PIRM2018 Challenge on Spectral Image Super-Resolution: Methods and Results

Mehrdad Shoeiby[1], Antonio Robles-Kelly[2], Radu Timofte[3], Ruofan Zhou[4], Fayez Lahoud[4], Sabine Süsstrunk[4], Zhiwei Xiong[5], Zhan Shi[5], Chang Chen[5], Dong Liu[5], Zheng-Jun Zha[5], Feng Wu[5], Kaixuan Wei[6], Tao Zhang[6], Lizhi Wang[6], Ying Fu[6], Koushik Nagasubramanian[7], Asheesh K. Singh[7], Arti Singh[7], Soumik Sarkar[7], and Baskar Ganapathysubramanian[7]

[1]DATA61 - CSIRO, Black Mountain Laboratories, Acton ACT 2601, Australia
[2]Faculty of Sci., Eng. and Built Env., Deakin University, VIC 3216, Australia
[3]Computer Vision Laboratory, D-ITET, ETH Zurich, Switzerland
[4]Image and Visual Representation Laboratory, EPFL, Switzerland
[5]University of Science and Technology of China
[6]Beijing Institute of Technology, China
[7]Iowa State University, Lab of Mechanics, IA 50011

Abstract. In this paper, we describe the Perceptual Image Restoration and Manipulation (PIRM) workshop challenge on spectral image super-resolution, motivate its structure and conclude on results obtained by the participants. The challenge is one of the first of its kind, aiming at leveraging modern machine learning techniques to achieve spectral image super-resolution. It comprises of two tracks. The first of these (Track 1) is about example-based single spectral image super-resolution. The second one (Track 2) is on colour-guided spectral image super-resolution. In this manner, Track 1 focuses on the problem of super-resolving the spatial resolution of spectral images given training pairs of low and high spatial resolution spectral images. Track 2, on the other hand, aims to leverage the inherently higher spatial resolution of colour (RGB) cameras and the link between spectral and trichromatic images of the scene. The challenge in both tracks is then to recover a super-resolved image making use of low-resolution imagery at the input. We also elaborate upon the methods used by the participants, summarise the results and discuss their rankings.

Keywords: Super-resolution, Multispectral, Hyperspectral, RGB, Stereo

1 Introduction

Image super-resolution (SR) aims at reconstructing details, that is high frequency information that was lost in an image due to various reasons such as camera sensor limitations, blurring, subsampling, and image manipulations. Hence,

Mehrdad Sheoiby, Antonio Robles-Kelly, and Radu Timofte are the PIRM2018 organizers, while the other authors participated in the challenge.

image SR is an important problem which has found application in areas such as video processing [1], light field imaging [2] and image reconstruction [3], and has attracted ample attention in the image processing and computer vision community [4]. Early approaches to SR were often based upon the rationale that higher-resolution images have a frequency domain representation whose higher-order components are greater than their lower-resolution analogues. Thus, such methods [5] exploited the shift and aliasing properties of the Fourier transform to recover a super-resolved image. Kim *et al.* [6] extended the method in [5] to settings where noise and spatial blurring are present in the input image. In a related development, in [7], SR in the frequency domain is effected using Tikhonov regularization.

Note that the methods above are not based upon learning, but rather they aim at improving an image metric which is often related to signal-to-noise ratio (SNR). Regarding learning-based approaches, Dong *et al.* [8] present a deep convolutional network for single-image SR which surpasses the state-of-the-art performance at that time represented by patch-based methods using sparse coding [9] or anchored neighborhood regression [10]. Kim *et al.* [11] go deeper with a network based on VGG-net [12]. The network in [11] is comprised of 20 layers so as to exploit the image context across image regions. In [13], a multi-frame deep network for video SR is presented. The network employs motion compensated frames as input and single-image pre-training. In addition, some of the recent challenges on example-based single image SR [14,15,16], through benchmarking and introduction of SR specific datasets promoted several methods for super-resolving images [17,18,19,20,21].

