

# Multi-Modal Spectral Image Super-Resolution

Fayez Lahoud\*, Ruofan Zhou\*, and Sabine Süsstrunk

School of Computer and Communication Sciences  
École Polytechnique Fédérale de Lausanne  
{ruofan.zhou,fayez.lahoud,sabine.susstrunk}@epfl.ch

**Abstract.** Recent advances have shown the great power of deep convolutional neural networks (CNN) to learn the relationship between low and high-resolution image patches. However, these methods only take a single-scale image as input and require large amount of data to train without the risk of overfitting. In this paper, we tackle the problem of multi-modal spectral image super-resolution while constraining ourselves to a small dataset. We propose the use of different modalities to improve the performance of neural networks on the spectral super-resolution problem. First, we use multiple downscaled versions of the same image to infer a better high-resolution image for training, we refer to these inputs as a multi-scale modality. Furthermore, color images are usually taken at a higher resolution than spectral images, so we make use of color images as another modality to improve the super-resolution network. By combining both modalities, we build a pipeline that learns to super-resolve using multi-scale spectral inputs guided by a color image. Finally, we validate our method and show that it is economic in terms of parameters and computation time, while still producing state-of-the-art results.<sup>1</sup>

**Keywords:** Spectral Reconstruction, Spectral Image Super-Resolution, Residual Learning, Image Completion, Multi-Modality

## 1 Introduction

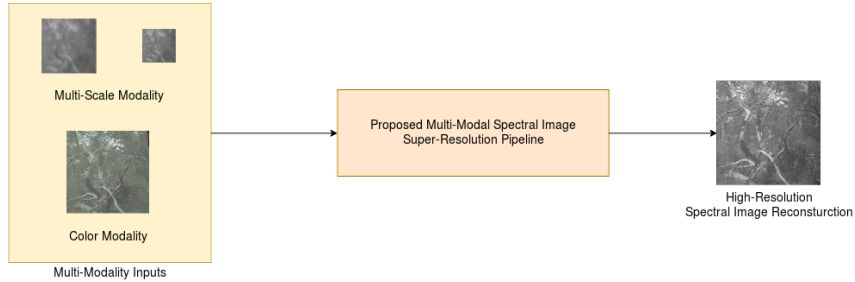
In this paper, we address spatial image super-resolution for spectral images. We tackle the problem posed by the PIRM2018 Spectral Image Challenge [20, 19] for reconstructing high-resolution spectral images from twice (LR2) and thrice (LR3) downscaled versions. The challenge has two tracks. The first (Track1) asks to super-resolve from only the spectral low-resolution images, and the second (Track2) provides a guided super-resolution challenge using a high-resolution 3-channel color image in addition to the low-resolution spectral data. Both tracks contain a small number of images, so one of the main obstacles in this challenge is to improve the generalization of the algorithms on a limited dataset.

Single-image super-resolution is an active research area with a wide range of applications in areas such as astronomy, medical imaging, or image enhancement.

---

\* Both authors contribute equally to this work.

<sup>1</sup> Code at <https://github.com/IVRL/Multi-Modal-Spectral-Image-Super-Resolution>.



**Fig. 1.** The proposed framework: our super-resolution algorithm is able to reconstruct high-quality, high-resolution spectral images by taking advantage of multi-modal data consisting of multi-scale spectral images and color images.

The goal is to infer, from a single low-resolution (LR) image, the missing high frequency content that would correspond to its high-resolution counterpart (HR). The problem itself is inherently ill-posed since there are multiple reconstructions that could lead to the same low-resolution observation.

Deep learning involves the design of large scale networks for a variety of image reconstruction problems. To this end, deep neural networks were applied to the super-resolution task. For example, in Dong *et al.* [4], the training set included LR inputs and their corresponding HR output images, where the inputs are upsampled to the correct resolution using bicubic interpolation. The network only takes one low-resolution image with a fixed downscaling factor as input. Here, we use an image completion algorithm [1] to fuse low-resolution spectral images with different downscaling factors to reconstruct a better upsampled input.

In addition, SRCNN [4] has other limitations such as slow convergence and a small receptive field because of its shallow architecture. Deep residual learning [9] was initially proposed to solve the performance degradation as network depth increases, and has shown to increase accuracy on image classification and object detection methods. Here, we use residual learning to reconstruct the residuals between the LR and HR images, rather than learning how to rebuild the HR image from LR. Our assumption is that learning the residual mapping is much easier than learning the original HR image. Furthermore, multiple image restoration tasks such as VDSR [13], DnCNN [25], and DWSR [7] use residual connections from the input to the output and reduce their training time through faster convergence. By combining the image completion upscaling method with residual learning, we build a model suited for multi-scale image super-resolution.

