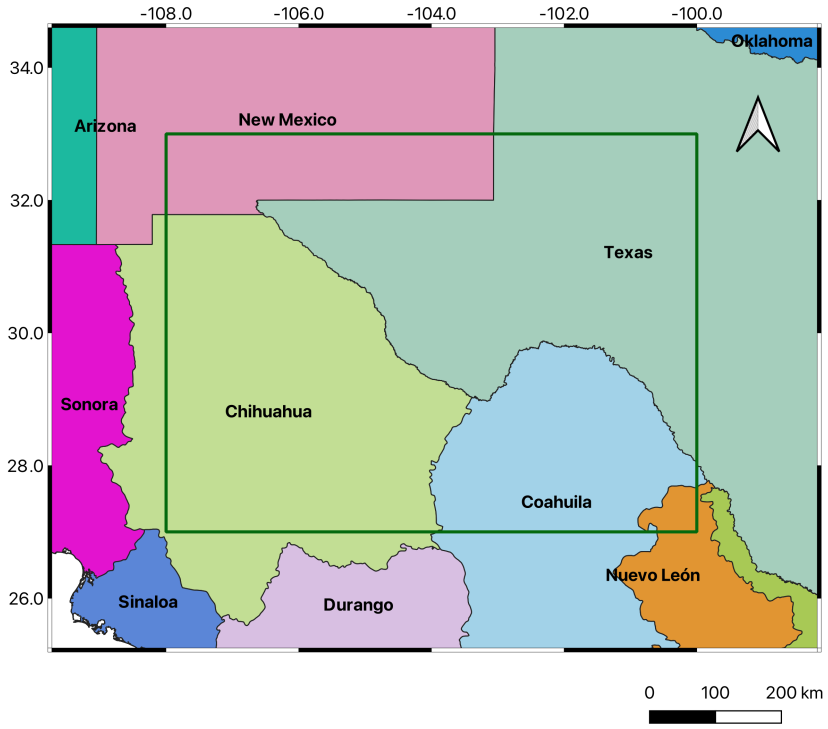



Figure 4.3: Location of the study area.

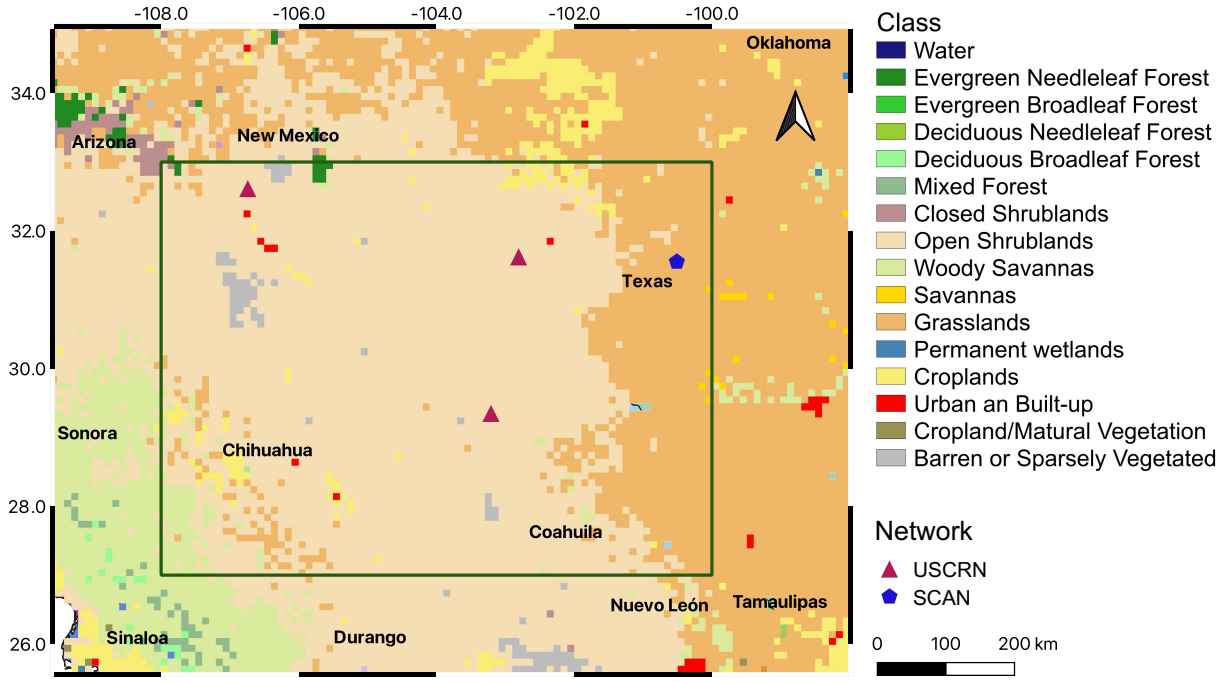



Figure 4.4: Distribution of the stations providing in situ SM measurements and Land cover map of the study area.

All data used in this chapter, including SMAP SM products at 9km spatial resolution, the MODIS products and the in-situ data, are collected on eight dates in 2020. The eight dates are January 25, Feburary 25, March 13, April 14, May 16, September 5, October 7 and December 26. The objective is to downscale SMAP SM maps to the 1km-resolution SM map. The MODIS products is utilized to provide more spatial texture information. The downscaling performance is evaluated by the in-situ SM data.

### SMAP data

The SMAP mission is an L-band satellite incorporating both a radiometer and a radar dedicated to global SM measurements [100]. The daily SMAP level-3 SM passive product at 9km (L3_SM_P_E) resolution acquired from NSIDC (National Snow and Ice Data Center) is downscaled to a 1km SM map. Only the descending data acquired at 6:00 am is used.

### MODIS products

NDVI and LST are physically related to SM [116] and commonly used for SM downscaling [100, 111, 112]. High-resolution auxiliary information, i.e., $\boldsymbol{M}$ in Eq. 4.9 , MODIS products (MYD13A2 and MYD11A2) collected from the Land Processes Distributed Active Archive Center (LPDAAC), are utilized to downscale the SMAP SM products at 9km resolution up to 1km resolution. The MYD13A2 Version 6 product provides the normalized difference vegetation index (NDVI) and the enhanced vegetation index with a 1km resolution. Only the NDVI layer from MYD13A2 is used in this study. The MYD11A2 Version 6 product provides an average 8-day per-pixel land surface temperature (LST) and emissivity with a 1km spatial resolution. Only the first layer "LST_Day_1km" from MYD11A2 is used.

### In-situ measurement

The international soil moisture network (ISMN) hosts in-situ SM measurements collected starting 1952 to present from a total of 35 international SM networks. SM data from two networks (i.e., USCRN and SCAN) are used to evaluate the downscaling quality because the stations in these two networks are distributed more densely in the study area. There are four stations in the study area. The in-situ SM observation measures the small point scale SM values and cannot be used directly in large-scale soil moisture application, the shortcoming of which can be improved by remote sensing-based SM mapping approaches. Considering that in-situ SM measures are more accurate than remote sensing SM products, here, we use these measures as ground truth to validate our downscaling results.

### Data pre-processing

For each time point, the MODIS NDVI and LST products, SMAP SM products, as well as the in-situ measurements are prepared. The MODIS MYD13A2 and MYD11A2 products at

1km resolution are downloaded and stitched together to achieve the global coverage for the further processing. The NDVI layer from MYD13A2, the LST layer from MYD11A2, and the SM layer from SMAP products layers are georeferenced and cropped by the longitude and latitude of the region of interest boundary. The image size of SMAP 9km SM, NDVI and LST, covering the study area are $74\times86$, $666\times774$ and $666\times774$, separately. The three-channel input of the network contains the 1km NDVI, 1km LST and the 1km interpolated SM, which is obtained from SM at 9km using a bilinear interpolation. Considering the fact that SMAP soil moisture range between 0 and 1, we address the negative-valued outliers as positive values using a neighborhood refilling method, in which to remove the outliers in SMAP SM products, we refill the pixels using median values of their 3x3 neighboring pixels.

### 4.4.2 Model implementation

A BDIP downscaling model is illustrated in Figure 4.5, where the FCNN $f(\cdot)$ performs the inverse model, and the downsampler $\mathrm{D}(\cdot)$ acts as the forward model. The inverse model will be trained while the forward model is known and fixed. In this manner, the FCNN can achieve the downscaling purpose by learning from the forward model and inverting the downsampling operation. Then, the relationship between SM maps at 1km ($\boldsymbol{Y}$), 9km ($\boldsymbol{X}_9$) can be expressed as follows,

$$\boldsymbol{X}_9 = \mathrm{D}_9(\boldsymbol{Y}), \tag{4.11}$$

where $\mathrm{D}_9(\cdot)$ is implemented using the "Lanczos" filtering with the downsampling factor 9.

