Hyperspectral Image Super-Resolution in Arbitrary Input-Output Band Settings

Zhongyang Zhang* Zhiyang Xu* Zia Ahmed† Asif Salekin‡ Tauhidur Rahman*
University of Massachusetts Amherst* University at Buffalo† Syracuse University‡
{zhongyangzha,zhiyangxu,trahman}@cs.umass.edu zahmed2@buffalo.edu asalekin@syr.edu

Abstract

Hyperspectral image (HSI) with narrow spectral bands can capture rich spectral information, but it sacrifices its spatial resolution in the process. Many machine-learning-based HSI super-resolution (SR) algorithms have been proposed recently. However, one of the fundamental limitations of these approaches is that they are highly dependent on image and camera settings and can only learn to map an input HSI with one specific setting to an output HSI with another. However, different cameras capture images with different spectral response functions and bands numbers due to the diversity of HSI cameras. Consequently, the existing machine-learning-based approaches fail to learn to super-resolve HSIs for a wide variety of input-output band settings. We propose a single Meta-Learning-Based Super-Resolution (MLSR) model, which can take in HSI images at an arbitrary number of input bands' peak wavelengths and generate SR HSIs with an arbitrary number of output bands' peak wavelengths. We leverage NTIRE2020 and ICVL datasets to train and validate the performance of the MLSR model. The results show that the single proposed model can successfully generate super-resolved HSI bands at arbitrary input-output band settings. The results are better or at least comparable to baselines that are separately trained on a specific input-output band setting.

1. Introduction

Hyperspectral imaging has proven effective in solving numerous computer vision tasks, including image segmentation, object recognition, material sensing, and surface characterization in different domains like remote sensing, astronomy, materials science, and biology [14, 19, 26, 42]. HSI can capture rich spectral information by capturing bands near multiple peak wavelengths with a narrow bandwidth. However, HSI is often in low spatial resolution [53], as more filters need to be accommodated in the optical sensor mosaic. This drawback hinders the use of HSI for applications that require high resolution (HR) HSI. Many efforts have been devoted to the study of HSI SR. Among all of them, the fusion-based HSI SR recently gains a lot of attention due to its outstanding performance and accessibility of HR RGB. It leverages simultaneously collected HR multispectral image to spatially super-resolve the low resolution (LR) HSI [18, 25, 40, 43, 45, 50]. Both optimization based and deep-learning based approaches have been proposed for fusing high spatial frequency information in RGB with LR HSI to generate HR HSI.

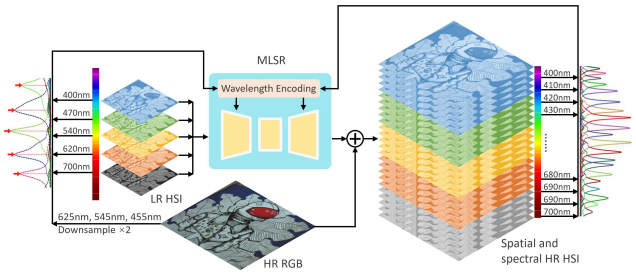


Figure 1: The overall architecture of our proposed MLSR. It takes in an HSI with one bands setting, super-resolve it with RGB guidance, and generate HR HSI with another bands settings. The Red arrows on the left spectral response curves represents the peak wavelengths.

Despite the moderate success of fusion-based HSI SR, it is limited to fixed input-output band settings, making it impractical in real-world settings. Different HSI cameras can image at different peak wavelengths. Hence, existing SR models that are trained to operate with a fixed set of input-output wavelengths are often fairly restrictive for broad application with other HSI cameras [10, 47, 55]. This project proposes a meta-learning-based SR model to address those two limitations. Meta-learning [20] is a broad definition of a wide range of methods that can learn to learn. In this paper, instead of MAML-like methods, "meta" refers to a module in the proposed structure whose weight is the output of another network that takes in other information as input. The question we want to answer in this paper is: "Can a single meta-learning-based machine learning model perform SR on HSI with bands at arbitrary input peak wavelengths and output SR image bands at another arbitrary wavelengths

setting." This work answers this question by demonstrating how a single SR model can be trained on HSIs with different input-output band numbers and peak wavelengths. The core component of MLSR is a meta weight prediction module that maps from wavelengths to the weights of the meta wavelengths convolution layer. The meta block takes the responsibility of fusing and unmixing spectrum information during the encoding and decoding of different input-output bands, which frees the backbone from taking care of both spatial SR and spectrum encoding/decoding.

This paper introduces a new task: RGB-guided HSI super-resolution with arbitrary input-output bands. This task requires flexible input formats, spectral-domain understanding, retrieval at arbitrary spectrum position, and super-resolved reconstruction. We systematically train and evaluate our proposed MLSR on multiple datasets (i.e., NTIRE 2020 and ICVL) and compare it with the state-of-the-art spatial and spectral joint super-resolution approaches [32]. The main highlight of our approach is that only one single trained model can achieve SR at arbitrary input-output HSI bands settings while the standard deep-learning methods need to train one model for each setting. In most cases, the proposed MLSR outperforms baseline methods.

2. Related Works

2.1. HSI Super-resolution

Existing HSI SR algorithms can be broadly classified into two approaches: single HSI SR and MSI-HSI fusion-based SR. The former assumes that no auxiliary MSI image like HR RGB is available and solely operate on the input LR HSI data [1, 5, 17, 21, 28, 33, 44, 49]. Unlike single-image scheme, fusion-based methods take advantage of both the high spatial resolution in the multispectral image (MSI), and the high spectral resolution in the HSI. The core idea is to guide the HSI SR with the high spatial resolution information captured by the MSI images. Fusion approaches include Bayesian-based approaches [2,8,11,31], tensor-based approaches [12, 18, 43, 51, 54], matrix factorization-based approaches [6,13,30,46], and deep-learning-based methods [10,24,47,55,56]. Moreover, researchers have been studying the application of image prior in this domain [24,39].