One often can obtain a high spatial resolution panchromatic image accompanying the multi-spectral low resolution image. The fusion of both images allows obtaining both high spatial and spectral resolution images. This is helpful for many remote-sensing applications like agriculture, earth exploration, and astronomy. We make use of a 3-color RGB high spatial resolution image to guide the super-resolution of the 14-band low-resolution spectral images in Track2 of the

challenge. Thus, we design our pipeline to incorporate the guiding images to achieve higher performance on top of our previous residual network results.

In this paper, we propose an efficient framework for multi-modal spectral image super-resolution shown in Fig. 1. The main contributions of this paper are the following: 1) We build a residual learning network suitable for super-resolution due to the sparse nature of the problem. 2) We design a data preprocessing approach that can fuse multi-scale images in order to create an upscaled input image to the network. This approach combines the information from multi-scale modalities with an image completion algorithm to provide a candidate image to the network that performs better than the typical bicubic interpolation. 3) We build a two-stage pipeline for guided super-resolution under consideration that very few data samples containing guiding information are available. The framework resembles transfer learning, as it allows to transfer information learned using one modality to another to compensate the lack of data.

## 2 Related Work

Single-image super-resolution corresponds is about upscaling a single low-resolution image to a higher spatial resolution. Typically, the image is in grayscale (1-channel) or in color (3-channel). This field has been studied for decades, so a large amount of literature exists. While early methods attempted to construct an efficient upscaling function using image statistics, recent trends have shown that learning to super-resolve using CNNs has a better performance than prior techniques [4, 11, 12, 16]. The architecture of the network affects the performance, as well as the loss function used. For instance, the authors in [26] have shown that L2 loss doesn't give the best PSNR results even though they are directly related.

Our work has some relation to the conventional problem of single-image super-resolution, however it is done for images with high spectral resolution (14 channels). While this does not change the nature of the problem, the fact that we are fusing multi-scale inputs and predicting on a larger number of channels requires adapting the model and loss functions to account for these factors.

Due to hardware limitations, high spectral resolution images come at the cost of lower spatial resolution. To mitigate this problem, they are often combined with higher spatial but lower spectral resolution images. Previous works [22–24] used statistical methods to mix spatial information from the high-spatial low-spectral resolution image with the color information from the multi-spectral bands. However, it is expensive and time-consuming to generate a large set of registered spectral and color images. To cope with the limited training data, a model can be trained on a large but related dataset, and then adapted to perform on the smaller given dataset. Prior work on domain adaptation [3, 6, 8] show the merit of these techniques to handle small or difficult-to-label datasets. Similarly, we use our original framework for super-resolving the multi-spectral images, and then use a small residual network to refine the result through a color image guide, which requires significantly less training examples compared to the whole model.

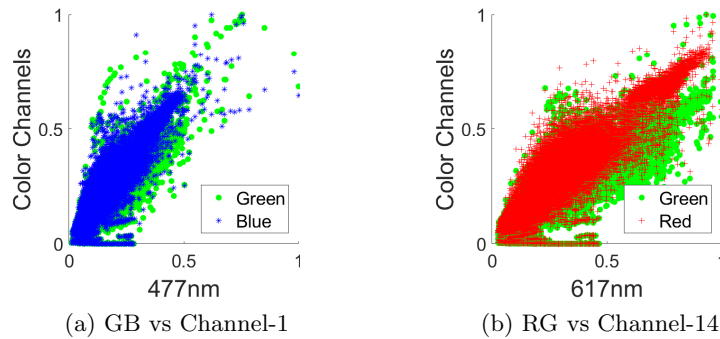
### 3 PIRM2018 Challenge

We use a dataset from PIRM2018 Spectral Image Challenge [20, 19]. The dataset consists of two tracks: Track1 contains 240 spectral images and Track2 contains 130 different image stereo pairs of spectral images and their corresponding aligned color images. The datasets are split according to Table 1, the test ground-truth is not available for download, so we report and compare on the validation dataset.

**Table 1.** PIRM2018 Spectral Image Challenge Dataset.

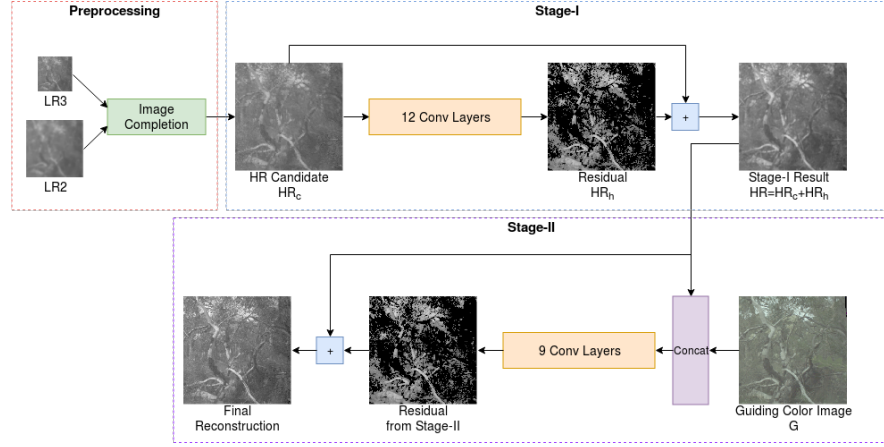
Track	Training	Validation	Test
1	200	20	20
2	100	10	20