The input $\boldsymbol{M}$ contains three layers, i.e., the NDVI from MYD13A2, LST from MYD11A2, and the interpolated 1km SM map. Then the final output of the network, which is the estimated 9km SM can be formulated as:

$$\hat{\boldsymbol{X}}_9 = (\mathrm{D}_9(f(\boldsymbol{M}, \boldsymbol{\beta})) \tag{4.12}$$

We minimize the loss function as follows to train the FCNN,

$$L = \delta(\boldsymbol{X}_9, \hat{\boldsymbol{X}}_9) \tag{4.13}$$

Unknown parameters $\boldsymbol{\beta}$ are network parameters of the FCNN, including weights and bias. Once the model is trained, the intermediate output $\boldsymbol{Y}$ can be obtained as the downscaled SM.

#### Forward model selection

The downsampler works as the forward model mapping the LR SM to the HR SM. To find an appropriate forward model with the best capability of preserving the spatial information

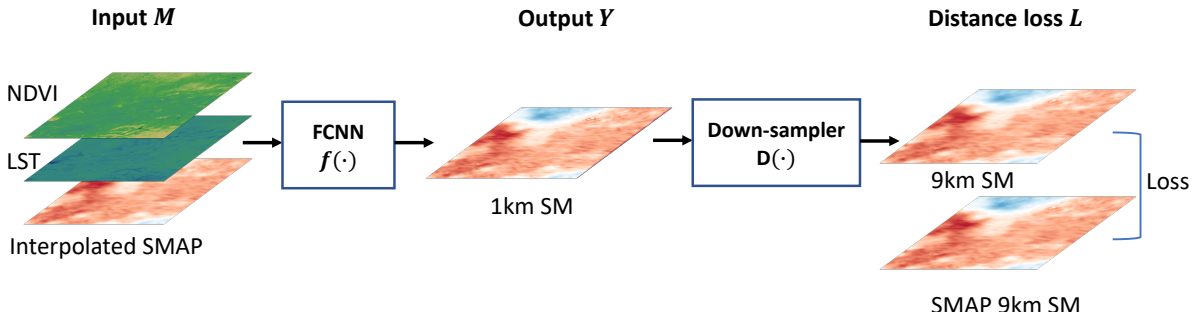



Figure 4.5: Overall model architecture for SMAP SM downscaling. The network input $\boldsymbol{M}$ includes MODIS products NDVI, LST, and the interpolated 1km SM from 9km SMAP SM. The output of the downscaling model is the downscaled SM $\boldsymbol{Y}$. Unknown parameters $\boldsymbol{\beta}$ are network parameters of the FCNN, including weights and bias.

and the highest downscaling accuracy, we try the average pooling, max-pooling and the downsampling with Lanczos kernel. Although it is claimed that no consistent differences are found among these downsampling methods for RGB images super-resolution [16], it is critical to find out their performance of downscaling on the RS imagery products.

### Loss function design

The reconstruction loss is initialized as a L2 loss, which is commonly used in image reconstruction tasks [16]. However, given its performance of blurring some detailed spatial information, L1 loss, L1 loss combined SSIM loss [117], as well as the combination of L1 loss, SSIM loss and perceptual loss [118] are tested to better reconstruct the structural spatial texture in HR SM maps.

### Skip connections

Skip connections in FCNNs solves the degradation problem and ensures the feature reusability by copying and concatenating features from shallower layers to deeper layers. To better preserve the spatial feature in the input data, besides the skip-connections existing in a classic U-Net architecture, we add a skip connection (indicated by the purple line in Figure 4.2) by concatenating the output feature of the input layer to the output feature of the last second layer (i.e, the last $3 \times 3$ convolution layer indicated by the yellow block in Figure 4.2.)

### Parameters configuration

The learning rate and training epochs for different models are summarized in Table. 4.1.

Table 4.1: Parameters configuration for different models

| Downsampler | Loss | Learning rate | Epoch |
|---|---|---|---|
| Average | L2 | 0.003 | 3000 |
| Max | L2 | 0.003 | 3000 |
| Lanczos | L2 | 0.003 | 3000 |
| Lanczos | L1 | 0.003 | 3000 |
| Lanczos | 0.1×L1+1×SSIM | 0.001 | 3000 |
| Lanczos | 0.1×L1+1×SSIM+0.5×perceptual | 0.001 | 3000 |

### 4.4.3 Methods comparison

The compared methods includes Bicubic, GFPCA [119], PCA [120] and CNMF [121]. Bicubic is a standard interpolation approach based on the cubic interpolation. GFPCA is designed for fusion of hyperspectral and RGB image based on PCA [119]. PCA, as a standard data transformation method, has been used for remote sensing data pansharpening [120, 122]. CNMF is developed based on nonnegative matrix factorization unmixing and applied to hyperspectral and multispectral data fusion and downscaling [121, 123, 124].

Compared methods are conducted using the downscaling toolbox from Github. The source code is available at https://github.com/codegaj/py_pansharpening. These methods all require two sets of inputs which are the high-resolution channels and the low-resolution channels. The summation of the NDVI and LST is used for the high-resolution input. The low-resolution channel is the 9km SMAP SM.

### 4.4.4 Evaluation strategy

Following the commonly used evaluation scheme for downscaling algorithms, the downscaled SM map is evaluated from three aspects, i.e.,

a) the consistency of the spatial variation pattern with the SMAP SM maps [112, 100],

b) the numerical accuracy of the SM values to in-situ SM measurements [112, 111, 100],

c) the amount of the spatial textural information compared to the SMAP SM map [111, 100].

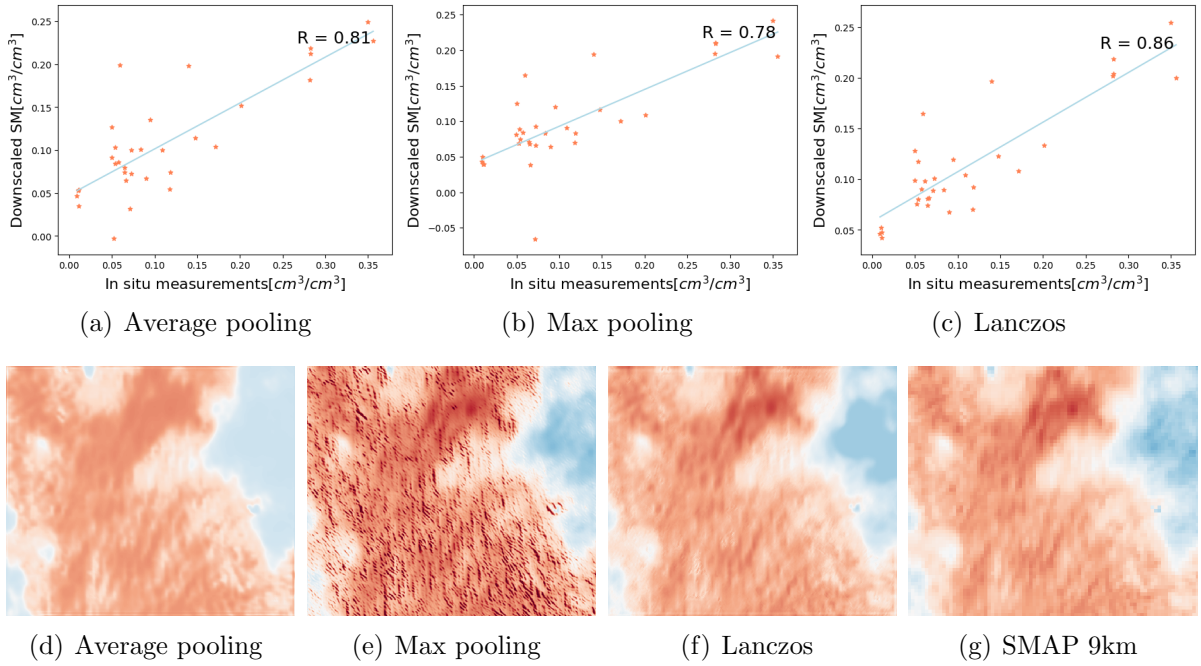
Downscaling results are evaluated in both visual and numerical ways. For the visual evaluation, the downscaled SM maps will be presented together with the SMAP SM maps at 9km spatial resolution, as well as the estimation by the other four downscaling methods. For the numerical evaluation, the in-situ groundtruth measurements on eight dates are used as the reference. The classical statistical metrics are calculated to represent the error

scores, including correlation coefficient (R), mean square error (MSE), the difference of the mean values (BIAS), root mean square error (RMSE), normalized root mean square error (nrRMSE), unbiased root mean square error (ubRMSE).