2.2. Spectral Interpolation

Spectral interpolation has been explored extensively in spectroscopy for estimating the reflectance or transmittance spectra where fine-grained measurement is not available. To this end, several interpolation methods including linear, cubic, PCA-based methods have proven effective for spectral interpolation [23,48]. Furthermore, according to [36], cubic interpolation has been established as an effective and adequate spectral interpolation technique, which shows that a more nonlinear model for interpolation does not boost the performance any further. More recently, deep-learning-based methods have also been proposed for jointly learning spectral-spatial super-resolution [32]. Recent works have also explored RGB to HSI mapping with deep-learning and achieved remarkable performance [4,27,38]. However, they can only work on RGB to produce HSI with specific settings and do not have any input-output flexibility.

2.3. Meta Strategy in Image Super-Resolution

Meta-learning [20] has attracted significant attention in recent years. Weight prediction network or hypernetworks is an essential component of meta-learning [20]. Hypernetworks [7, 15] are neural networks that predict the weight of another neural network based on some auxiliary information or embeddings. Hypernetworks are usually used in multi-task learning and synthesizing predictive models by conditioning an embedding of the support dataset [34, 35]. In this scenario, the target network's weights are not learned during the training process; instead, it is predicted by another network. For example, Hu *et al*. [22] proposed a meta-upsample module for SR tasks and made the upsample scale arbitrary and continuous. Cai *et al*. [9] used a hypernetwork to predict the classifier's weight to migrate to new categories without re-train few-shot learning. Our proposed MLSR also applied this mechanism which will be explained in the next section.

2.4. End-to-End SSSR

There are several works [32, 52] exploring spatial and spectral SR recently. They find that compared to spatial and spectral SR separately, the end-to-end model can avoid spectral distortion and spatial inconsistency and lead to better overall results. Yi's method [52] bases on optimization, while Mei *et al*. [32] focuses on deep-learning models and discusses about the best way to combine spatial and spectral SR stages. As they show us some good results, these methods are fixed to a certain strict setting and cannot be applied on multiple datasets with different settings.

3. MLSR

In this section, we introduce the proposed MLSR. Fig. 2 shows the overall architecture. It mainly consists of five parts: Spectrum-Oriented Feature Extraction (SOFE), Wavelength-2-Weight Network (W2WNet), Spectrum-Oriented Bands Reconstruction (SOBR), a backbone network, and an upsampler.

3.1. Formulation and Motivation

The overall architecture is motivated by performing SR on HSIs with different input settings and output HR bands on demand. A feature extraction layer with fixed parameters does not have the flexibility to handle input with variable channels. To precisely extract features in each band,

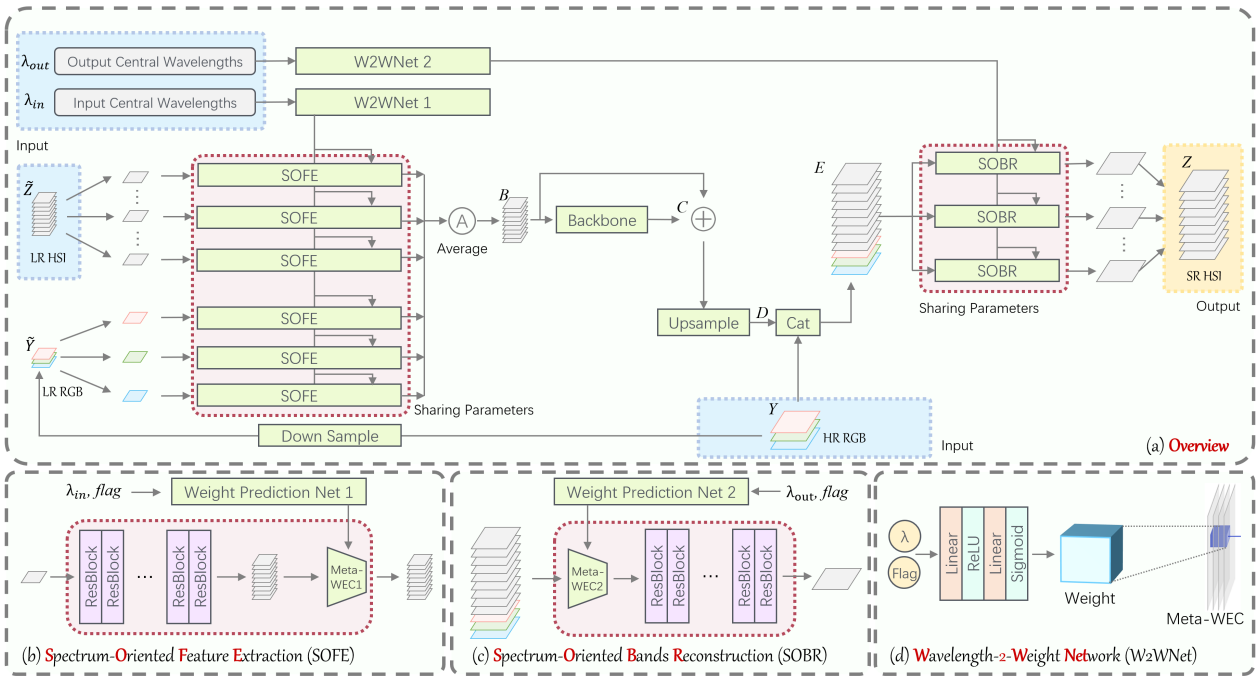


Figure 2: The architecture of MLSR. (a) The overview of the whole model. (b) Structure of SOFE. (c) Structure of SOBR. (d) Structure of W2WNet. MLSR consists of three stages: SOFE in the head, backbone and upsample in the body and SOBR in the tail. Light blue dashed rectangles surround inputs, and the golden dashed rectangle marks the output.

we need a meta module with a full understanding of wavelengths information to predict the feature extractor corresponding to each image band. Each band has a unique spectral response curve. Since the peak wavelengths are the most prominent properties for each band, we use it as a identifier in this work for simplicity. Our design also takes advantage of fusion-based HSI SR by incorporating high-resolution spatial information from HR RGB which guides LR HSI during the SR process. The backbone is formed by residual dense blocks [57] which shows superior performance on RGB SR. Like the feature extraction layer, the output layer also leverages meta-learning to effectively reconstruct an arbitrary number of image bands given the corresponding peak wavelengths.