For Track1, each data sample  $i$  contains a triplet of 14-channel images  $C_I^i = (HR^i, LR2^i, LR3^i)$ , where  $HR^i$  is the high resolution ground-truth image, and  $LR2^i$  and  $LR3^i$  are the low resolution images obtained by 2 and 3 times downscaling, respectively. The downscaling technique used in this dataset is nearest-neighbors downscaling, *i.e.*, the pixels in the low resolution images are taken at alternating indices from the original image. Even though the 3 times downsampled signal contains less information, it still can cover part of the missing information from the 2 times downsampled signal. This implies that we can make use of a combination of multi-scale downsampled images to obtain a better representation of the high resolution version.



**Fig. 2.** Statistical analysis on different input modalities.

Track2 provides the same information as Track1 with an additional color guiding image  $G^i$  of the same size as the high-resolution spectral image, giving us data samples of the form  $C_{II}^i = (HR^i, LR2^i, LR3^i, G^i)$ . The same downscaling



**Fig. 3.** Illustration of our proposed stacked residual learning framework for spectral image super-resolution. It contains three steps: preprocessing, Stage-I, and Stage-II. Image completion is done in preprocessing to generate a HR candidate. Then Stage-I reconstruct the HR using a 12-layer residual learning network. Stage-II refines Stage-I results using guiding color image  $G$  through a 9-layer residual learning network.

technique is used here. The color image is a 3-channel image already registered to its spectral counterpart with the same resolution as the target high resolution image. The registration is done using FlowNet [5]. Fig. 2 show the distributions of pixel values from the first and last channel of the spectral image with respect to the color channels from the guide. The first demonstrates the correlation between channel-1 (close to blue) with respect to the green and blue channels from the color guide, and the second shows the correlation between channel-14 (close to orange) and the red and green channels. The correlation of these values indicate that the color channels can help predict the spectral pixel values. Both plots have multiple color pixels with zero value, this is due to the image warping done by the registration algorithm.

## 4 Method

We propose a residual learning framework for multi-modal spectral image super-resolution as shown in Fig. 3.

Similar to bicubic interpolation adopted in many super-resolution algorithms [13, 4], we first upscale the low-resolution spectral inputs LR2 and LR3, which are subsampled from the full resolution spectral image by a factor of 2 and 3. We use an image completion algorithm [1] on the multi-scale inputs to generate a high-resolution spectral image candidate with the desired size. Then we train residual learning networks for spectral image super-resolution.

For Track1, Stage-I uses one 12-layer residual learning network to reconstruct high-resolution results from the image candidate. These reconstructions are used



**Fig. 4.** Illustration of downscaling and upscaling.

to generate the solution for Track1. In Track2, we have less training data. So we design our solution to take advantage of Stage-I. Stage-II takes the concatenation of Stage-I’s proposed output and the higher-resolution color image as inputs. It is trained on the small dataset of image pairs and refines Stage-I results through guiding color images.

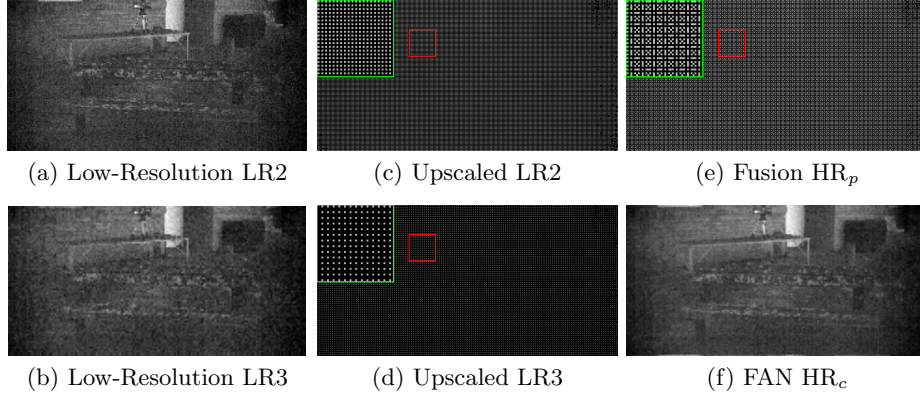
#### 4.1 Image Completion

LR2 and LR3 are both obtained by downscaling the original HR version using nearest-neighbor downscaling. Therefore, a large amount of pixel information is preserved, which means we can already recover part of the ground-truth immediately from the low-resolution samples. In fact, we can recover  $\frac{1}{4}$  of the data from LR2 and  $\frac{1}{9}$  from LR3 by simply upscaling the image and setting the new pixels to black (unfilled). Together, LR2 and LR3 give us  $\frac{1}{3}$  of the original image pixels. Figure. 4 shows how we recover the partial high-resolution image, named  $HR_p$ , from both low-resolution examples.