## 4.5 Result and discussion

### Forward model selection

Figure 4.6 shows the downscaling results obtained by different downsampler. Using the downsampler with Lanczos kernel gives the sharpest SM map with the most spatial texture preserved. R value over the time series corresponding to Lanczos is also the highest. The average pooling smooths some linear spatial features and the max pooling brings fake spatial features. The results indicate that using different downsamplers significantly affects the downscaling performance.




(a) Average pooling (b) Max pooling (c) Lanczos

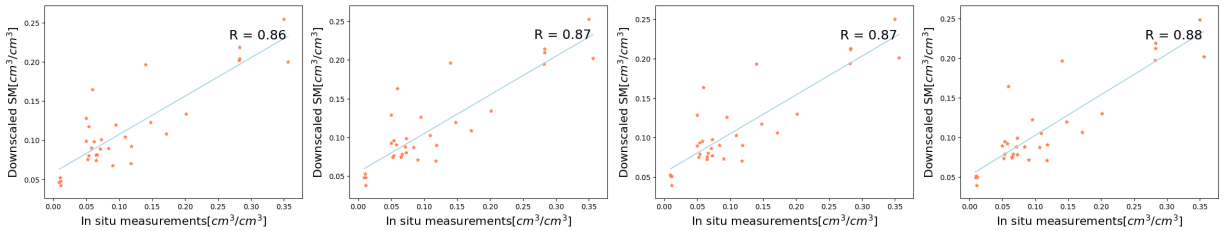
(d) Average pooling (e) Max pooling (f) Lanczos (g) SMAP 9km

Figure 4.6: Scatters of downscaled 1km SM against in-situ SM measurements over eight dates and the downscaled 1km SM maps on Jan 25th by models with different downsamplers.
Using the downsampler with Lanczos kernel gives the sharpest SM map with the most spatial texture preserved.Average pooling smooths linear spatial feature and max pooling introduce fake spatial features.

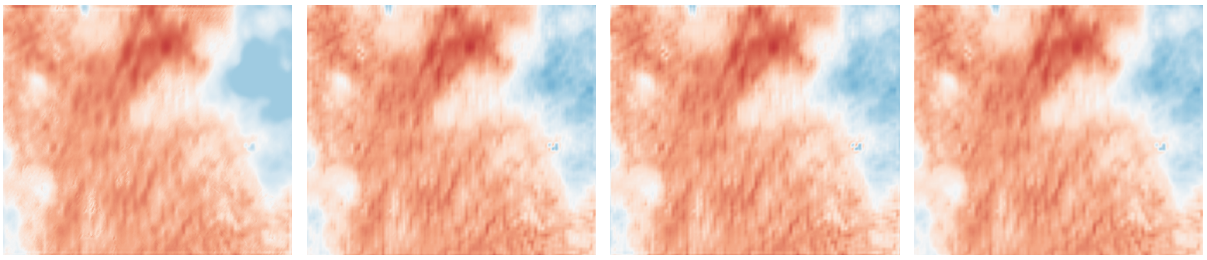
### Performance of loss functions

Once the downsampler with Lancnos kernel is selected, we fix the downsampler and change the loss function. The 1km downscaled SM map obtained with L2 loss is blurred, especially

on the right side of the image Figure 4.7(a), although the R value is high. Given that L2 loss is sensitive to high-frequency signals and tends to smooth the image, we tried L1 loss instead, which can better accommodate high-frequency information. As a result, Figure 4.7(f) is sharper than Figure 4.7(e) and shows more spatial texture, with R value increasing. Then, the SSIM loss using for preserving image structural feature and the perceptual loss using for extracting spatial information from feature domains are added one-by-one. As a result, Figure 4.7(h) shows the richest spatial information and its corresponding R value achieved 0.88. The results indicate the importance of designing loss functions for downscaling visual performance. Although the R value does not highly increased, the spatial information shown in the downscaled SM map gets sharper and richer.




(a) L2 (b) L1 (c) L1+SSIM (d) L1+SSIM+perceptual

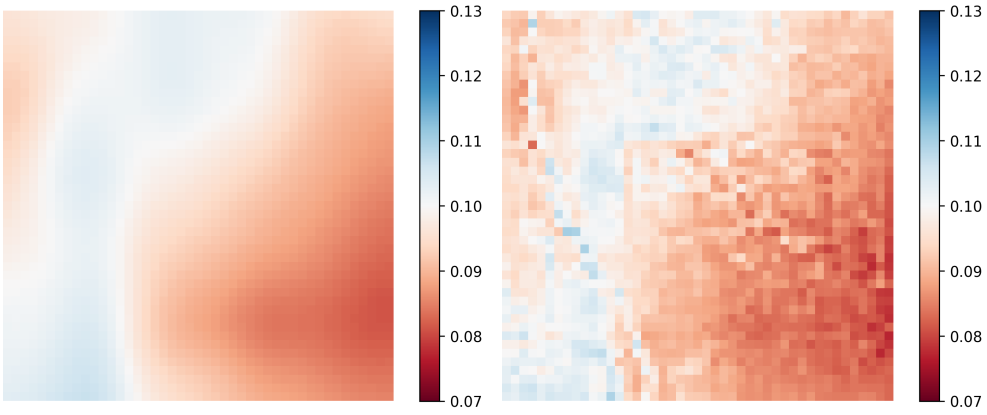


(e) L2 (f) L1 (g) L1+SSIM (h) L1+SSIM+perceptual

Figure 4.7: Downscaled 1km SM maps by networks with different loss implementations. L2 loss tends to smooth the image. L1 loss can better accommodate high-frequency information. Although (f) is sharper than (e), the R value does not increase. Then the SSIM loss using for preserving image structural feature and the perceptual loss using for extracting spatial information from the feature domains are added one-by-one. As a result, (h) shows the richest spatial information and its corresponding R value achieved 0.88.

### Performance of skip connection

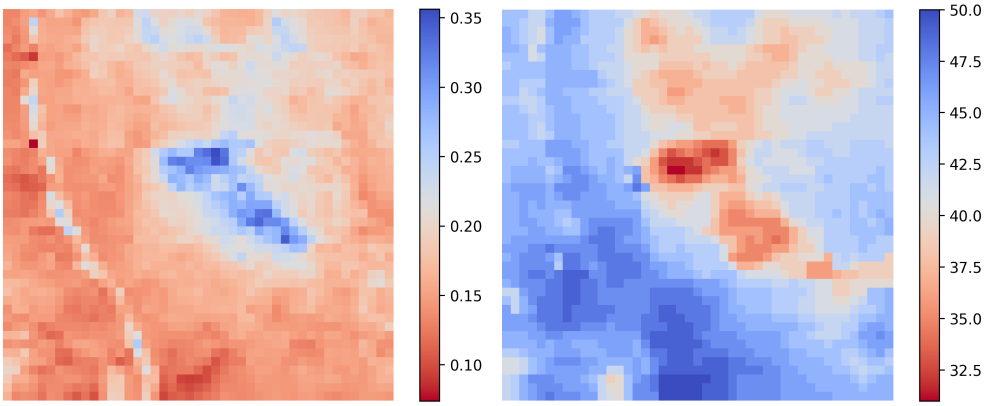
Downscaling performances are compared between the U-Net with and without the additional skip connection. The result is shown in Figure 4.8. The downscaling result with the skip connection (Figure 4.8(b)) shows much richer spatial information than that without the skip connection (Figure 4.8(a)). So, the skip connection in the U-Net can better preserve the low-level feature in NDVI and LST.




(a) without skip




(b) with skip




(c) NDVI

(d) LST

Figure 4.8: Downscaled 1km SM maps (unit: $cm^3/cm^3$) by networks (a) with and (b) without the additional skip-connection (indicated by the purple line in Figure 4.2). (b) fuses spatial information from the (c) NDVI and (d) LST better than (a).

### Spatial detail restoration

To check the downscaling effectiveness of the proposed method, the downscaled 1km SM maps and the 9km SMAP SM maps are zoomed in different scales, shown in Figure 4.9. The downscaled SM map shows not only the consistent variation pattern with the 9km SMAP SM, but also much more spatial detail information which is consistent with satellite RGB images. For example, the green linear region in Figure 4.9(e) is the cropland with higher water content which is indicated by the blue linear feature in Figure 4.9(d).

### Methods comparison

Table 4.3 summaries R values, BIAS and RMSE values between the 1km downscaled SM map and the in-situ groundtruth over eight dates. 32 points (i.e., 4 stations $\times$ 8 days) in total are used to calculate metrics.