However, the focus of the above methods/challenges is SR for trichromatic (colour)images, despite the fact that spectral cameras with modern complementary metal-oxide-semiconductor (CMOS) or charge-coupled device (CCD) detectors have more resolution constraints. This is mainly due to the larger number of wavelength channels that the spectral image sensors need to cover compared to RGB image sensors that only need to cover three channels. Note that these cameras (CCD/CMOS) are attractive imaging devices because they offer major advantages such as full integrability of the imaging sensors, high speed, and mobility. While the spectral SR has been subject to study for decades [22,23,24], example-based learning methods are limited [25] mainly due to the lack of spectral SR benchmarking platforms and difficulty accessing suitable SR spectral datasets. For example, among one of the few example-based spectral SR methods, [25] was developed by putting together three different hyperspectral datasets [26,27,28] with total combined 110 hyperspectral images.

Considering the above existing limitations for example-based spectral image SR, this challenge is motivated by three notions. (I) Inherent lower resolution of spectral imaging systems compared to their RGB counterpart renders spectral SR substantially crucial in improving the spatial resolution of imaging spectroscopy data. (II) The lack of a suitable spectral dataset is constraining the development of example based SR methods. (III) Lack of a benchmarking platform is making it difficult to assess and compare various spectral SR methodologies. Thereby, this challenge, while benchmarking example-based spectral SR, utilizes a novel dataset named StereoMSI to develop deep learning based SR methods[1]. The dataset consists of 350 multispectral images and their stereo RGB pairs. Since the dataset offers registered RGB images, the challenge also aims at leveraging the inherent higher resolution of RGB images to further improve the resolution of the spectral images.

For the rest of the paper, we first briefly introduce the StereoMSI dataset [29], before reviewing the challenge structure and the evaluation metrics. We then go through methods of the teams with performance above bicubic interpolation. Finally, we discuss the results of the winners of the challenge and conclude the paper.

2 Tracks and Dataset

The challenge consists of two tracks. Track 1 aims at using machine learning techniques to train single spectral image SR systems to obtain reliable multispectral super-resolved images at testing. The objective of Track 2 is to exploit the higher resolution of the RGB images as registered onto their corresponding spectral images to further boost the performance of the algorithms at testing.

2.1 Track 1: Spectral Image Super-Resolution

As mentioned above, Track 1 focuses on to the problem of super-resolving the spatial resolution of spectral images given training pairs of low spatial resolution (LR) and high spatial resolution (HR) or ground truth images for training. The main idea is to apply modern machine learning techniques to the problem of spectral SR and train a system that, at testing, can obtain a super-resolved version of a single LR image at input.

Thus, the computational objective of the track is to obtain ×3 spatially super-resolved spectral images making use of training imagery which has been down-sampled with the factors of ×2, and ×3 using nearest neighbour interpolation. For Track 1, 240 spectral images have been split into 200 for training, 20 for validation and 20 for testing.

2.2 Track 2: Colour-guided Spectral Image Super-Resolution

Track 2 of the challenge aims at leveraging the link between spectral and trichromatic images of the scene to facilitate the use of on-sensor filter arrays. The motivation here is that, by using machine learning techniques and the increased spatial resolution of colour cameras, a system can be trained to obtain reliable spectral super-resolved images at testing.

Thus, the computational objective of the track is to obtain ×3 spatially super-resolved spectral images making use of spectral-colour stereo pairs. Since

[1] Refer to `https://pirm2018.org/` for the spectral SR challenge and the dataset download links.

the RGB images are pixel-wise aligned to spectral images, the inherent higher resolution of ×4 in colour imagery can be exploited to better train the system so as to improve the SR results. In this case, 120 stereo image pairs are used, with 100 of these employed for training, 10 for validation and 10 for testing.