We denote the low spatial and spectral resolution image as $\tilde{Z} \in \mathbb{R}^{h\times w\times s}$ and the LR RGB image as $\tilde{Y} \in \mathbb{R}^{h\times w\times 3}$ where h,w, and s stand for height, width and number of LR HSI bands. We denote the high spatial and spectral resolution HSI as $Z \in \mathbb{R}^{H\times W\times S}$ and the HR RGB image as $Y \in \mathbb{R}^{H\times W\times 3}$, where H,W, and S stand for height, width and number of HR HSI bands. Moreover, as peak wavelengths of the input and output spectral bands are made variable in our method, they are denoted as $\lambda_{in} \in \mathbb{R}^{s+3}$, where λ_{in} =$[\lambda_{hsi}, \lambda_{rgb}]$, $\lambda_{hsi} \in \mathbb{R}^{s}$, $\lambda_{rgb} \in \mathbb{R}^{3}$, [,] denotes concatenation operation, and $\lambda_{out} \in \mathbb{R}^{S}$. The λ_{hsi} and λ_{rgb} are input wavelengths of the LR HSI and RGB bands.

The objective of MLSR is solving:

$$Z = g(\tilde{Z}, Y, \lambda_{in}, \lambda_{out}|\phi, \theta) \tag{1}$$

where g stands for the overall model inference, ϕ represents the weight of the hypernetwork W2WNet, and θ represents parameters of all other parts in MLSR.

3.2. W2WNet

Given peak wavelengths and bandwidth indicators, W2WNet is to predict the weights of the meta wavelength embedding convolution (Meta-WEC) module. During training, the spectrum information naturally embedded in each image band can be encoded into the parameters of the W2WNet and extracted at inference. This design allows flexibility in the input/output formats. It enables the Meta-WEC to filter out noise from other bands and keep distinct features for a particular band in extraction and reconstruction. The meta module in W2WNet proves the possibility of building a general cross-dataset model for HSI images.

The input of W2WNet is two scalars: corresponding peak wavelength of an HSI band and a binary digit indicating whether this band is from an HSI or RGB image. In our experiment, a 1x1 convolution layer in Meta-WEC can filter out noisy information and extract bands' features. To make the model compact and highly efficient, we decide to use the 1x1 convolution layer. Better performance can be achieved by swapping in a larger module. We use this

simple network to demonstrate the advances of our meta-network. Since the required storage of information is small, we used two fully-connect layers to store the bands' knowledge. Fig. 2(d) shows the structure of the W2WNet, where the "weight" block represents the predicted weights for the Meta-WEC. W2WNet1 and W2WNet2 take input and output wavelengths, respectively.

3.3. SOFE

SOFE generates feature representation from each input band corresponding to its specific spectrum information (i.e., wavelength and bandwidth). It consists of a feature extraction module and a meta wavelength embedding convolution (Meta-WEC) module. Meta-WEC follows a similar Meta mechanism as [22]. The weights of Meta-WEC are the output of W2WNet.

We loop through the input LR HSI bands and apply our feature extractor to each band independently. The feature extractor module consists of eight Residual Blocks. After extracting the generic features, we use the Meta-WEC module, a 2D convolution with a 1×1 kernel, to encode wavelengths information into the feature of each band correspondingly. This operation essentially embeds a band's spectrum information into its feature representation. This design makes the model more robust in terms of input-output formats.

Additionally, as the bands in RGB image has wider bandwidth and embeds spectral information in a wider range than narrow-band HSI bands, following the previous work [21], we pass each band of the paired LR RGB to SOFE in parallel as well. After SOFE, we merge the spectrum-informed feature maps for each band by taking the average to form a general embedding. When we train our model on five input bands, LSTM learns that band 0 is followed by bands 7. However, when we train our model on seven input bands, band 0 is followed by band 5. The LSTM is not able to correctly capture the correlation between image bands.

We can calculate the spectrum-informed feature maps generated by SOFE modules as follow:

$$B = \frac{1}{s+3} \sum_{i=1}^{s+3} Conv(R_0([\tilde{Z}, \tilde{Y}]_i), W_{in}^i) \qquad (2)$$

where $Conv$ denotes Meta-WEC1, $B \in \mathbb{R}^{h \times w \times G}$ stands for the spectrum-informed feature map, W_{in}^i represents the weights of Meta-WEC for each LR HSI or LR RGB band where i is the index of the concatenated LR HSI and LR RGB bands, R_0 means the feature extraction module in SOFE, and G represents the channel number of the feature map extracted in the last step. $Conv$ denotes the Meta-WEC applied on the features of i^{th} band and takes matrix W_{in}^i as its weights. W_{in}^i is the output from W2WNet1 with wavelength λ_{in}^i as its input.

3.4. Backbone and Upsampler

The backbone uses a deep neural network to learn deep features and map LR representation B to HR representation D (figure 2). We use eight Residual Dense Blocks (RDB) proposed by Zhang *et al*. [57] as our backbone module. This module is widely adopted in all kinds of SR tasks and has shown strong performance. We add a skip connection over the whole backbone to resolve the gradient vanishing problem caused by the deep network [16, 57]. We use eight layers of RDB in the backbone, which is half of the layers than the original proposed Residual Dense Network.

We use sub-pixel upsampling layer proposed by Shi *et al*. in [37] as our upsampler. After upsampling, the general embedding C becomes $D \in \mathbb{R}^{H \times W \times G}$, which have the same spacial size as Z. Then D and Y are concatenated together along the channel dimension, and the result feature map E contains both high spectral resolution information from D and high spatial information from Y.

The feature map D is a general embedding which contains the HR bands' knowledge for all covered spectrum wavelength range, and the module SOBR can be viewed as a filter to extract a single band at a given wavelength.

After the upsampling module, the feature map D already has the same spatial size as HR RGB. As HR RGB contains more fine-grained HR information and can provide guidance for the final output, feature map D is concatenated with it before going to SOBR.

3.5. SOBR

SOBR reconstructs the output image bands guided by required spectrum (i.e., wavelength and band) information, backbone generated HR embedding, and high-resolution spatial information from HR RGB. SOBR consists of Meta-WEC2 and a reconstruction module consisting of four Residual Blocks.