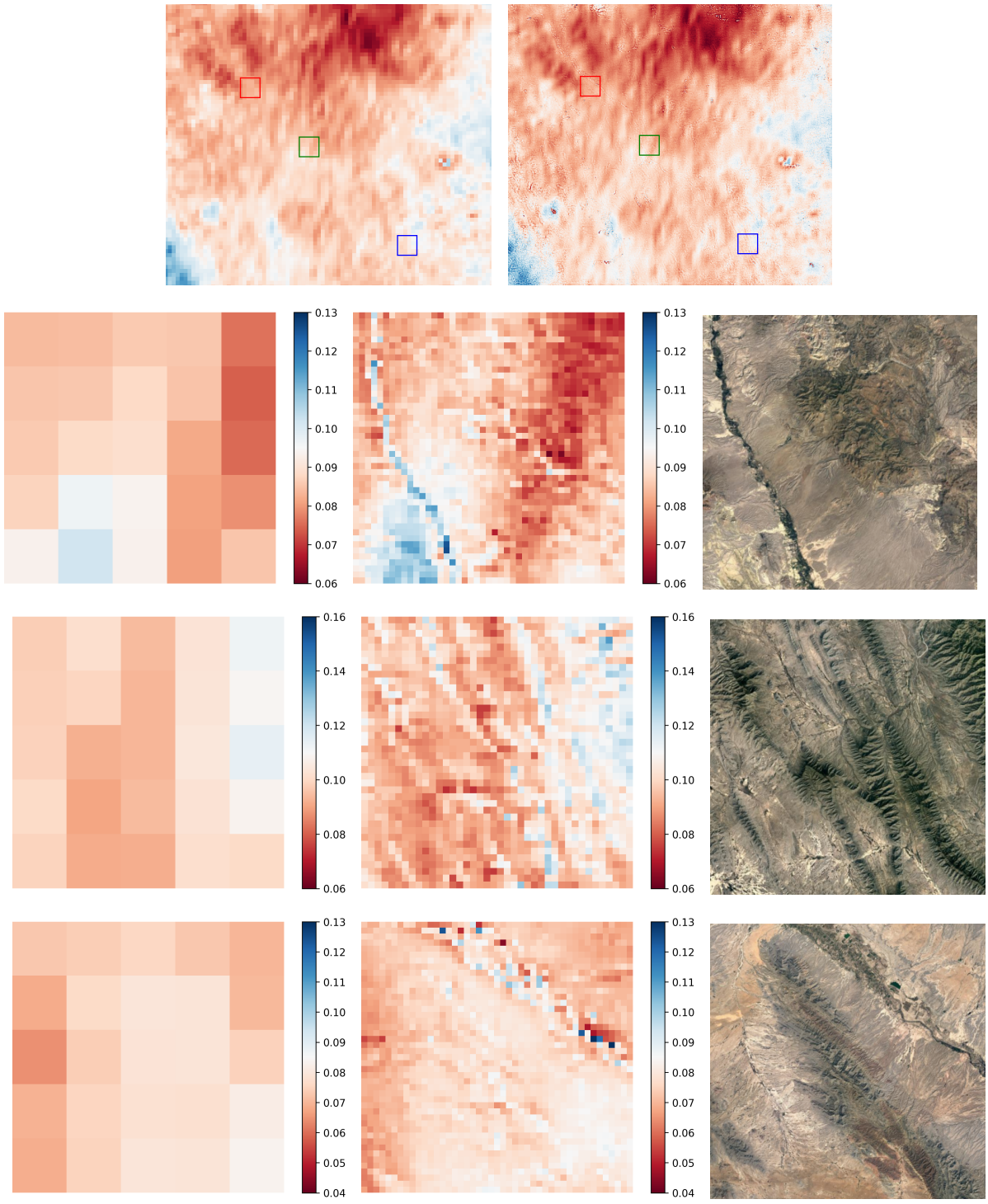



Figure 4.9: Comparison between the 9km SMAP SM maps and downscaled 1km SM maps (unit: $cm^3/cm^3$) on Dec 26th. Figures in the 1st line are separately the 9km SMAP (left) and the 1km downscaled SM (right). Figures in the 2nd line are separately the zoomed region on 9km SMAP and 1km SM indicated with the green box in the 1st line. Figure on the right is the corresponding area clipped from the google earth. Figures in 3rd and 4th lines are separately correspond to the blue box and red box in the 1st line. The downscaled SM map shows not only the consistent variation pattern with the 9km SMAP SM, but also much more spatial detail information. For example, the green linear region in the rgb image in the 2nd line is the cropland with higher water content which is indicated by the blue linear feature in the zoomed 1km SM.

Table 4.2: Methods assessment.

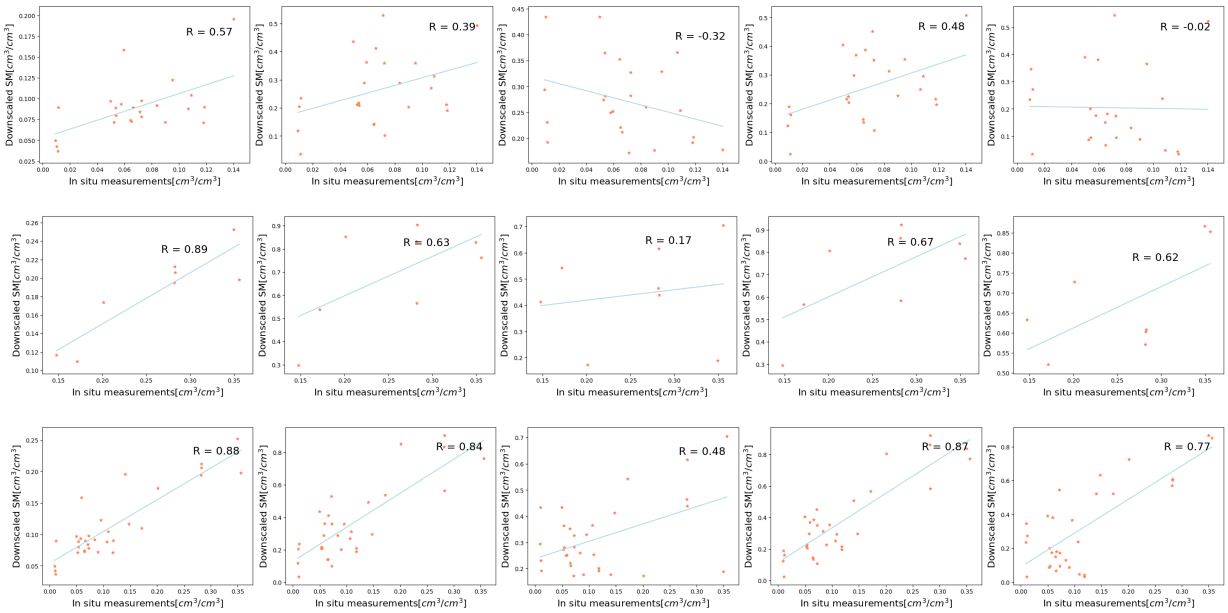
| Method | MODIS used? | Performance | | |
|---|---|---|---|---|
| | | Variation consistent? | Spatial texture increased? | SM value range correct? |
| BDIP | Yes | Yes | Yes | Yes |
| Bicubic | No | Yes | Yes | No |
| GFPCA[119] | Yes | Yes | No | No |
| PCA[120] | Yes | No | Yes | No |
| CNMF[121] | Yes | No | Yes | No |

The validation scatterplots over eight dates shown in Figure 4.10 show consistent correlation degrees with Table 4.3. Scatters of the proposed BDIP method show a obvious linear relationship between the downscaled SM and groundtruth data.

By observing the scatters, we found that the measurements within the USCRN network are generally smaller than the SCAN network because the USCRN stations distributed in the shrublands, and the SCAN station is in the grassland, where the soil normally contains more water. It is found that the score of SCAN network is generally better than USCRN network. The possible reasons are listed as follows. (i) The better statistical score could be caused by fewer station points. (ii)The soil moisture is overall higher at SCAN station than that at USCRN stations. (iii)Sensors of these two networks could be different.

Figure 4.11 displays the downscaled SM maps by different methods on two dates. SM maps generated by the Bicubic and GFPCA share consistent variation patterns with SMAP SM maps, but with large bias. GFPCA SM maps get more blurred than 9km SMAP SM. The proposed method, on the contrary, estimates the high-resolution SM map with sharp and clear boundaries. SM maps achieved by PCA and CNMF preserve much information in the NDVI and LST than the SMAP SM map, which is the fake SM texture. They fail to properly extract and balance the spatial feature information from the SMAP SM and MODIS products. The above results description is summarized in Table 4.2.