Image completion is the task of completing an image with a percentage of pixels missing. This has a wide range of applications such as noise-removal, demosaicing, inpainting, artifact removal as well as image editing. One particular usage is image-scaling and super-resolution. There have been multiple approaches to fill the missing parts of an image.

One main category of methods relies on matrix completion [10, 15, 17]. While these methods are well suited for large number of retained pixels, they do not work when the input matrix has fully missing columns and rows such as ours. We also do not have many connected pixels to form patches, so patch-based methods are not suited [14, 21].

The extreme image completion [1] method FAN is able to complete a 1% pixel image with low computation time, and returning visually interpretable images. FAN relies on an efficient implementation of a modified truncated Gaussian filter. The sparse image is filtered with a Gaussian to interpolate missing entries with Gaussian weights assigned to available pixels in a window surrounding the missing entry, on which the Gaussian filter is centered. The modification is that



**Fig. 5.** Illustration of Image Completion on channel 1 of one example from the validation set: (a-b) are the low-resolution images, (c-d) their upscaled version, (e) the fusion of both upscaled versions and (f) the image completion result. Images (c-e) have been gamma corrected for visual clarity.

the Gaussian weights are adjusted to account for the number of locally available pixels.

We use FAN to obtain our input  $HR_c$ . Note, that we keep the ground truth pixels in  $HR_c$  even though FAN outputs different values for them. Fig. 5 shows the steps to obtain the completed image from both inputs.

#### 4.2 Stage-I: Residual Learning

The input of our Stage-I network  $HR_c$  is a low-frequency estimation with partially correct high-frequencies  $HR_h$ . Thus we can formulate it as  $HR_c = HR - HR_h$ , where  $HR_h$  contains information of the high-resolution spectral image, such as textures and edges. We adopt a residual learning formulation to train a residual mapping  $f(HR_c) = HR_h$ . The architecture of the residual learning network is shown in the Stage-I part of Figure. 3. By adopting residual learning, the network only learns to predict the high-frequency details without preserving all low-frequency details. This allows us to use a smaller model and train faster than conventional CNN methods. In our residual learning network (Stage-I) for spectral image super-resolution, we use 12 convolutional layers of the same setting except for the last layer: 64 filters of size  $3 \times 3$  and followed by a ReLU activation. The last layer for generating residual images, consists of 14 filters of size  $3 \times 3$ .

As shown in [26], the loss function in an image restoration task is very important when the resulting image is going to be shown to a human observer. Typical losses include the  $L1$  and  $L2$  distance measures. However, these methods are not well suited to deal with multi-spectral data. The spectral information divergence [2] (SID) compares the similarity of two pixels by measuring the discrepancy between their spectral signatures. This measure has been widely used

in hyper-spectral data processing. By defining the relative entropy of the prediction  $P$  with respect to the ground-truth  $G$  containing  $N$  pixels as:

$$D(P||G) = \sum_{i=0}^N P_i \log\left(\frac{P_i}{G_i}\right) \quad (1)$$

The SID can then be defined as the symmetric sum of both relative entropy measures:

$$\text{SID} = D(P||G) + D(G||P) \quad (2)$$

Additionally, the pixel values are in the range  $[0, 65536]$ , so a relative error measure is well suited to reduce the large error that an absolute measure could have at the higher end of the range. The mean relative absolute error (MRAE) does exactly that by punishing errors relative to the value of the ground-truth. The MRAE is calculated as:

$$\text{MRAE} = \left| \frac{P - G}{G} \right| \quad (3)$$

To better optimize along both metrics, we use a loss function of a sum of MRAE and SID to train our network:

$$\text{Loss} = \text{SID} + \text{MRAE} \quad (4)$$

### 4.3 Stage-II: Color Guided Super-Resolution

We propose a further improvement by using registered pairs of spectral and color images. In fact, mixing information from both modalities allows obtaining both high spatial and spectral resolution images. However, due to the difficulty of obtaining a large set of registered image pairs, we introduced a transfer learning method built on top of the previous residual network. We build a new residual learning network that takes as input the previous super-resolved image (obtained from Stage-I) concatenated with a 3-channel color image. The new network acts as a fine-tuner for the super-resolution based on the new color data accompanying its input. The network architecture is shown in Stage-II part of Fig. 3. Here we use 8 convolutional layers with 64 filters of size  $3 \times 3$  each followed by a ReLU activation, and we use a final convolutional layer with 14 filters of size  $3 \times 3$  to produce the residual image. We use the same loss function to train this network as discussed above.