SOBR extract features from the upsampled general embedding for each output band given its peak wavelength. W2WNet generates the weights for Meta-WEC2 based on the wavelengths. The output of SOBR is a single band of HR HSI, and SOBR generates S bands for the SR HSI. The operation in SOBR can be represented as follow:

$$Z = Cat_{k=1}^{S} R_1(Conv([D, Y], W_{out}^k)) \qquad (3)$$

where Cat denotes the concatenate operation, $Conv$ denotes Meta-WEC2. D denotes the upsampled output of the backbone module, R_1 stands for the reconstruction module in SOFE, and W_{out}^k is the weight of Meta-WEC2 for k^{th} output band. We apply the reconstruction model after Meta-WEC2 to alleviate the high-level discrepancies in the reconstructed features.

4. Experiments

4.1. Datasets

Most of the popular HSI datasets are poorly organized and lack diversity due to their limited size and object types. Because of its generalizability and flexibility, the proposed MLSR can not be properly trained on those tiny datasets. Thus mainly two datasets are compared on: NTIRE2020 [4] and ICVL [3], both of which contain more than 200 images.

NTIRE2020: The NTIRE 2020 dataset [4] is currently considered as the most comprehensive HSI dataset. The HSIs are captured by a Specim IQ mobile hyperspectral camera. At every 10nm between 400nm and 700nm, an image band is captured by a distinct filter. Thus an image in NITRE has 31 bands with 512×482 spatial dimension.

ICVL: The ICVL dataset was released in 2016 [3]. The hyperspectral images were captured by the Specim PS Kappa DX4 hyperspectral camera. Each original HSI consists of 519 bands from 400nm to 700nm with a step size of 1.25nm. The 519 bands are downsampled to 31 bands, and the step size becomes 10nm. The dimension of processed HSI in the ICVL dataset is $1392 \times 1300 \times 31$.

4.2. Training Details

As this paper focuses on HSI SR with arbitrary input-output settings, the experiments are carefully designed to support the claim. There are three types of experiments in total, which are: 1. Use evenly sampled LR input bands to predict all possible SR bands. 2. Use completely randomly sampled input bands to predict all possible SR bands. 3. Use central LR bands to predict SR bands on the sides. Since the first experiment is more standardized and easier to compare with baseline methods, it is selected as the main experiment, while the other two are studied in the auxiliary experiments subsection. The rest of this subsection are details of the main experiment.

MLSR is trained and evaluated on two datasets, NTIRE2020 and ICVL. For NTIRE2020, we follow the same organization of training and evaluation datasets of the NTIRE2020 competition [4]. For ICVL, 40 images are selected as evaluation set, while the rest are training set.

For each dataset, five sub-datasets are made by evenly sampling five to nine bands from all 31 bands in the original LR HSI images. A series of sub-datasets are made for each dataset to train MLSR as they share the same domain knowledge while yet have different input band number and peak wavelengths.

Since each output band is generated independently, the output bands' peak wavelengths are set to the same as HR HSI, 31 in our case, to let W2WNet2 learn as much output pattern as possible. For one trained model, it can be applied to various input bands' combinations and output HR HSI bands at any requested peak wavelengths. All sub-datasets of a dataset are trained simultaneously and each batch is randomly sampled from one of these sub-datasets. During the evaluation, one sub-dataset is used each time to generate a PSNR/SSIM pair. The SR capability of model is evaluated on multiple scales: $\times 2$, $\times 3$ and $\times 4$.

Each LR input image pair(RGB and HSI) is cropped to a 50×50 patch randomly while keeping the contents of RGB and HSI the same. For data augmentation, each patch is applied 90° rotation, horizontal and vertical flip randomly with 50% possibility. All the models are trained from scratch since the input channel number is not compatible with normal RGB models, and thus pre-train technique cannot be applied. The learning rate is set to 10^{-4} initially and will decrease by half every 20 epochs. Moreover, for loss function, we followed previous works [22, 29] and applied L1 loss for better convergence. The graphic cards we use are eight RTX 2080ti or RTX 1080ti with 11GB of video memory. The corresponding batch size for $\times 2$, $\times 3$ and $\times 4$ are 64, 32, and 32. Each model is trained for approximately 200 epochs, and training takes one and a half days, two days, and two and a half days for these three scales.

4.3. Baseline methods

To the best of our knowledge, there is no HSI super resolution algorithm that can support arbitrary input-output band settings. To compare the performance of MLSR, we carefully developed several baseline methods that have been trained and tested on specific input-output band settings (i.e., 5 or 7 specific input bands to all 31 bands). Baselines can learn a specific input-output setting better since it does not have to learn to do SR in a new setting. In Table 1, baselines can be classified into two categories: two-stage methods and one-stage methods. The former do spectral and spatial SR separately, while the latter do both at the same time.

For deep-learning-based models, SepSSJSR1 and SimSSJSR [32] are selected. SepSSJSR1 is a two-stage method, while SimSSJSR is a one-stage method. They are completely retrained on the same datasets using the same training method during the experiments. Since the two deep-learning-based methods have fixed structures and can only be trained and evaluated on a fixed input setting, one separate model needs to be trained for each band setting, although it is not fair for MLSR.

As similar works are limited even with simplification, besides methods mentioned above, several original baselines are designed as well. Each baseline method's input needs to be flexible to compare impartially, especially the band's number. However, all existing parametric methods, including deep-learning-based ones, cannot provide this flexibility. Therefore, some non-parametric methods have to be selected. They interpolate inputs with various bands number to a fixed higher spectral resolution LR HSI