The PCA and CNMF methods were designed for multispectral, hyperspectral images pansharpening, where the HR images and the LR images share the similar spatial texture. PCA and CNMF are also used to enhance the contrast of the original image. However, for the SM downscaling guided by NDVI and LST, the HR NDVI, LST and the LR SMAP SM have different spatial texture. So, simply extracting the spatial textural information from all bands leads to the failure of data fusion. GFPCA performs better than PCA and CNMF because a transformation from NDVI and LST to the SMAP SM was conducted instead of extracting information from all of MODIS and SMAP products. However, the downscaled SM still gets blurred, which could be caused by the transformation or upsampling procedure. Bicubic interpolates the SMAP SM directly without using MODIS products, leading to insufficient spatial details.




(a) BDIP (b) Bicubic (c) PCA (d) GFPCA (e) CNMF

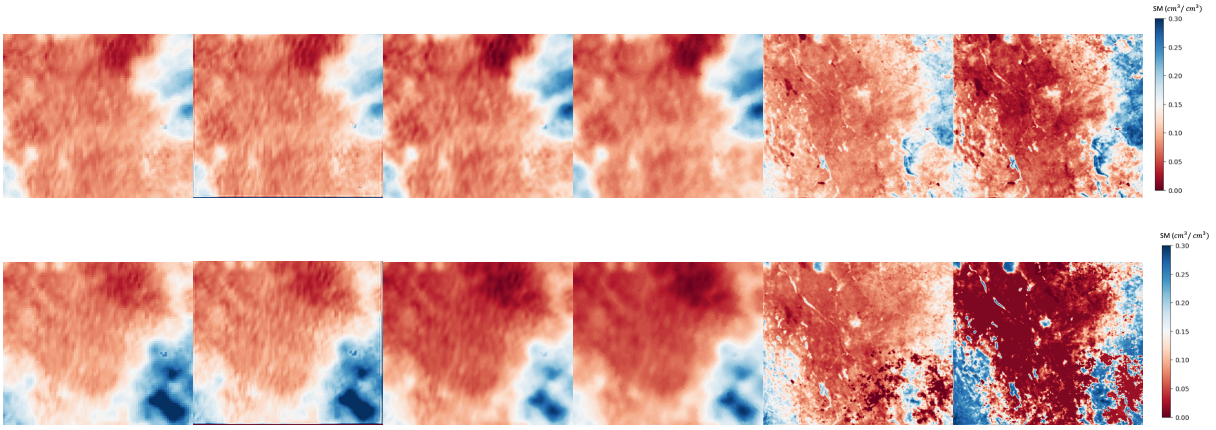
Figure 4.10: Scatters of the 1km SM estimated by the different method and the in-situ groundtruth over eight dates with R values. Two networks are separately calculated. The first row is USCRN network and the second row is the SCAN network. The last row is the result obtained by all stations from two networks.

To sum up, the downscaled 1km SM by our proposed method not only has the consistent variation pattern with the SMAP maps, but also restores more spatial details than other methods with higher accuracy.

## 4.6 Conclusion

We proposed a Bayesian DIP downscaling model for SMAP SM products by integrating the FCNN into a Bayesian framework. MODIS products was used as the model input to guide the downscaling procedure. An hourglass FCNN was adopted to map the nonlinear relationship between MODIS products and high-resolution SM map and to better construct the spatial heterogeneous information in SM map. The MAP inverse problem was solved by back propagation instead of EM iterations, which makes the model optimization simpler and faster. Experiments on the time series data showed that SM maps estimated by the proposed method provided more spatial texture details than other existing unsupervised downscaling methods, and the estimated SM was very close to in-situ measurements with a high overall R value 0.88. The proposed Bayesian downscaling model are very effective for SMAP SM downscaling.

Despite of the successful of the BDIP downscaling approach based on the in situ and visual validation, the chapter has shortcoming of insufficient result analysis from the ge-




(a) SMAP (b) BDIP (c) Bicubic (d) GFPCA (e) PCA (f) CNMF

Figure 4.11: SMAP SM map at 9km resolution in column (a), the downscaled 1km SM maps by the different methods from column (b) to (f), and the input NDVI map in column (g) at April 14 (first row) and Sep 5 (second row). The proposed method estimates the high-resolution SM map with sharp and clear boundaries. SM maps generated by the Bicubic and GFPCA share consistent variation patterns with SMAP SM maps, but with large bias. GFPCA SM maps get more blurred than 9km SMAP SM. SM maps achieved by PCA and CNMF preserve much information in the NDVI and LST than the SMAP SM map, which is fake SM texture. They are not able to properly extract and balance the spatial feature information from the SMAP SM and MODIS products.

ographical perspective, such as how the downscaled SM map correlates in fine-scale with the land cover types, precipitation and the elevation. Another shortcoming of the chapter is the insufficient comparison with more advanced downscaling methods considering that the unsupervised downscaling approach is limited. However, this unsupervised approach has larger potential than supervised ones to be widely used without high-resolution maps required. Moreover, since the proposed model is flexible to fuses multi-source remote sensing products and its downsampler part can be adjust according to the resolution of existing SM products, it has the big potential to be applied to more SM products with different spatial resolutions and to fuse more remote sensing products, such as precipitation and terrain products.

Table 4.3: R, MSE($cm^6/cm^6$), BIAS($cm^3/cm^3$), RMSE($cm^3/cm^3$), nrRMSE($cm^3/cm^3$) and ubRMSE($cm^3/cm^3$) of the validation for the 1km downscaled SM with the measurement of in-situ stations from two networks.

| | | R | MSE | BIAS | RMSE | nrRMSE | ubRMSE |
|---|---|---|---|---|---|---|---|
| PCA | All(4 stats) | 0.475 | 0.054 | -0.198 | 0.232 | 0.334 | 0.121 |
| | SCAN (1 stat) | 0.165 | 0.065 | -0.183 | 0.256 | 0.460 | 0.179 |
| | USCRN (3 stats) | -0.315 | 0.050 | -0.203 | 0.224 | 0.363 | 0.526 |
| GFPCA | All | 0.871 | 0.090 | -0.252 | 0.301 | 0.330 | 0.163 |
| | SCAN | 0.668 | 0.224 | -0.447 | 0.473 | 0.611 | 0.156 |
| | USCRN | 0.475 | 0.046 | -0.188 | 0.215 | 0.431 | 0.104 |
| CNMF | All | 0.765 | 0.076 | -0.205 | 0.276 | 0.322 | 0.185 |
| | SCAN | 0.623 | 0.180 | -0.414 | 0.424 | 0.590 | 0.094 |
| | USCRN | -0.018 | 0.041 | -0.135 | 0.204 | 0.381 | 0.153 |
| Bicubic | All | 0.842 | 0.092 | -0.256 | 0.304 | 0.340 | 0.164 |
| | SCAN | 0.631 | 0.218 | -0.438 | 0.467 | 0.617 | 0.160 |
| | USCRN | 0.387 | 0.051 | -0.195 | 0.226 | 0.434 | 0.113 |
| Ours | All | **0.882** | **0.002** | **0.003** | **0.053** | **0.155** | **0.053** |
| | SCAN | **0.891** | **0.007** | **0.076** | **0.085** | **0.346** | **0.038** |
| | USCRN | **0.568** | **0.001** | **-0.020** | **0.037** | **0.202** | **0.031** |

# Chapter 5

# Conclusion

In summary, this thesis proposed three task-specific methods based on a Bayesian DIP framework, which disentangle mixed pixels into application-dependent components for RS images inversion. The proposed SU method (in Chapter 2) incorporates a linear mixture forward model, the DIP accounting for the abundance spatial correlation, the noise heterogeneity of HSIs and a purified-means endmember constraint to the Bayesian framework. A designed EM algorithm solves the resulting maximum a posteriori problem to achieve accurate SU results. The SPM approach (in Chapter 3) adopts a discrete spectral mixture model as the forward equation to model the subpixel labels. The SMD research (in Chapter 4) fuses higher-resolution RS products to improve the spatial resolution of the current SM product. It is a successful generalization of the Bayesian DIP framework to applications of RS data fusion and environmental monitoring. All these above proposed methods are unsupervised, i.e., not requiring high-resolution ground truth, by making use of the forward models to reconstruct the observed low-resolution data. They outperform other unsupervised state-of-arts methods on the tested datasets.