2.3 Dataset

Both tracks of the challenge are based on a novel dataset, which we have named StereoMSI (Stereo Multispectral Image) dataset. The dataset consists of 350 stereo RGB-spectral image pairs which were collected in-house. The images in the dataset depict a wide variety of scenes, under natural and artificial illuminants in the city of Canberra, the capital of Australia. The nature of the images ranges from natural settings to industrial and office environments. At acquisition time, special attention was paid to the exposure settings as related to the image quality. This is important since the stereo pairs were captured using two different cameras, one with a colour sensor (RGB XiQ camera model MQ022CG-CM) and the other one based upon the IMEC snapshot sensor (a XiQ multispectral camera model MQ022HG-IM-SM4x4) covering the visible range between 470nm and $620nm$. For more information on the StereoMSI dataset, we refer the reader to [29].

3 Challenge Structure and Evaluation Metrics

3.1 Challenge Phases

The challenge was structured similarly for both tracks. It comprised of three phases. These were the development, the validation and the testing phase. During the development phase, the participants gained access to all the training and the LR validation images so as to be able to develop their solutions offline. In the validation phase, the participants had the opportunity to test their solutions on the Codalab[2] server that would, later on, be used for the testing and benchmarking. This gave the participants the opportunity to fine-tune their methods and evaluate them using the same quality metrics that would be used for the final testing phase. During the testing phase, the participants were given access to the LR testing images and were required to submit super-resolved HR image results, code, and a fact sheet describing their solutions for the final evaluation of their methods.

3.2 Evaluation Protocol

During the testing phase, the submitted super-resolved images were evaluated with respect to the fidelity of the reconstruction of the spectral images at testing.

[2] Refer to `https://competitions.codalab.org/competitions/19226` for Track 1, and `https://competitions.codalab.org/competitions/19227` for Track 2

Regarding the quantitative assessment of the fidelity of the spectral images, this was effected by comparing the super-resolved hyperspectral images with their corresponding ground truth. For purposes of ranking the participants in the challenge, we used the mean of relative absolute error (MRAE) and the spectral information divergence (SID). Besides, the per-band mean squared error (MSE), the average per-pixel spectral angle (APPSA), the average per-image structural similarity index (SSIM) and the mean per-image peak signal-to-noise ratio (PSNR) were also computed. Nonetheless, these were not used for purposes of ranking but were obtained since they provide a full set of metrics and scores to evaluate the quality of the submitted results.

3.3 Assessment Measures

As mentioned above, we have used a wide variety of quality measures to assess the results submitted by the participants during the challenge. The first of these is the MRAE [30], which is given by

$$MRAE = \frac{1}{M \times W \times H} \sum_{i=1}^{M} \mathbf{1}^T \frac{|\mathbf{I}_i^* - \mathbf{I}_i|}{\mathbf{I}_i} \mathbf{1} \tag{1}$$

where $\mathbf{I}_i^*$ is the matrix corresponding to the i^{th} wavelength-indexed channel in the super-resolved image, $\mathbf{I}_i$ is the array for the channel indexed i in the reference image *i.e.* the ground truth, $\mathbf{1}$ is an all-ones column vector whose length depends on the context, W, H and N are the width, height and the number of wavelengths channels in the image, respectively.

For ranking the participants, we also used SID, which is an information theoretic measure for spectral similarity and discriminability [31]. We have computed the mean SID (MSID) as follows

$$MSID = \frac{1}{M} \sum_{i=1}^{M} SID_i \tag{2}$$

where the spectral information divergence for the i^{th} wavelength-indexed band is given by

$$SID_i = D(\boldsymbol{x}||\boldsymbol{y}) + D(\boldsymbol{y}||\boldsymbol{x}) \tag{3}$$

and

$$\begin{aligned} D(\boldsymbol{x}||\boldsymbol{y}) &= \sum_{n=1}^{W \times H} p_n log(p_n/q_n) \\ D(\boldsymbol{y}||\boldsymbol{x}) &= \sum_{n=1}^{W \times H} q_n log(q_n/p_n). \end{aligned} \tag{4}$$

Here, p_n and q_n are the normalized values at the wavelength-indexed band under consideration for the n^{th} pixel in the reference (ground truth), and super-resolved images, respectively.