Stage1	Stage2	Input Bands	NTIRE2020 Scale ×2 PSNR/SSIM	NTIRE2020 Scale ×3 PSNR/SSIM	NTIRE2020 Scale ×4 PSNR/SSIM	ICVL Scale ×2 PSNR/SSIM	ICVL Scale ×3 PSNR/SSIM	ICVL Scale ×4 PSNR/SSIM
Linear	EDSR	5	37.26/0.973	32.92/0.929	31.31/0.897	44.69/0.991	42.80/0.983	42.18/0.978
Linear	RDN	5	38.62/0.981	35.76/0.956	35.41/0.948	44.80/0.992	43.06/0.984	42.43/0.979
Cubic	EDSR	5	35.66/0.960	31.72/0.917	30.33/0.882	43.48/0.990	41.66/0.981	40.77/0.975
Cubic	RDN	5	36.35/0.968	34.05/0.941	33.94/0.931	43.62/0.990	41.47/0.981	40.71/0.976
SepSSJSR1*		5	40.48/0.972	31.68/0.848	32.29/0.853	**48.08**/0.992	44.28/0.980	43.18/0.973
SimSSJSR*		5	37.63/0.949	30.27/0.796	30.23/0.786	44.24/0.980	39.82/0.955	39.27/0.943
MLSR		5	**41.77/0.985**	**39.56/0.966**	**38.07/0.966**	47.89/**0.993**	**46.59/0.992**	**45.94/0.991**
Linear	EDSR	7	38.52/0.976	33.38/0.930	31.56/0.900	45.02/0.993	43.04/0.984	42.37/0.979
Linear	RDN	7	39.30/0.984	36.89/0.959	36.55/0.951	45.34/0.993	43.33/0.985	42.65/0.980
Cubic	EDSR	7	39.06/0.975	33.47/0.930	31.60/0.898	45.34/0.993	43.21/0.984	42.51/0.979
Cubic	RDN	7	40.37/0.984	36.97/0.959	36.70/0.951	45.55/0.993	43.34/0.985	42.72/0.980
SepSSJSR1*		7	41.23/0.973	31.94/0.852	32.51/0.856	**48.85/0.993**	44.31/0.980	43.23/0.972
SimSSJSR*		7	39.13/0.960	30.79/0.812	30.75/0.805	44.83/0.982	41.04/0.959	39.90/0.950
MLSR		7	**42.06/0.985**	**39.57/0.966**	**38.12/0.966**	48.03/0.993	**46.58/0.992**	**46.10/0.991**

Table 1: Evaluation results on NTIRE2020 and ICVL. *Step1* and *Step2* stand for the spectral interpolation and spatial super-resolution method, respectively. * denotes methods that train one model for each setting.

and then send the result to a following spatial SR model to generate the HSI with high spectral and spatial resolution. As a result, those original baselines are all two-stage methods in spectral and spatial SR order.

In the first stage, Cubic and Linear are chosen since they can guarantee the "one pipeline for all settings" target. Each pixel in the input HSI is treated as a spectrum curve, and the value in different bands are some sample points on the curve. The spectral interpolation aims to resample more densely on the curve. In the second stage, a modified version of Enhanced Deep Super-Resolution network (EDSR) [29], and Residual Dense Network (RDN) [57] are applied. Since they do not use HR RGB, for the fairness of comparison, they are both modified to fusion-based methods similarly to MLSR. LR RGB bands are concatenated with LR HSI initially, and the HR RGB bands are concatenated with the output feature map after the upsampling module in the tail. EDSR and RDN are selected because they have shown superior RGB SR performance, and both are widely used as baselines in recent SR papers. Moreover, as the backbone of MLSR is a modified version of RDN, using it as the second stage baseline model can better support the effectiveness of the proposed structure.

4.4. Model Performance

Following the common practice, we evaluate the MLSR on three scales, i.e., $2\times$, $3\times$, $4\times$. For each scale and each dataset, one model is trained and evaluated. For evaluation, peak signal-to-noise ratio (PSNR), and structure similarity (SSIM), which are the two most widely used quantitative picture quality indices (PQIs), are adopted.

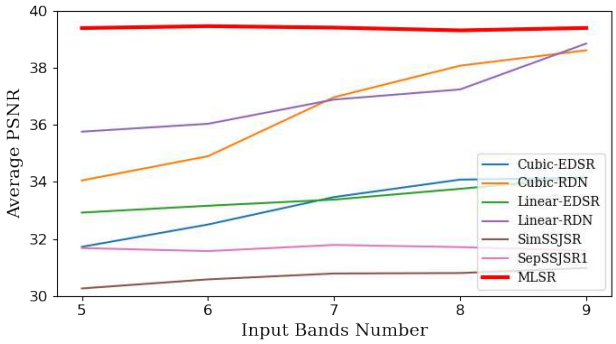


Figure 3: Average PSNR over input band number with scaling factor $\times 3$.

Table 1 shows the evaluation results on MLSR and baseline methods. It is clear that for both datasets and under most settings, MLSR gets the same level or better results than baseline models. Fig. 4 shows two examples of the resulting SR bands comparison and the corresponding residual error map. The figure clearly shows that though the two baselines proposed by [32] are good at predicting low-frequency parts, they are not working well on high-frequency parts like edges, which leads to bad results.

It can be observed that the number of input bands largely influences the performance of baseline methods. With fewer input bands, the spectral-spatial SR becomes more challenging. To evaluate the impact of the input band number on the performance, the input band number varies from 5 to 9 using equidistant sampling. As shown in Fig. 3, the