## 5.1 Summary of contributions

This thesis improved mixed pixel disentangling in three key applications, i.e., SU, SPM and SMD by integrating the DIP approach and other prior information into a Bayesian framework to allow comprehensive usage of different prior knowledge for enhanced data inversion. This thesis has the following main contributions.

1. To improve **the decomposition of mixed pixels into endmembers and abundances in SU**, a designed DCNN and a new spectral mixture model with heterogeneous noise are integrated into a Bayesian framework that is efficiently solved by a new iterative optimization algorithm.

2. To improve **the decomposition of mixed pixels into class labels of subpixels in SPM**, a dedicated DCNN architecture and a new discrete spectral mixture model are integrated into the Bayesian framework to allow the use of both spatial prior and the forward model.

3. To improve **the decomposition of mixed pixels into soil moisture concentrations of subpixels in SMD**, a new DIP architecture and a forward degradation model are integrated into the Bayesian framework that is solved by the stochastic gradient descent approach.

Some specific contributions in this thesis include:

1. A skip-connection FCNN is designed for the estimation of underlying quantities, where DIP is used to model the spatial correlation of the abundance field, subpixel label field, and HR SM field. Compared to NNLS or fully connected network, the FCNN is able to efficiently and accurately estimate the desired quantities by leveraging GPUs and the large-scale spatial correlation in RS images.

2. The noise is modelled as a multivariate Gaussian distribution to account for the noise variance heterogeneity in HSI for SU. As a result, the loss function of BCUN is designed based on the M-distance rather than MSE loss. The designed conditional distribution of spectral observations also enables the incorporation of the SMM into the FCNN training process for effectively leveraging the knowledge in the forward spectral model.

3. The endmember is modelled and estimated by a "purified means" approach which can be seamlessly integrated into the Bayesian framework by a designed conditional distribution of the endmembers given the abundance.

4. Different forward models describing the mixed pixels generation process in RS images are integrated into the Bayesian framework to solve the different inverse problems.

5. The key components are coherently integrated into a Bayesian framework, and the resulting MAP problem is solved by a designed EM algorithm for SU and SPM.

6. The proposed framework is trained by reconstructing HSIs or the LR soil moisture without requiring any ground-truth data for model training.

## 5.2 Future work directions

1. More complex physical forward models can be integrated to the framework. For example, most SU studies assume the linear SMM as the generative model of mixed pixels while non-linear models should be considered to better explain the generation process.

2. Endmembers are modelled by a purified-means approach and estimated in the M-step in an EM approach. How to estimate endmembers by treating endmembers as nodes of a network is a promising study direction.

3. The hourglass UNet architecture is adopted as the inverse model in the current methods. More network architectures can be studied and compared, such as ResNet and the recently proposed transformers. Although the DIP is efficient to capture the spatial correlation in RS images, the convolutions tend to smooth the spatial edge features. Therefore, state-of-arts methods that can better model the non-local spatial correlation should be studied.

4. The proposed framework is currently applied to the single-image processing. It is of great significance to train the model with RS image datasets and apply the pre-trained model to more test images.

5. When the forward model is not fully known or it contains unknown parameters, how to utilize the known information of the forward model and estimate the unknown parameters are of interest.

# References


[1] Claire Thomas, Thierry Ranchin, Lucien Wald, and Jocelyn Chanussot. Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1301–1312, 2008.

[2] Hassan Ghassemian. A review of remote sensing image fusion methods. *Information Fusion*, 32:75–89, 2016.

[3] Telmo Adão, Jonáš Hruška, Luís Pádua, José Bessa, Emanuel Peres, Raul Morais, and Joaquim Sousa. Hyperspectral imaging: A review on uav-based sensors, data processing and applications for agriculture and forestry. *Remote Sensing*, 9(11):1110, 2017.

[4] Robert Jackisch, Sandra Lorenz, Robert Zimmermann, Robert Möckel, and Richard Gloaguen. Drone-borne hyperspectral monitoring of acid mine drainage: An example from the sokolov lignite district. *Remote Sensing*, 10(3):385, 2018.

[5] Renbao Lian, Weixing Wang, Nadir Mustafa, and Liqin Huang. Road extraction methods in high-resolution remote sensing images: A comprehensive review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:5489–5507, 2020.

[6] Xiaoping Liu, Guohua Hu, Bin Ai, Xia Li, and Qian Shi. A normalized urban areas composite index (NUACI) based on combination of DMSP-OLS and MODIS for mapping impervious surface area. *Remote Sensing*, 7(12):17168–17189, 2015.

[7] Sabah Sabaghy, Jeffrey P Walker, Luigi J Renzullo, Ruzbeh Akbar, Steven Chan, Julian Chaubell, Narendra Das, R Scott Dunbar, Dara Entekhabi, Anouk Gevaert, et al. Comprehensive analysis of alternative downscaled soil moisture products. *Remote Sensing of Environment*, 239:111586, 2020.

[8] Xulong Liu, Ruru Deng, Jianhui Xu, and Feifei Zhang. Coupling the modified linear spectral mixture analysis and pixel-swapping methods for improving subpixel water mapping: application to the Pearl River Delta, China. *Water*, 9(9):658, 2017.

[9] Qunming Wang, Peter M Atkinson, and Wenzhong Shi. Fast subpixel mapping algorithms for subpixel resolution change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 53(4):1692–1706, 2014.

[10] Andrew Blake, Pushmeet Kohli, and Carsten Rother. *Markov Random Fields for Vision and Image Processing*. MIT Press, 2011.

[11] Kevin P Murphy. Undirected graphical models (Markov random fields). *Machine Learning: A Probabilistic Perspective; MIT Press: Cambridge, MA, USA*, pages 661–705, 2012.

[12] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. Boston, USA, 2005.

[13] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018.

[14] Xiao Bai, Fan Xu, Lei Zhou, Yan Xing, Lu Bai, and Jun Zhou. Nonlocal similarity based nonnegative tucker decomposition for hyperspectral image denoising. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):701–712, 2018.

[15] Zezhou Cheng, Matheus Gadelha, Subhransu Maji, and Daniel Sheldon. A Bayesian perspective on the deep image prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5443–5451, 2019.

[16] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[17] Savas Ozkan and Gozde Bozdagi Akar. Deep spectral convolution network for hyperspectral unmixing. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3313–3317. IEEE, 2018.

[18] Kuang Gong, Ciprian Catana, Jinyi Qi, and Quanzheng Li. Pet image reconstruction using deep image prior. *IEEE transactions on medical imaging*, 38(7):1655–1665, 2018.

[19] Emrah Bostan, Reinhard Heckel, Michael Chen, Michael Kellman, and Laura Waller. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. *Optica*, 7(6):559–562, 2020.

[20] Kevin C Zhou and Roarke Horstmeyer. Diffraction tomography with a deep image prior. *Optics express*, 28(9):12872–12896, 2020.

[21] Xiaofeng Ma, Youtang Hong, and Yongze Song. Super resolution land cover mapping of hyperspectral images using the deep image prior-based approach. *International Journal of Remote Sensing*, 41(7):2818–2834, 2020.

[22] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019.

[23] Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pages 524–533. PMLR, 2019.

[24] Cheng Jiang, Hongyan Zhang, Huanfeng Shen, and Liangpei Zhang. Two-step sparse coding for the pan-sharpening of remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(5):1792–1805, 2013.

[25] Bouthayna Msellmi, Daniele Picone, Zouhaier Ben Rabah, Mauro Dalla Mura, and Imed Riadh Farah. Sub-pixel mapping model based on total variation regularization and learned spatial dictionary. *Remote Sensing*, 13(2):190, 2021.

[26] Yong Chen, Ting-Zhu Huang, and Xi-Le Zhao. Destriping of multispectral remote sensing image using low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(12):4950–4967, 2018.

[27] Junchang Ju, Eric D Kolaczyk, and Sucharita Gopal. Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing. *Remote Sensing of Environment*, 84(4):550–560, 2003.

[28] Hao Wang and Dit-Yan Yeung. A survey on Bayesian deep learning. *ACM Computing Surveys (CSUR)*, 53(5):1–37, 2020.

[29] Yuan Fang, Linlin Xu, Junhuan Peng, Honglei Yang, Alexander Wong, and David A Clausi. Unsupervised Bayesian classification of a hyperspectral image based on the spectral mixture model and Markov random field. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9):3325–3337, 2018.

[30] Tingting Wang, Faming Fang, Fang Li, and Guixu Zhang. High-quality Bayesian pansharpening. *IEEE Transactions on Image Processing*, 28(1):227–239, 2018.

[31] Yuan Fang, Yuxian Wang, Linlin Xu, Rongming Zhuo, Alexander Wong, and David A Clausi. BCUN: Bayesian fully convolutional neural network for hyperspectral spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[32] Geonho Cha, Minsik Lee, and Songhwai Oh. Unsupervised 3d reconstruction networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3849–3858, 2019.