As mentioned above, we have also assessed the results using other measures that, despite not being used for ranking, are widely employed elsewhere in the literature for purposes of evaluating the performance of image enhancement methods in general. One measure is the APPSA given by

$$APPSA = \frac{1}{W \times H}\mathbf{1}^T \left[\arccos \left(\frac{\sum_{i=1}^{M} (\mathbf{I}_i^* \odot \mathbf{I}_i)}{\sqrt{\sum_{i=1}^{M} (\mathbf{I}_i^* \odot \mathbf{I}_i^*)}\sqrt{\sum_{i=1}^{M} (\mathbf{I}_i \odot \mathbf{I}_i)}} \right) \right] \mathbf{1} \tag{5}$$

whereby, in the equation above, we have use the same notation as earlier in the section.

The equations for MSE and PSNR are expressed as

$$MSE = \frac{1}{M \times\ W \times\ H} \sum_{i=1}^{M} ||\mathbf{I}_i^* - \mathbf{I}_i||_2^2 \tag{6}$$

, and

$$PSNR = 20 \times \log_{10} \left(\frac{p\_max}{MSE} \right) \tag{7}$$

where $p\_max = 2^{16} - 1$, *i.e.* 65535, corresponds to the maximum possible value of each pixel.

Note that, when comparing images, MRAE or MSE measures have the advantage of ease of implementation, however, they are not aimed at measuring perceptual similarity. Thus, we have used the structural similarity index (SSIM)[32]. The SSIM is a perceptual metric that quantifies the image quality by taking texture into account [33]. Following the authors, we have calculated the SSIM across several windows over the image and averaged them to compute the final results. The per-window SSIM is given by

$$SSIM_{i,n} = \frac{(2\mu\mu^* + C_1)(2\hat{\sigma} + C_2)}{(\mu^{*2} + \mu^2 + C_1)(\sigma^{*2} + \sigma^2 + C_2)} \tag{8}$$

where μ^* and ${\sigma^*}^2$ are the mean and variance for the n^{th} $N \times N$ window in the i^{th} wavelength-indexed band on the super-resolved image. Similarly, μ and σ^2 account for the mean and variance of the window in the reference image. Also, $C_1 = k_1 L$, and $C_2 = k_2 L$ are introduced to avoid division by zero when the mean or covariance values are close to zero; L is the dynamic range of the pixel values (65535 for 16-bit images) with $k_1 \ll 1$, and $k_2 \ll 1$ being a small constant [33].

With the $SSIM_{i,n}$ in hand, the mean SSIM per image can be computed in a straightforward manner as follows

$$MSSIM = \frac{1}{M \times W \times H} \sum_{i=1}^{M} \sum_{n=1}^{W \times H} SSIM_{i,n} \tag{9}$$

For the computation of the $SSIM$ results presented here we used the command *compare_ssim* from the python package *scikit-image*[3] with default parameter settings and a window size of 7×7, *i.e.* $N = 7$.

[3] For more information on scikit-image toolkit, go to `http://scikit-image.org`

4 Challenge Methods and Teams

In total, the challenge had four participating teams that, having subscribed on the Codalab website, completed the three phases and submitted testing results that improved upon the bicubic upsampling baseline as measured using the MRAE and SID. Of these four teams, all participated in Track 1 while three of the teams competed in Track 2. In this section, we elaborate further on the approach taken by each of these teams. The name of each team appears in parentheses.