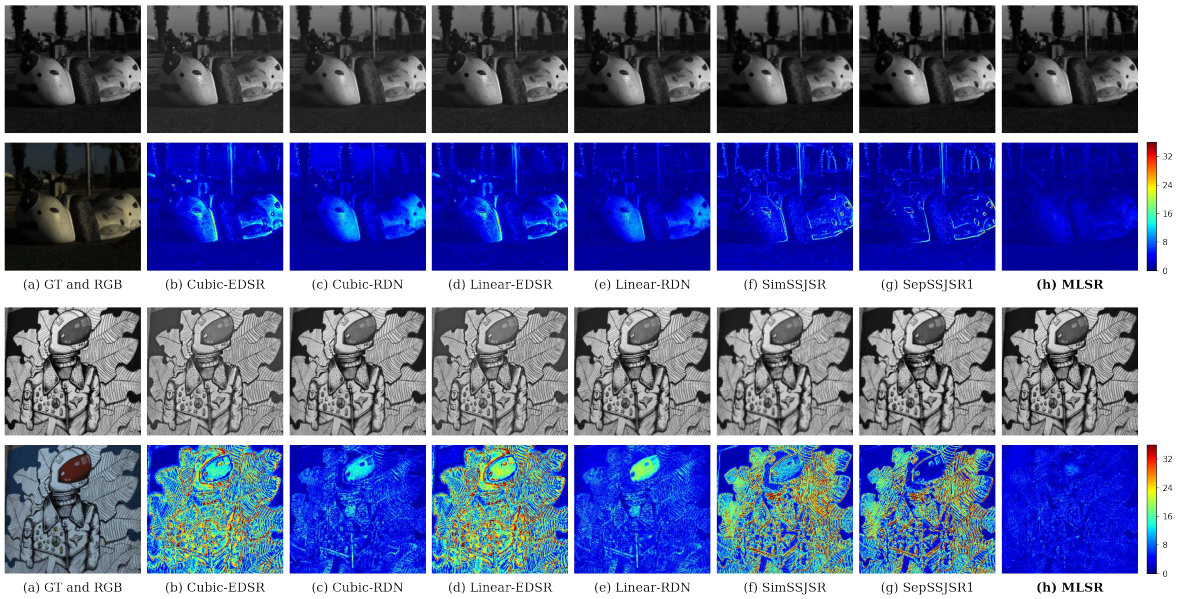


Figure 4: Predicted HSI bands by different methods and their absolute error map. Band 3 is selected in the upper image, band 20 selected in the lower image. The the scaling factor is $\times 3$.

average PSNR of the output HR HSIs from MLSR is much higher than the baselines. As the number of input bands decreases, the PSNRs of the baselines decrease rapidly. However, MLSR only drops slightly and can maintain high performance across the numbers of input bands.

4.5. Auxiliary Experiments: Spectral Extrapolation and Randomized Input Bands

Extrapolate:	Experiment 1	Experiment 2
Baseline	30.26/0.675	29.27/0.671
MLSR	**36.40/0.970**	**35.99/0.970**

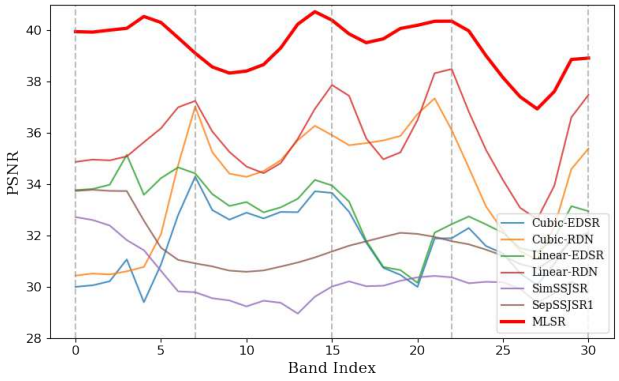
Table 2: The performance (PSNR/SSIM) of baseline and MLSR for spectral extrapolation and spatial SR task in two experimental settings. The scaling factor is $\times 2$.

Since the proposed task requires the model to understand the spectrum and reconstruct HR bands at requested peak wavelengths, it is necessary to know how the model predicts bands whose peak wavelengths are outside of all the training bands' wavelengths. This part is similar to extrapolation, which is generally considered a more challenging problem than interpolation due to the lack of guidance on both sides [41]. Two experiments are designed to evaluate the performance of MLSR on spatial SR with spectral extrapolation on LR HSI. In the first experiment, MLSR is asked to generate the last ten bands using only the first 21 bands of the LR input. The second experiment uses the middle 21 bands of the LR input to generate the five bands on each side of the HR output. Both experiments are done on NTIRE with the scaling factor $\times 2$. MLSR is compared with the best performing baseline: the Cubic+RDN, and Table 2 shows the results. Although compared to spectral interpolation, the performance decreases, MLSR is still much better than the baseline on extrapolation.

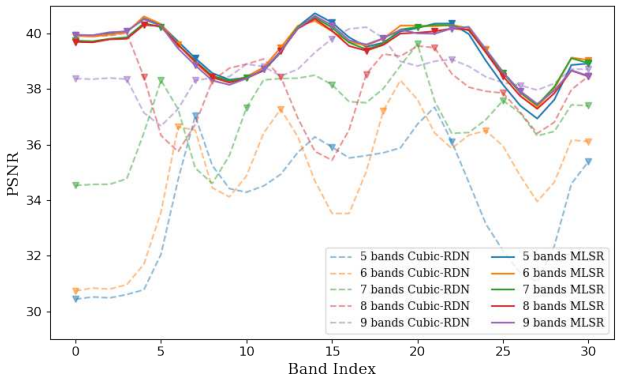
Input Bands Index	Baseline	MLSR
2,5,11,20,28	29.75/0.860	**38.64/0.976**
7,19,21,25,29	23.57/0.649	**38.38/0.975**
8,12,15,20,21	22.27/0.437	**38.02/0.973**
1,4,9,12,19,28	30.91/0.875	**38.63/0.976**
5,6,8,17,19,20,25	28.09/0.663	**38.36/0.975**
6,7,9,10,11,12,15,17,22	24.53/0.587	**38.07/0.973**

Table 3: The performance (PSNR/SSIM) of baseline and MLSR with randomly selected HSI bands. The scaling factor is $\times 2$, and the baseline method is Cubic+RDN.

Another experiment is conducted to evaluate further the model's ability to perform spectral HR reconstruction on more diverse input configurations. The model outputs the whole 31 bands with randomly selected 5 to 9 bands from the spectrum of LR HSI as input. With the random selection, the input bands can cluster together and can be unevenly distributed. As a result, interpolation and extrapolation are conducted at the same time. As shown in Table 3, the performance of MLSR is consistently better than the baseline on all randomly selected input formats with the scaling factor $\times 2$. Moreover, a single MLSR model to perform all the experiments in Table 3 is trained. The results prove that the proposed W2WNet can learn the knowledge



(a) The PSNR of different methods across different bands with LR HSIs with 5 input bands. The gray vertical dash lines highlights the band's indices that is part of the input LR HSI.



(b) Illustration of PSNR curves for different number of input bands for MLSR and Cubic-RDN baseline. The input HSI bands' indices are highlighted by the small triangles with the same color.