[33] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. *IEEE transactions on medical imaging*, 38(2):394–405, 2018.

[34] Shima Kamyab, Zohreh Azimifar, Rasool Sabzi, and Paul Fieguth. Deep learning methods for inverse problems. *PeerJ Computer Science*, 8:e951, 2022.

[35] Licheng Jiao, Miaomiao Liang, Huan Chen, Shuyuan Yang, Hongying Liu, and Xianghai Cao. Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10):5585–5599, 2017.

[36] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[37] Dongxian Wu, Yisen Wang, Shu-Tao Xia, James Bailey, and Xingjun Ma. Skip connections matter: On the transferability of adversarial examples generated with resnets. *arXiv preprint arXiv:2002.05990*, 2020.

[38] Zhaofan Qiu, Ting Yao, and Tao Mei. Learning spatio-temporal representation with pseudo-3d residual networks. In *proceedings of the IEEE International Conference on Computer Vision*, pages 5533–5541, 2017.

[39] Joe McGlinchy, Brian Johnson, Brian Muller, Maxwell Joseph, and Jeremy Diaz. Application of unet fully convolutional neural network to impervious surface segmentation in urban environment from high resolution satellite imagery. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3915–3918. IEEE, 2019.

[40] Burkni Palsson, Jakob Sigurdsson, Johannes R Sveinsson, and Magnus O Ulfarsson. Hyperspectral unmixing using a neural network autoencoder. *IEEE Access*, 6:25646–25656, 2018.

[41] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2012.

[42] Xiaoxia Sun, Qing Qu, Nasser M Nasrabadi, and Trac D Tran. Structured priors for sparse-representation-based hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 11(7):1235–1239, 2013.

[43] Yujia Chen, Linlin Xu, Yuan Fang, Junhuan Peng, Wenfu Yang, Alexander Wong, and David A Clausi. Unsupervised Bayesian subpixel mapping of hyperspectral imagery based on band-weighted discrete spectral mixture model and Markov random field. *IEEE Geoscience and Remote Sensing Letters*, 2020.

[44] Antonio Plaza, Qian Du, José M Bioucas-Dias, Xiuping Jia, and Fred A Kruse. Foreword to the special issue on spectral unmixing of remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing*, 49(11):4103–4110, 2011.

[45] José M Bioucas-Dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2):354–379, 2012.

[46] Linlin Xu, Alexander Wong, Fan Li, and David A Clausi. Intrinsic representation of hyperspectral imagery for unsupervised feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2):1118–1130, 2015.

[47] Charles L Lawson and Richard J Hanson. *Solving Least Squares Problems*, volume 15. Siam, 1995.

[48] José M Bioucas-Dias and Mário AT Figueiredo. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. In *2010 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, pages 1–4. IEEE, 2010.

[49] Ying Qu and Hairong Qi. Udas: An untied denoising autoencoder with sparsity for spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 57(3):1698–1712, 2018.

[50] Yuanchao Su, Jun Li, Antonio Plaza, Andrea Marinoni, Paolo Gamba, and Somdatta Chakravortty. Daen: Deep autoencoder networks for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[51] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[52] Xiaomei Zhao, Yihong Wu, Guidong Song, Zhenye Li, Yazhuo Zhang, and Yong Fan. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis*, 43:98 – 111, 2018.

[53] Yunsong Li, Jing Hu, Xi Zhao, Weiying Xie, and JiaoJiao Li. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 266:29–41, 2017.

[54] Jiawei Zhang, Jinshan Pan, Wei-Sheng Lai, Rynson WH Lau, and Ming-Hsuan Yang. Learning fully convolutional networks for iterative non-blind deconvolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3817–3825, 2017.

[55] Behnood Rasti, Bikram Koirala, Paul Scheunders, and Pedram Ghamisi. Undip: Hyperspectral unmixing using deep image prior. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[56] Dong Xu, Lei Sun, and Jianshu Luo. Noise estimation of hyperspectral remote sensing image based on multiple linear regression and wavelet transform. *Boletim de Ciências Geodésicas*, 19(4):639–652, 2013.

[57] José MP Nascimento and José MB Dias. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE transactions on Geoscience and Remote Sensing*, 43(4):898–910, 2005.

[58] Joseph W Boardman et al. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, volume 1, pages 11–14, 1993.

[59] Linlin Xu, Jonathan Li, Alexander Wong, and Junhuan Peng. Kp-means: A clustering algorithm of k "purified" means for hyperspectral endmember estimation. *IEEE Geoscience and Remote Sensing Letters*, 11(10):1787–1791, 2014.

[60] Linlin Xu, Alexander Wong, Fan Li, and David A Clausi. Extraction of endmembers from hyperspectral images using a weighted fuzzy purified-means clustering model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2):695–707, 2015.

[61] Daniel C Heinz et al. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(3):529–545, 2001.

[62] Xiangyang Kong, Yongqiang Zhao, Jize Xue, and Jonathan Cheung-Wai Chan. Hyperspectral image denoising using global weighted tensor norm minimum and nonlocal low-rank approximation. *Remote Sensing*, 11(19):2281, 2019.

[63] Behnood Rasti, Paul Scheunders, Pedram Ghamisi, Giorgio Licciardi, and Jocelyn Chanussot. Noise reduction in hyperspectral imagery: Overview and application. *Remote Sensing*, 10(3):482, 2018.

[64] Nicolas Dobigeon, Jean-Yves Tourneret, and Chein-I Chang. Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery. *IEEE Transactions on Signal Processing*, 56(7):2684–2695, 2008.

[65] Konstantinos E Themelis, Athanasios A Rontogiannis, and Konstantinos D Koutroumbas. A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing. *IEEE Transactions on Signal Processing*, 60(2):585–599, 2011.

[66] Savas Ozkan and Gozde Bozdagi Akar. Improved deep spectral convolution network for hyperspectral unmixing with multinomial mixture kernel and endmember uncertainty. *arXiv preprint arXiv:1808.01104*, 2018.

[67] Morten Arngren, Mikkel N Schmidt, and Jan Larsen. Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior. *Journal of Signal Processing Systems*, 65(3):479–496, 2011.

[68] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.

[69] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[70] Feiyun Zhu, Ying Wang, Shiming Xiang, Bin Fan, and Chunhong Pan. Structured sparse method for hyperspectral unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88:101–118, 2014.

[71] Hector M Vargas and Henry Arguello Fuentes. Colored coded-apertures for spectral image unmixing. In *Image and Signal Processing for Remote Sensing XXI*, volume 9643, page 964320. International Society for Optics and Photonics, 2015.

[72] Lei Tong, Jun Zhou, Yuntao Qian, Xiao Bai, and Yongsheng Gao. Nonnegative-matrix-factorization-based hyperspectral unmixing with partially known endmembers. *IEEE Transactions on Geoscience and Remote Sensing*, 54(11):6531–6544, 2016.

[73] Veera Senthil Kumar Ganesan and Vasuki S. Clustering based band selection for endmember extraction using simplex growing algorithm in hyperspectral images. *Multimedia Tools and Applications*, 76(6):8355–8371, 2017.

[74] Tatsumi Uezato, Mathieu Fauvel, and Nicolas Dobigeon. Hyperspectral image unmixing with lidar data-aided spatial regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 56(7):4098–4108, 2018.

[75] Michael E Winter. N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *Imaging Spectrometry V*, volume 3753, pages 266–275. International Society for Optics and Photonics, 1999.

[76] Xinyu Wang, Yanfei Zhong, Liangpei Zhang, and Yanyan Xu. Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 55(11):6287–6304, 2017.

[77] Ying Qu, Rui Guo, and Hairong Qi. Spectral unmixing through part-based nonnegative constraint denoising autoencoder. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 209–212. IEEE, 2017.

[78] Min Zhao, Xiuheng Wang, Jie Chen, and Wei Chen. A plug-and-play priors framework for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[79] Xiuheng Wang, Min Zhao, and Jie Chen. Hyperspectral unmixing via plug-and-play priors. In *2020 IEEE International Conference on Image Processing (ICIP)*, pages 1063–1067. IEEE, 2020.

[80] Ulugbek S Kamilov, Hassan Mansour, and Brendt Wohlberg. A plug-and-play priors approach for solving nonlinear imaging inverse problems. *IEEE Signal Processing Letters*, 24(12):1872–1876, 2017.

[81] Mertens, KC, De, Baets, B, Verbeke, LPC, De, Wulf, and RR. A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models. *Int J Remote Sens*, 2006.