4.1 Residual Learning *(IVRL_Prime)*

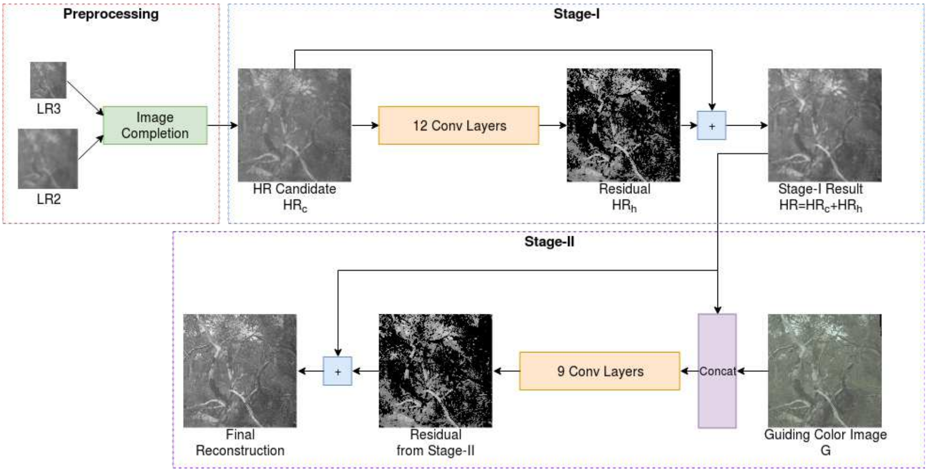

Fig. 1. Illustration of the proposed stacked residual learning framework for spectral image super-resolution, it contains three steps: preprocessing, Stage-I and Stage-II. Image completion is done in preprocessing to generate a HR candidate. Then Stage-I reconstruct the high-resolution spectral image using a 12-layer residual learning network. Finally Stage-II refines Stage-I results using guiding color image G through a 9-layer residual learning network.

This framework [34] contains preprocessing and two residual learning networks[4]. The image compression algorithm [35] was first used on the given LR×2 and LR×3 inputs to generate an HR candidate with the desired size. Then two residual learning networks [36] follows. As depicted in Fig. 1, Stage-I uses one 12-layer residual learning network to reconstruct a primary results from the HR candidate. Stage-II is built upon Stage-I results and it also takes the HR color image as inputs. One 7-layer residual learning network refines the outputs from Stage-I and produces the final results.

[4] Code at `https://github.com/IVRL/Multi-Modal-Spectral-Image-Super-Resolution`.

As the color images are not provided in Track 1, Stage-II is ignored in the solution for Track 1. Since both tracks contains spectral images, both Track 1 and Track 2 datasets were used for the training of Stage-I. Overlapping patches of size 96 × 96 with a stride of 24 were cropped from the dataset. Adam [37] is used for optimizing the network with weight decay of $1e-5$ and a learning rate of 0.001. The learning rate is decayed by 10 every 30 epochs. During the training of the second track, Track 2 dataset was utilized. Overlapping patches of 48×48 were cropped with a stride of 16 from the dataset. The training strategy for Stage-II was similar to Stage-I. Sum of SID and MRAE was used for the loss function for the training of both stages.

According to the authors, this method is the first stacked residual learning framework for spectral image SR. It is also a novel solution to transfer knowledge from a large dataset into a smaller one with different modalities where training data is limited.