Figure 5: Performance of MLSR and baselines across different parts of the spectrum with the scaling factor $\times 3$.

embedded in the spectrum instead of the statistical bias in the predefined wavelengths and settings.

4.6. Performance over Output Bands

So far, it is well proved that the total number of input bands impacts the performance. In this subsection, how the input bands' distribution affect the output bands is studied. Fig. 5a shows the band-wise PSNR of MLSR and baselines with five input bands. The scaling factor is set to $\times 3$. The baseline models find local maxima at the same bands used in the input (highlighted by the vertical dash line in Fig. 5a). As the distance to the nearest input band index increases, there is a fairly sharp decrease in the baseline models' performance. Overall, the performances of the baseline models are dictated by the input LR HSI bands' locations. The proposed MLSR consistently outperforms all the baseline models across all the bands. While the difference in performance between MLSR and the baseline models is relatively low at the input bands' indices, the main advantage of MLSR can be observed at indices further away from these bands. Overall the proposed MLSR gives a relatively more flat response curve and can generate superior spectral-spatial super-resolved HSI at intermediate wavelengths.

Fig. 5b shows the PSNR of MLSR and one top-performing baseline (i.e., Cubic RDN) across all the bands as the number of input band number changes with the scaling factor $\times 3$. Again, the performance of the baseline model is primarily dictated by choice of the input bands, and the shape of the PSNR curve varies significantly as the input band number changes. Surprisingly, the shape of the PSNR curve of the proposed MLSR model almost remains the same as the number of input bands and the input band indices changes. Overall, it indicates that MLSR is agnostic to the input band settings and provides a consistent, flatter, and higher response across the spectrum.

4.7. Ablation Study

To prove the effectiveness of W2WNet and Meta-WEC, Meta-WECs are replaced with an ordinary 1×1 convolution, removes the W2WNet, and keeps all other parts the same. The scaling factor is set to $\times 2$, the input bands number is five, and the dataset is NTIRE2020. The first experiment in Table 4 only replaces the Meta-WEC in SOFE, while the second replaces Meta-WECs in both SOFE and SOBR. Also, to prove that 1×1 kernel is a better choice for Meta-WEC than other settings like widely-used 3×3, the same experiment as above is done with the Meta-WEC kernel's size set to 3×3.

Experiment	Results
Replace Meta-WEC in SOFE	39.13/0.978
Replace Meta-WEC in SOFE and SOBR	27.76/0.917
Replace Meta-WEC 1×1 kernel by 3×3	39.93/0.980
MLSR Original	**41.77/0.985**

Table 4: Ablation study results(PSNR/SSIM).

5. Conclusion

This paper proposes a Meta fusion-based framework for spectral understanding, retrieval, and high-resolution HSI reconstruction. The proposed W2WNet and Meta-WEC modules adaptively learn spectral information in each image band and enrich its feature map. Our model has shown significant flexibility and robustness in generating arbitrary input-output bands within the same domain knowledge datasets. It is the first step toward cross-dataset HSI SR with diverse domain knowledge. Extensive experiments show the superiority of our novel design.

References

[1] T. Akgun, Y. Altunbasak, and R.M. Mersereau. Super-resolution reconstruction of hyperspectral images. *IEEE Transactions on Image Processing*, 14(11):1860–1875, nov 2005. 2

[2] Naveed Akhtar, Faisal Shafait, and Ajmal Mian. Bayesian sparse representation for hyperspectral image super resolution. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2015. 2

[3] Boaz Arad and Ohad Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *Computer Vision – ECCV 2016*, pages 19–34. Springer International Publishing, 2016. 5

[4] Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 446–447, 2020. 2, 5

[5] P. V. Arun, Krishna Mohan Buddhiraju, Alok Porwal, and Jocelyn Chanussot. CNN-based super-resolution of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6106–6121, sep 2020. 2

[6] Ricardo Augusto Borsoi, Tales Imbiriba, and Jose Carlos Moreira Bermudez. Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability. *IEEE Transactions on Image Processing*, 29:116–127, 2020. 2

[7] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Smash: one-shot model architecture search through hypernetworks. *arXiv preprint arXiv:1708.05344*, 2017. 2

[8] Leon Bungert, David A Coomes, Matthias J Ehrhardt, Jennifer Rasch, Rafael Reisenhofer, and Carola-Bibiane Schönlieb. Blind image fusion for hyperspectral imaging with the directional total variation. *Inverse Problems*, 34(4):044003, mar 2018. 2

[9] Qi Cai, Yingwei Pan, Ting Yao, Chenggang Yan, and Tao Mei. Memory matching networks for one-shot image recognition. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, jun 2018. 2

[10] Yi Chang, Luxin Yan, Houzhang Fang, Sheng Zhong, and Wenshan Liao. HSI-DeNet: Hyperspectral image restoration via convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2):667–682, feb 2019. 1, 2

[11] Yi Chang, Luxin Yan, Xi-Le Zhao, Houzhang Fang, Zhijun Zhang, and Sheng Zhong. Weighted low-rank tensor recovery for hyperspectral image restoration. *IEEE Transactions on Cybernetics*, 50(11):4558–4572, nov 2020. 2

[12] Renwei Dian, Leyuan Fang, and Shutao Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5344–5353, 2017. 2

[13] Renwei Dian and Shutao Li. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Transactions on Image Processing*, 28(10):5135–5146, 2019. 2

[14] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013. 1

[15] David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016. 2

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2016. 4

[17] Shiying He, Haiwei Zhou, Yao Wang, Wenfei Cao, and Zhi Han. Super-resolution reconstruction of hyperspectral images via low rank tensor modeling and total variation regularization. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 6962–6965. IEEE, 2016. 2

[18] Wei He, Yong Chen, Naoto Yokoya, Chao Li, and Qibin Zhao. Hyperspectral super-resolution via coupled tensor ring factorization. *arXiv preprint arXiv:2001.01547*, 2020. 1, 2

[19] Hien Van Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 44–51, 2010. 1

[20] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *arXiv preprint arXiv:2004.05439*, 2020. 1, 2

[21] Jing Hu, Xiuping Jia, Yunsong Li, Gang He, and Minghua Zhao. Hyperspectral image super-resolution via intrafusion network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7459–7471, 2020. 2, 4