[82] Bouthayna Msellmi, Daniele Picone, Zouhaier Ben Rabah, Mauro Dalla Mura, and Imed Riadh Farah. Sub-pixel mapping model based on total variation regularization and learned spatial dictionary. *Remote Sensing*, 13(2):190, 2021.

[83] Da He, Yanfei Zhong, Xinyu Wang, and Liangpei Zhang. Deep convolutional neural network framework for subpixel mapping. *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[84] Koen C Mertens, Bernard De Baets, Lieven PC Verbeke, and Robert R De Wulf. A sub-pixel mapping algorithm based on sub-pixel/pixel spatial attraction models. *International Journal of Remote Sensing*, 27(15):3293–3310, 2006.

[85] Peng Wang, Yiquan Wu, and Henry Leung. Subpixel land cover mapping based on a new spatial attraction model with spatial-spectral information. *International Journal of Remote Sensing*, 40(16):6444–6463, 2019.

[86] Qunming Wang, Wenzhong Shi, and Peter M Atkinson. Sub-pixel mapping of remote sensing images based on radial basis function interpolation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92:1–15, 2014.

[87] Shangrong Wu, Zhongxin Chen, Jianqiang Ren, Wujun Jin, Wenqian Guo, Qiangyi Yu, et al. An improved subpixel mapping algorithm based on a combination of the spatial attraction and pixel swapping models for multispectral remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 15(7):1070–1074, 2018.

[88] Koen Mertens, Lieven Verbeke, Els Ducheyne, and Robert De Wulf. Using genetic algorithms in sub-pixel mapping. *International Journal of Remote Sensing*, 24(21):4241–4247, 2003.

[89] Xiaohua Tong, Xiong Xu, Antonio Plaza, Huan Xie, Haiyan Pan, Wen Cao, and Dong Lv. A new genetic method for subpixel mapping using hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4480–4491, 2016.

[90] Yanfei Zhong, Yunyun Wu, Xiong Xu, and Liangpei Zhang. An adaptive subpixel mapping method based on map model and class determination strategy for hyperspectral remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3):1411–1426, 2014.

[91] Xiaodong Li, Yun Du, and Feng Ling. Super-resolution mapping of forests with bitemporal different spatial resolution images based on the spatial-temporal Markov random field. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(1):29–39, 2013.

[92] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3341–3350, 2020.

[93] Liangpei Zhang, Ke Wu, Yanfei Zhong, and Pingxiang Li. A new sub-pixel mapping algorithm based on a bp neural network with an observation model. *Neurocomputing*, 71(10-12):2046–2054, 2008.

[94] Andrew J Tatem, Hugh G Lewis, Peter M Atkinson, and Mark S Nixon. Super-resolution land cover pattern prediction using a hopfield neural network. *Remote Sensing of Environment*, 79(1):1–14, 2002.

[95] Xiaofeng Ma, Youtang Hong, Yongze Song, and Yujia Chen. A super-resolution convolutional-neural-network-based approach for subpixel mapping of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(12):4930–4939, 2019.

[96] Yanna Bai, Wei Chen, Jie Chen, and Weisi Guo. Deep learning methods for solving linear inverse problems: Research directions and paradigms. *Signal Processing*, page 107729, 2020.

[97] Feng Ling, Yun Du, Fei Xiao, and Xiaodong Li. Subpixel land cover mapping by integrating spectral and spatial information of remotely sensed imagery. *IEEE Geoscience and Remote Sensing Letters*, 9(3):408–412, 2011.

[98] Peter M Atkinson. Sub-pixel target mapping from soft-classified, remotely sensed imagery. *Photogrammetric Engineering & Remote Sensing*, 71(7):839–846, 2005.

[99] Peng Wang, Liguo Wang, and Jocelyn Chanussot. Soft-then-hard subpixel land cover mapping based on spatial-spectral interpolation. *IEEE Geoscience and Remote Sensing Letters*, 13(12):1851–1854, 2016.

[100] Zushuai Wei, Yizhuo Meng, Wen Zhang, Jian Peng, and Lingkui Meng. Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the Tibetan Plateau. *Remote Sensing of Environment*, 225:30–44, 2019.

[101] Jian Peng and Alexander Loew. Recent advances in soil moisture estimation from remote sensing. *Water*, 9(7):530, 2017.

[102] Yan Jin, Yong Ge, Jianghao Wang, Yuehong Chen, Gerard BM Heuvelink, and Peter M Atkinson. Downscaling AMSR-2 soil moisture data with geographically weighted area-to-area regression kriging. *IEEE Transactions on Geoscience and Remote Sensing*, 56(4):2362–2376, 2017.

[103] Seyed Hamed Alemohammad, Jana Kolassa, Catherine Prigent, Filipe Aires, and Pierre Gentine. Global downscaling of remotely sensed soil moisture using neural networks. *Hydrology and Earth System Sciences*, 22(10):5341–5356, 2018.

[104] Eni G Njoku, William J Wilson, Simon H Yueh, Steve J Dinardo, Fuk K Li, Thomas J Jackson, Venkat Lakshmi, and J Bolten. Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during sgp99. *IEEE Transactions on Geoscience and Remote Sensing*, 40(12):2659–2673, 2002.

[105] Yasir H Kaheil, M Kashif Gill, Mac McKee, Luis A Bastidas, and Enrique Rosero. Downscaling and assimilation of surface soil moisture using ground truth measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 46(5):1375–1384, 2008.

[106] Ujjwal Narayan and Venkat Lakshmi. Characterizing subpixel variability of low resolution radiometer derived soil moisture using high resolution radar data. *Water resources research*, 44(6):W06425, 2008.

[107] Jennifer Pellenq, Jetse Kalma, Gilles Boulet, G-M Saulnier, Scott Wooldridge, Yann Kerr, and Abdelghani Chehbouni. A disaggregation scheme for soil moisture based on topography and soil depth. *Journal of Hydrology*, 276(1-4):112–127, 2003.

[108] Gwangseob Kim and Ana P Barros. Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data. *Remote Sensing of Environment*, 83(3):400–413, 2002.

[109] Jinyoung Rhee, Jungho Im, and Gregory J Carbone. Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data. *Remote Sensing of Environment*, 114(12):2875–2887, 2010.

[110] Prashant K Srivastava, Dawei Han, Miguel Rico Ramirez, and Tanvir Islam. Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application. *Water resources management*, 27(8):3127–3144, 2013.

[111] Wei Xu, Zhaoxu Zhang, Zehao Long, and Qiming Qin. Downscaling SMAP soil moisture products with convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:4051–4062, 2021.

[112] Fengping Wen, Wei Zhao, Qunming Wang, and Nilda Sánchez. A value-consistent method for downscaling SMAP passive soil moisture with MODIS products using self-adaptive window. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2):913–924, 2019.

[113] Yunsong Li, Jing Hu, Xi Zhao, Weiying Xie, and JiaoJiao Li. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 266:29–41, 2017.

[114] Vittala K Shettigara. A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set. *Photogrammetric Engineering and remote sensing*, 58(5):561–567, 1992.

[115] Bin Fang, Venkataraman Lakshmi, Rajat Bindlish, Thomas J Jackson, and Pang-Wei Liu. Evaluation and validation of a high spatial resolution satellite soil moisture product over the continental United States. *Journal of Hydrology*, page 125043, 2020.

[116] Abdalhaleem Abdalla HASSABALLA, Abdul Nasir MATORI, and Helmi Zulhaidi Mohd SHAFRI. Surface moisture content retrieval from visible/thermal infrared images and field measurements. *Caspian Journal of Applied Sciences Research*, 2(AICCE'12 GIZ' 12):182–189, 2013.

[117] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[118] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[119] Wenzhi Liao, Xin Huang, Frieke Van Coillie, Guy Thoonen, Aleksandra Pižurica, Paul Scheunders, and Wilfried Philips. Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter. In *2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4. Ieee, 2015.

[120] P Kwarteng and A Chavez. Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis. *Photogramm. Eng. Remote Sens*, 55(1):339–348, 1989.

[121] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 50(2):528–537, 2011.

[122] Gemine Vivone, Luciano Alparone, Jocelyn Chanussot, Mauro Dalla Mura, Andrea Garzelli, Giorgio A Licciardi, Rocco Restaino, and Lucien Wald. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2565–2586, 2014.

[123] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization. In *2011 3rd workshop on hyperspectral image and signal processing: Evolution in remote sensing (WHISPERS)*, pages 1–4. IEEE, 2011.

[124] Naoto Yokoya, Norimasa Mayumi, and Akira Iwasaki. Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 6(2):419–426, 2012.