## 5 Experiments

### 5.1 Comparative Results

We train the two stages separately. For Stage-I, we use spectral patches of size  $96 \times 96$  with a stride of 24 cropped from the fused LR2 and LR3 images following the described image completion scheme. We use spectral images from both tracks



to obtain a larger training set for Stage-I. We use Adam for optimizing the network with weight decay =  $1e - 5$  and a learning rate of 0.001. We decay the learning rate by 10 every 30 epochs. We set the minibatch size to 64. After Stage-I converges, we use the Track2 dataset for training Stage-II. We crop  $48 \times 48$  overlapping patches with a stride of 16 from the dataset. We use the same training strategy as Stage-I for Stage-II. We use a sum of SID and MRAE for the loss function for the training of both stages.

Evaluation with MRAE, SID and PSNR metrics is conducted on two validation sets: Validation-I includes 20 spectral images and Validation-II includes 10 pairs of spectral images and corresponding guided color images. Note that as Validation-I does not have a guiding color image as input, only results from Stage-I are shown.

**Table 2.** Test results on Validation-I. The bold values indicate the best performance.

Metric	Bicubic	Stage-I Results	EDSR
MRAE	0.11	<b>0.08</b>	0.10
SID	57.39	<b>43.48</b>	43.57
PSNR	36.07	<b>37.44</b>	37.27

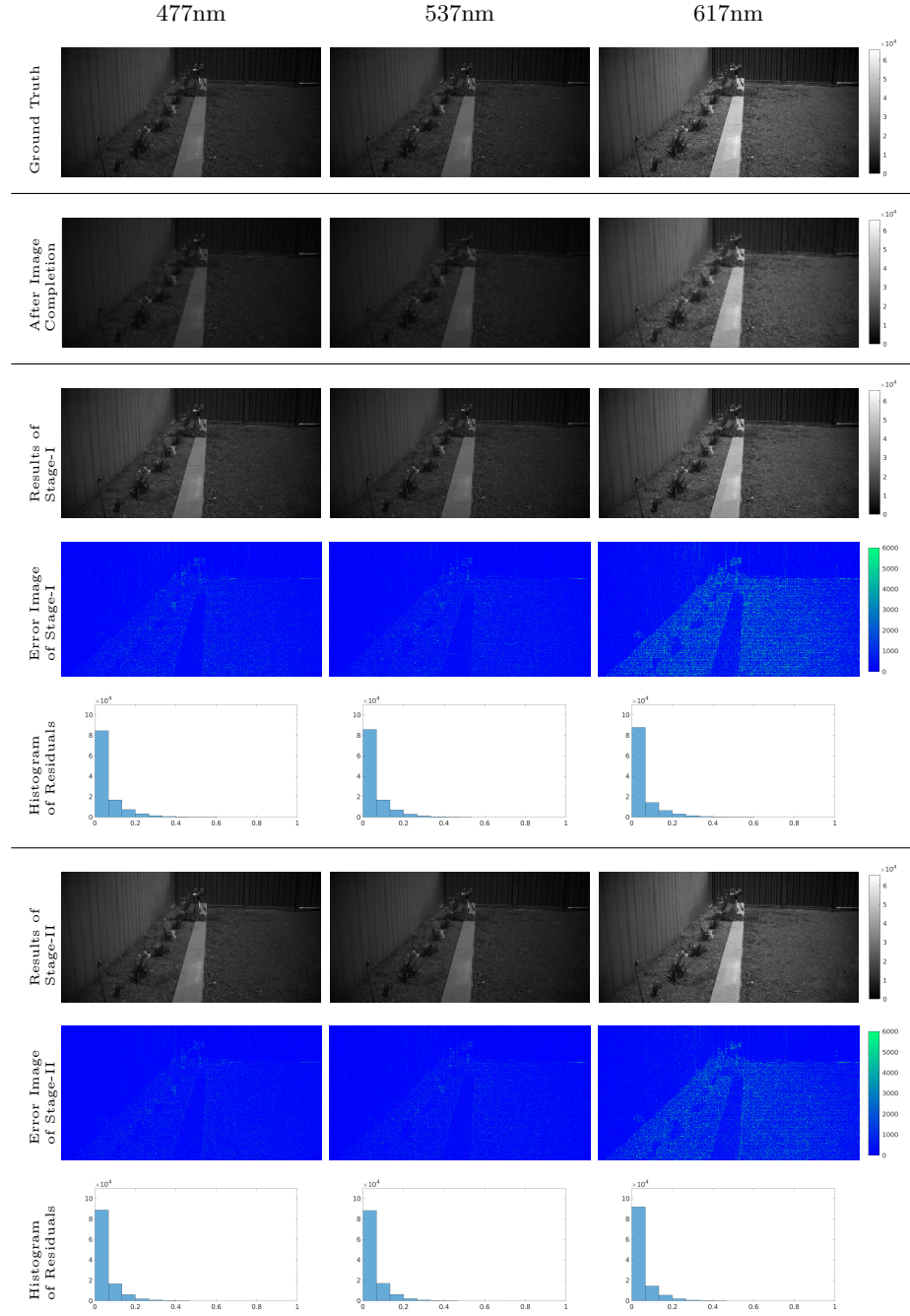
**Table 3.** Test results on Validation-II. The bold values indicate the best performance.