4.2 Deep Residual Attention Network *(VIDAR)*

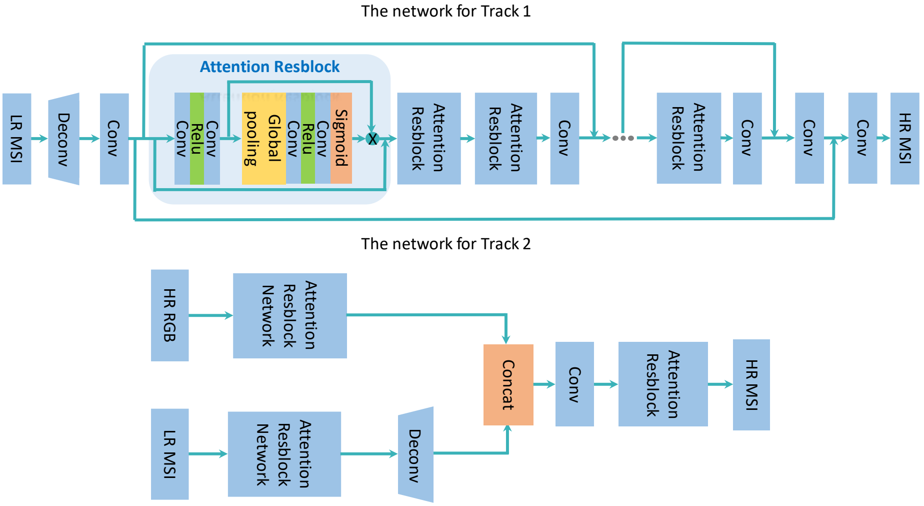

Fig. 2. Illustration of deep residual attention network

This method proposes a novel deep residual attention network [38] for the spatial SR of MSIs. The proposed method extends the classic residual network by I) directly using the 3D LR MSI as input instead of upsampling the 2D bandwise images separately, and II) integrating the channel attention mechanism into the residual network. These two operations fully exploit the correlations across both the spectral and spatial dimensions of MSIs and significantly promote the performance of MSI SR. Furthermore, In Track 2, a fusion framework was designed based on the proposed network. The spatial resolution of the MSI input is enhanced in one branch, while the spectral resolution of the RGB input is enhanced in the other. These two branches are then fused together by concatenating the features of high spatial resolution RGB images and low spatial resolution spectral images using a channel attention mechanism. This is to achieve further improvements in the results compared to using the single MSI input. Note that to avoid zero points in the registered RGB images, the images were cropped with 12 pixels on each border in the network. For the super-resolved images, reconstructed images of Track 1 were used to make up the cropped borders.

For Track 1, the network was trained using spectral input patches of size 20×20 and corresponding HR 60×60 spectral patches. Batch size is set to 64 and the optimizer is ADAM by setting $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10e - 8$. The initial learning rate is set to $1 \times 10 - 4$ and then decreases to half every $1 \times 10e5$ iterations of back-propagation. The training is stopped when no notable decay of training loss is observed. For the loss function, modified MARE was utilised which converts numbers below 1/65536 to zeros. Compared to the training process of Track 1, in Track 2, HR registered RGB patches were added as inputs. The other training settings are consistent with Track 1.

4.3 Enhanced Memory-Persistent Network *(Grit)*

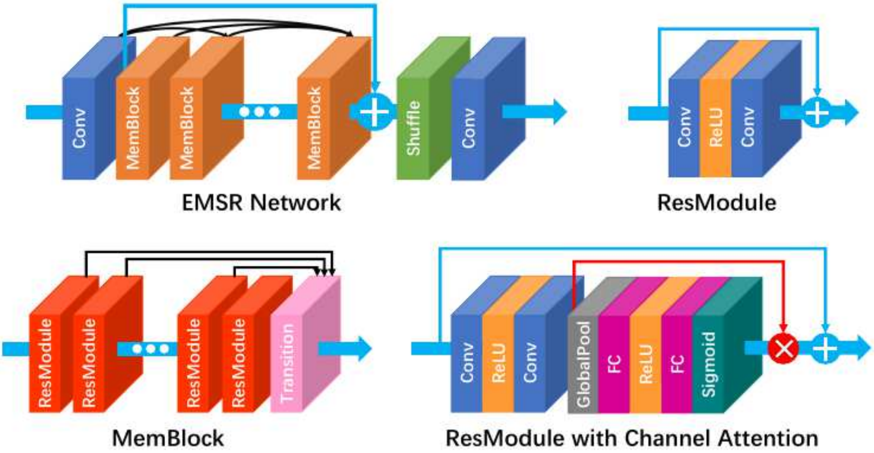

Fig. 3. The architecture of the proposed Enhanced Memory-Persistent Network (EMSR) and its variant (EMSR-CA)

The authors inspect the potential bottlenecks still existing in the strong baseline EDSR [39] and propose two solutions to boost the performance further.