[22] Xuecai Hu, Haoyuan Mu, Xiangyu Zhang, Zilei Wang, Tieniu Tan, and Jian Sun. Meta-SR: A magnification-arbitrary network for super-resolution. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jun 2019. 2, 4, 5

[23] Cong Phuoc Huynh and Antonio Robles-Kelly. A comparative evaluation of spectral reflectance representations for spectrum reconstruction, interpolation and classification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, jun 2013. 2

[24] Junjun Jiang, He Sun, Xianming Liu, and Jiayi Ma. Learning spatial-spectral prior for super-resolution of hyperspectral imagery. *IEEE Transactions on Computational Imaging*, 6:1082–1096, 2020. 2

[25] Charilaos I. Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma. Hyperspectral super-resolution: A coupled tensor factorization approach. *IEEE Transactions on Signal Processing*, 66(24):6503–6517, dec 2018. 1

[26] C. Kwan, B. Ayhan, G. Chen, Jing Wang, Baohong Ji, and Chein-I Chang. A novel approach for spectral unmixing, classification, and concentration estimation of chemical and biological agents. *IEEE Transactions on Geoscience and Remote Sensing*, 44(2):409–419, 2006. 1

[27] Fayez Lahoud, Ruofan Zhou, and Sabine Susstrunk. Multi-modal spectral image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 2

[28] Yong Li, Lei Zhang, Chen Dingl, Wei Wei, and Yanning Zhang. Single hyperspectral image super-resolution with grouped deep recursive residual network. In *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pages 1–4. IEEE, 2018. 2

[29] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, jul 2017. 5, 6

[30] Jianjun Liu, Zebin Wu, Liang Xiao, Jun Sun, and Hong Yan. A truncated matrix decomposition for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 29:8028–8042, 2020. 2

[31] Laetitia Loncan, Luis B De Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geoscience and remote sensing magazine*, 3(3):27–46, 2015. 2

[32] Shaohui Mei, Ruituo Jiang, Xu Li, and Qian Du. Spatial and spectral joint super-resolution using convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4590–4603, 2020. 2, 5, 6

[33] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, Shuai Wan, and Qian Du. Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, 9(11):1139, nov 2017. 2

[34] Kate Rakelly, Aurick Zhou, Chelsea Finn, Sergey Levine, and Deirdre Quillen. Efficient off-policy meta-reinforcement learning via probabilistic context variables. In *International conference on machine learning*, pages 5331–5340. PMLR, 2019. 2

[35] Andrei A Rusu, Dushyant Rao, Jakub Sygnowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. *arXiv preprint arXiv:1807.05960*, 2018. 2

[36] Norbert Sándor, Tamás Ondró, and János Schanda. Spectral interpolation errors. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(5):348–353, 2005. 2

[37] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. 4

[38] Mehrdad Shoeiby, Antonio Robles-Kelly, Ran Wei, and Radu Timofte. PIRM2018 challenge on spectral image super-resolution: Dataset and study. In *Lecture Notes in Computer Science*, pages 276–287. Springer International Publishing, 2019. 2

[39] Oleksii Sidorov and Jon Yngve Hardeberg. Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*. IEEE, oct 2019. 2

[40] Miguel Simoes, José Bioucas-Dias, Luis B Almeida, and Jocelyn Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2014. 1

[41] Jukka Talvitie, Markku Renfors, and Elena Simona Lohan. Distance-based interpolation and extrapolation methods for rss-based localization with indoor wireless signals. *IEEE transactions on vehicular technology*, 64(4):1340–1353, 2015. 7

[42] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 40(5):1267–1279, 2010. 1

[43] Wei Wan, Weihong Guo, Haiyang Huang, and Jun Liu. Nonnegative and nonlocal sparse tensor factorization-based hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12):8384–8394, 2020. 1, 2

[44] Qi Wang, Qiang Li, and Xuelong Li. Spatial-spectral residual network for hyperspectral image super-resolution. *arXiv preprint arXiv:2001.04609*, 2020. 2

[45] Qi Wei, José Bioucas-Dias, Nicolas Dobigeon, and Jean-Yves Tourneret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7):3658–3668, 2015. 1

[46] Qi Wei, José Bioucas-Dias, Nicolas Dobigeon, Jean-Yves Tourneret, Marcus Chen, and Simon Godsill. Multiband image fusion based on spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7236–7249, 2016. 2

[47] Wei Wei, Jiangtao Nie, Yong Li, Lei Zhang, and Yanning Zhang. Deep recursive network for hyperspectral image super-resolution. *IEEE Transactions on Computational Imaging*, 6:1233–1244, 2020. 1, 2

[48] Stephen Westland. Interpolation of spectral data. *Encyclopedia of Color Science and Technology*, pages 794–797, 2014. 2

[49] Weiying Xie, Xiuping Jia, Yunsong Li, and Jie Lei. Hyperspectral image super-resolution using deep feature matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):6055–6067, 2019. 2

[50] Yang Xu, Zebin Wu, Jocelyn Chanussot, Pierre Comon, and Zhihui Wei. Nonlocal coupled tensor cp decomposition for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 58(1):348–362, 2019. 1

[51] Yang Xu, Zebin Wu, Jocelyn Chanussot, and Zhihui Wei. Nonlocal patch tensor sparse representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 28(6):3034–3047, jun 2019. 2

[52] Chen Yi, Yong-qiang Zhao, Jonathan Cheung-Wai Chan, and Seong G Kong. Joint spatial-spectral resolution enhancement of multispectral images with spectral matrix factorization

and spatial sparsity constraints. *Remote Sensing*, 12(6):993, 2020. 2

[53] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, 2017. 1

[54] Kai Zhang, Min Wang, Shuyuan Yang, and Licheng Jiao. Spatial–spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1030–1040, 2018. 2

[55] Lei Zhang, Jiangtao Nie, Wei Wei, Yong Li, and Yanning Zhang. Deep blind hyperspectral image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 2020. 1, 2

[56] Tao Zhang, Ying Fu, Lizhi Wang, and Hua Huang. Hyperspectral image reconstruction using deep external and internal learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8559–8568, 2019. 2

[57] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 3, 4, 6