Metric	Bicubic	Stage-I Results	Stage-II Results	Residual Net	EDSR
MRAE	0.13	0.10	<b>0.09</b>	0.23	0.16
SID	43.32	38.04	<b>24.51</b>	36.29	30.67
PSNR	36.48	37.02	<b>39.17</b>	36.62	37.13

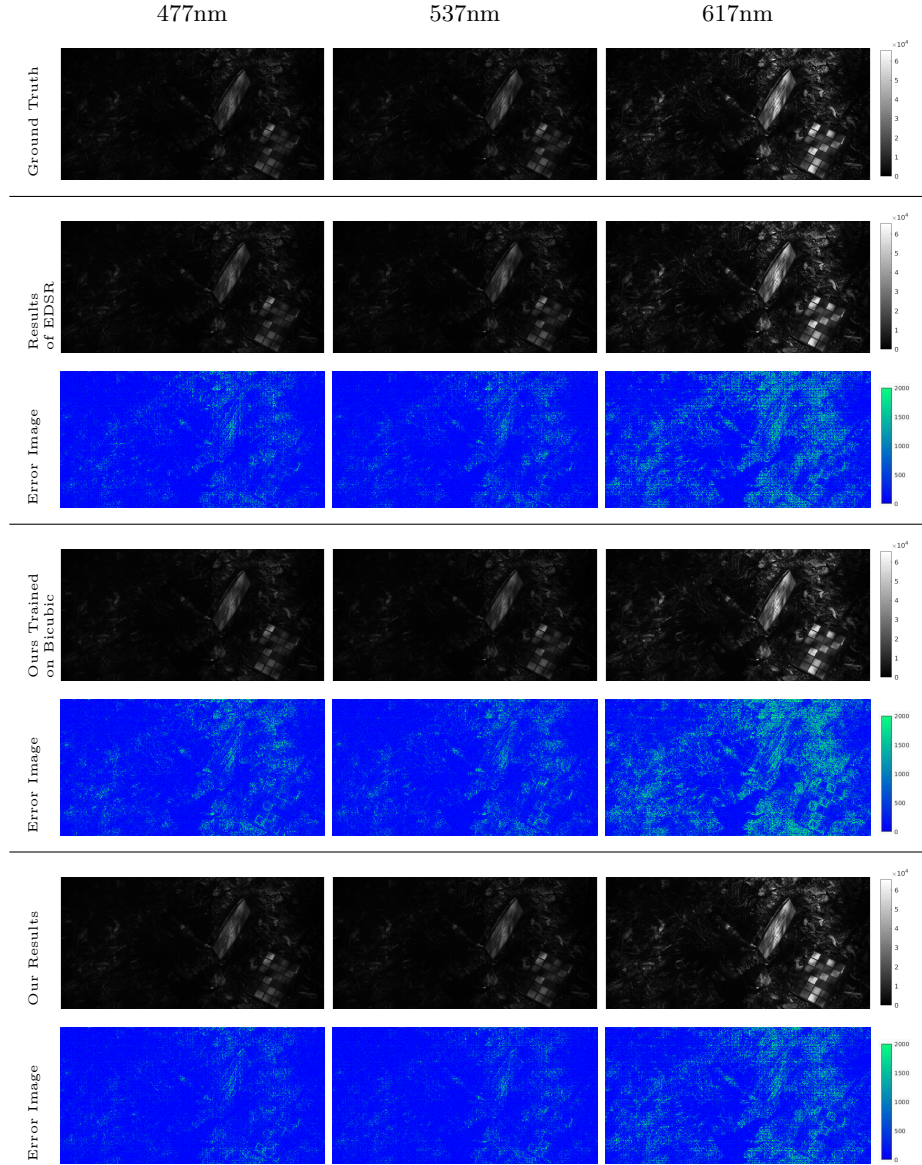
Table 2 shows the results on Validation-I, we compare our image completion method by training the same architecture on inputs from bicubic upscaled images taken from LR2. Our image completion input outperforms this commonly used upscaling method on all metrics. This also applies to the Validation-II dataset.

We show an example of results from different stages of our pipeline on Validation-II in Fig. 6. The error images in Fig. 6 clearly show that with the help of guiding color image, Stage-II is able to improve the results from Stage-I.

We display the comparison with other methods on Validation-II are displayed in Table 3. To show the merit of our transfer learning model, we train a residual learning network [13] and the state-of-the-art super-resolution network EDSR [16] using both spectral images (after applying image completion on LR2 and LR3) and guiding color images as inputs. For the residual network, we use 21 convolutional layers to obtain the equivalent size of our stacked stages. We set all convolutional layers of the residual network with a configuration of 64 filters of size  $3 \times 3$  and ReLU activation except the last layer which has 14 filters of size  $3 \times 3$  with no activation function. For EDSR, we use the same configuration



**Fig. 6.** Example of results from different stages. Error images show the absolute difference from our reconstruction to the ground truth spectral image. The histograms of residuals show the histogram of related absolute errors on the error images.