The proposed solutions are derived from two levels of hierarchy, *i.e.* model level and building block level.

In the model level, the authors build a new level of hierarchical structure over EDSR by stacking multiple residual blocks into a new block named memory block [40]. The proposed network still follows the topology like the EDSR model but contains multiple memory blocks instead of residual block. This hierarchical structure enables equipment of the dense connection [41] to encourage feature reuse (or persist memory), which may be a potential bottleneck neglected by EDSR model. It's worth noting that the "dense connection" inner the memory block is a simplified version of the one introduced in [41], since only the final transition stage (see Fig. 3) receives the outputs of all preceding layers, while the intermediate layers not. This simplification, allows the authors to adopt the residual module inner the memory block without worrying about the drastic increment of feature maps. Through such modification, a new architecture is proposed named Enhanced Memory-Persistent network (EMSR) to super-resolve the spectral image.

In the building block level, the residual scaling [42] used in EDSR is replaced by Squeeze and Excitation module [43] (*a.k.a* channel-wise attention). The residual scaling was originally used in EDSR to tackle with unstable training phenomenon. The authors find that instead of explicit scaling each channel of output with a fixed constant (0.1 in EDSR), utilizing channel-wise attention (CA) would gain great benefit, especially in spectral SR case. The CA module models the channel correlation in an adaptive way. It uses a set of learnable parameters to calculate the scaling factors *w.r.t.* each channel/feature maps, then multiply its original input by such scaling factors. This module can not only stabilize the training procedure but can also explicitly extract the spectral correlation from the data, such that significantly enhancing the performance. By integrating the two solutions together, a novel architecture named Enhanced Memory-Persistent network with Channel-wise Attention (EMSR-CA) has been proposed for SR.

The EMSR trained on the training set of track 2 was adopted without utilizing the RGB images.

During testing, a geometric self-ensemble strategy is adopted to maximize the potential performance of the models. During the test time, the input is flipped and rotated to generate 7 geometric transformed inputs (8 in total) for each sample. With those augmented images, and using the network, the corresponding super-resolved images were produced. The inverse transform is applied to those output images to get the original geometry. Finally, all obtained outputs are averaged to reach the final self-ensemble result. The methods surpass the baseline EDSR in all the metrics, demonstrating the effectiveness of the proposed solutions.

4.4 HYPER_CNN *(Spectral_SR)*

The architectures consist of 6 2D convolutional layers for the spatial SR of the spectral images. A residual block [44] comprised of two convolutional layers were

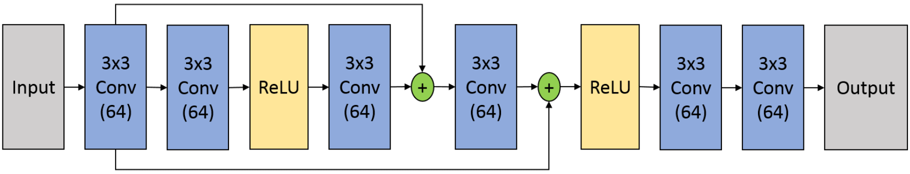

Fig. 4. Illustration of convolution neural network architecture for spectral super-resolution. The number of convolutional filters used in each layer is shown within brackets

stacked between convolutional layers. Each convolutional layer consisted of 64 filters with a kernel size of 3×3. A skip connection was used between the output of the first convolution layer and the residual block. 6400 patches of $20 \times 20 \times 14$ size were extracted from the LR spectral images and were used as input to the deep learning model. The architecture learns to spatially upsample the input data by a factor of three ($60 \times 60 \times 14$). Mean absolute error was used as loss function. The model was initialized with HeUniform [45] and trained with Adam [37] optimizer for 300 epochs with a learning rate of 0.001 and mini batches of size 32.