**Fig. 7.** Visual comparison of results from different methods: EDSR and our method trained on bicubic interpolated inputs and the completed HR candidates. Error images show the absolute difference from our reconstruction to the ground truth spectral image.

as the original paper except we ignore the Pixel Shuffle (since we already use an upscaled input) layer [18] and modify the last layers to have 14 filters to

reconstruct the 14-band spectral image. EDSR has 32 residual blocks with 256 filters for each convolutional layers. We train both networks using only Stage-II dataset, and we also do image completion before feeding LR2 and LR3 inputs to the networks. All networks are trained for 300 epochs. Although trained without guiding color images, our Stage-I gives slightly better results than the residual network and EDSR trained on pairs. With guiding color images, Stage-II gains significant improvements on all three metrics.

We also show in Fig 7 the visual comparison of EDSR [16] and our method trained on bicubic interpolated input and the completed HR candidates. The error images show that our method outperforms the other two methods.

In addition to performance, we also evaluate the memory and time consumption of the proposed model. For a  $240 \times 480$  spectral image (with LR2 size of  $120 \times 240$ ), our method only takes 0.5 seconds (0.3 seconds on Stage-I and 0.2 seconds on Stage-II) and 800MB memory on Titan X GPU. While for EDSR, it takes 1.1 seconds and 8000MB memory on the same device.

## 5.2 Ablation Studies

We run ablation studies on our Stage-I network to study how different factors affect the architecture’s performance. First, we study the effect of using different upscaling factors together and alone. Second, we study the effect of the depth on the network on its ability to generalize. Finally, we experiment with changing the loss metrics between MRAE, SID and their sum.

In all the experiments, we train the same residual network with the previously stated configurations, while varying only the one factor in question. We use Adam for optimizing the network with weight decay  $= 1e - 5$  and a learning rate of 0.001. We decay the learning rate by 10 every 20 epochs, and we train all the networks for 100 epochs. We report our results on the Track1 validation set.

**Upscaling Factors** In this section, we change the input of the network to understand how different scales affect its performance. We separate the LR2 and LR3 images, and create image completions from each one of these and train two networks separately using those inputs. Both networks are using the sum of MRAE and SID as loss function. We compare both of them against the original network trained on the completed LR2 and LR3 images together. Table 4 shows the performance of each network given different inputs. All networks achieve the best performance on the type of input they were trained on, we use those values to compare across models. The completed LR2 includes more original pixels than the completed LR3, the network trained on LR2 outperforms the network trained on LR3. Naturally, the network trained on image completion on both LR2 and LR3 obtains better results than the network trained on LR2 only. This also demonstrates that although LR3 has a lower resolution than LR2, it contains extra original pixels that help to reconstruct a higher-quality high-resolution spectral image.

**Table 4.** Test results on Validation-I. The rows represent the type of input the networks were trained on, the columns show the results on inputs taken with different downscaling factors. The bold values indicate the best performance.

	LR2		LR3		LR2 + LR3	
	MRAE	SID	MRAE	SID	MRAE	SID
LR2	0.12	55.50	0.25	192.0	0.12	62.97
LR3	0.16	106.24	0.18	117.17	0.14	123.63
LR2+LR3	0.13	58.10	0.24	178.04	<b>0.10</b>	<b>47.20</b>

**Depth Effect** We study the effect of the depth on the network accuracy and generalization. We empirically determine the best depth for the residual network architecture on the Stage-I problem. We vary the depth between 8 and 16 by steps of 2 and report the progress of this networks during training, as well as their best performances on the validation set. Table 5 shows the metrics for these 5 networks. We can see that at depth 12, we obtain the best performance in terms of MRAE and PSNR.

**Table 5.** Test results on Validation-I based on network depth. Numbers in the header row indicate the number of convolutional layers.

Metric	8	10	12	14	16
MRAE	0.11	0.11	<b>0.10</b>	0.11	0.12
SID	47.27	<b>46.94</b>	47.20	47.26	50.44
PSNR	35.07	35.13	<b>35.15</b>	35.05	31.21

**Table 6.** Test results on Validation-I based on loss metric. Metrics in the header row indicate the loss used during the training of the network. All networks have a similar structure.

Metric	MRAE	SID	MRAE+SID
MRAE	<b>0.09</b>	0.11	0.10
SID	87.75	<b>47.20</b>	<b>47.20</b>
PSNR	31.14	34.94	<b>35.15</b>

**Loss Metrics** In this section, we train multiple residual networks with the same parameters using different loss functions. We train with only MRAE, only SID, and a combination of both. We show that using both provides better super-resolved spectral images than using a single metric. Table 6 shows the results from these three models. While the network trained on MRAE only outperforms

the others on the MRAE metric, its results have a high SID loss. Combining both MRAE and SID losses during training gives the best of both metric results while also scoring high on PSNR.

## 6 Conclusion

Our work presents a spectral super-resolution technique based on the fusion of information from multiple sources. First, we introduce an upscaling scheme to combine multi-scale downsampled images based on image completion, and demonstrate it performs better than the commonly used bicubic method. We feed our upscaled images into a two-stage residual network pipeline. In the first stage, we infer original high-resolution images from the upscaled input. In the second stage, we further fine-tune the prediction by appending color guided images and input it into a smaller residual network. Both networks are economical in time and memory consumption while achieving competitive results.

In conclusion, we demonstrated different schemes combining multi-modal inputs for spectral super-resolution. While this work limited itself to the data provided by the challenge, it can be expanded into other modalities, namely different scales, near-infrared, or even depth inputs.

## References

1. Achanta, R., Arvanitopoulos, N., Susstrunk, S.: Extreme image completion. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2017). <https://doi.org/10.1109/icassp.2017.7952373>
2. Chang, C.I.: An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on information theory* **46**(5), 1927–1932 (2000)
3. Damodaran, B.B., Kellenberger, B., Flamary, R., Tuia, D., Courty, N.: Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. *arXiv preprint arXiv:1803.10081* (2018)
4. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: *European conference on computer vision*. pp. 184–199. Springer (2014)
5. Dosovitskiy, A., Fischer, P., Ilg, E., Hausser, P., Hazirbas, C., Golkov, V., Van Der Smagt, P., Cremers, D., Brox, T.: Flownet: Learning optical flow with convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2758–2766 (2015)
6. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495* (2014)
7. Guo, T., Mousavi, H.S., Vu, T.H., Monga, V.: Deep wavelet prediction for image super-resolution. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2017)
8. Gupta, S., Hoffman, J., Malik, J.: Cross modal distillation for supervision transfer. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2827–2836 (2016)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
10. Hu, Y., Zhang, D., Ye, J., Li, X., He, X.: Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE transactions on pattern analysis and machine intelligence* p. 1 (2012)
11. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1646–1654 (2016)
12. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 1637–1645 (2016)
13. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). <https://doi.org/10.1109/cvpr.2016.182>
14. Levin, A., Zomet, A., Weiss, Y.: Learning how to inpaint from global image statistics. In: *null*. p. 305. IEEE (2003)
15. Li, W., Zhao, L., Lin, Z., Xu, D., Lu, D.: Non-local image inpainting using low-rank matrix completion. In: *Computer Graphics Forum*. vol. 34, pp. 111–122. Wiley Online Library (2015)
16. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2017). <https://doi.org/10.1109/cvprw.2017.151>

17. Liu, Q., Lai, Z., Zhou, Z., Kuang, F., Jin, Z.: A truncated nuclear norm regularization method based on weighted residual error for matrix completion. *IEEE Transactions on Image Processing* **25**(1), 316–330 (2016)
18. Shi, W., Caballero, J., Huszar, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). <https://doi.org/10.1109/cvpr.2016.207>
19. Shoeiby, M., Robles-Kelly, A., Timofte, R., Zhou, R., Lahoud, F., Süsstrunk, S., Xiong, Z., Shi, Z., Chen, C., Liu, D., Zha, Z.J., Wu, F., Wei, K., Zhang, T., Wang, L., Fu, Y., Zhong, Z., Nagasubramanian, K., Singh, A.K., Singh, A., Sarkar, S., Baskar, G.: PIRM2018 challenge on spectral image super-resolution: Methods and results
20. Shoeiby, M., Robles-Kelly, A., Wei, R., Timofte, R.: PIRM2018 challenge on spectral image super-resolution: Dataset and study
21. Sun, J., Yuan, L., Jia, J., Shum, H.Y.: Image completion with structure propagation. In: *ACM Transactions on Graphics (ToG)*. vol. 24, pp. 861–868. ACM (2005)
22. Wei, Q., Dobigeon, N., Tournieret, J.Y.: Fast fusion of multi-band images based on solving a sylvester equation. *IEEE Transactions on Image Processing* **24**(11), 4109–4121 (2015)
23. Wycoff, E., Chan, T.H., Jia, K., Ma, W.K., Ma, Y.: A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In: *Acoustics, Speech and Signal Processing (ICASSP)*, 2013 IEEE International Conference on. pp. 1409–1413. IEEE (2013)
24. Yokoya, N., Yairi, T., Iwasaki, A.: Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing* **50**(2), 528–537 (2012)
25. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing* **26**(7), 31423155 (2017). <https://doi.org/10.1109/tip.2017.2662206>
26. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging* **3**(1), 47–57 (2017)