During testing 640 patches of $20 \times 20 \times 14$ were extracted from 20 test images for model prediction. The HR images were generated by aligning the model predictions.

5 Challenge Results and Discussion

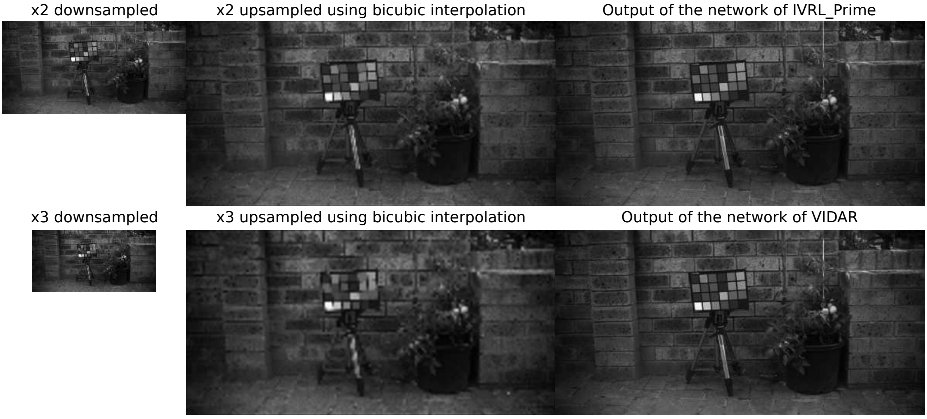
x2 downsampled
x2 upsampled using bicubic interpolation
Output of the network of IVRL_Prime
x3 downsampled
x3 upsampled using bicubic interpolation
Output of the network of VIDAR

Fig. 5. Performance of IVRL_Prime and VIDAR teams on testing image 118, compared to bicubic upsampled LR×2 and LR×3 images. Note that, for IVRL_Prime, inputs are LR×2 and LR×3 images, and for VIDAR the input is only the LR×3 image.

Table 1. Track 1 results. In the table, we show the mean and standard deviation (in parenthesis) for all the quality metrics used in our spectral image SR benchmarking.

Team	MRAE	SID	APPSA	MSE	PSNR	SSIM
IVRL_Prime	0.07	0.00006	0.06	1246673	36.7	0.82
VIDAR	0.11	0.00018	0.08	3414849	32.2	0.62
Grit	0.12	0.00016	0.08	3061024	32.8	0.63
Spectral_SR	0.17	0.00020	0.09	3712957	31.7	0.58
Bicubic upsample (×3)	0.21	0.00035	0.11	6353052	29.3	0.47

Table 2. Track 2 results. In the table, we show the mean and standard deviation (in parenthesis) for all the quality metrics used in our colour-guided SR benchmarking.

Team	MRAE	SID	APPSA	MSE	PSNR	SSIM
IVRL_Prime	0.07	0.00005	0.05	852268	38.2	0.86
VIDAR	0.09	0.00011	0.08	1940939	34.5	0.75
Grit	0.12	0.00017	0.08	3131840	32.7	0.63
Bicubic upsample (×3)	0.21	0.00035	0.11	6353052	29.3	0.47

We now turn our attention to the results of the challenge and the wining architectures. Note that, for both challenges, IVRL_Prime and VIDAR were the best performers, *i.e.* the winners of the challenge. The top 2 participants of the challenge and Grit (the third on both tracks) were trained using Adam [37]. In contrast to the two top performers, each of the networks used in Grit for the two tracks are trained from scratch using the ×2 downsampled imagery provided in the dataset. After the model converges, it is used as a pre-trained network for the ×3 scale.

It is also worth noting that IVRL_Prime team has employed both, the ×2 and ×3 downsampled imagery for training. This contrasts with the other teams, which employed only the ×3 downsampled imagery. Figure 5 depicts the performance of IVRL_Prime and VIDAR teams on image 118 from the testing split for Track 2. For the results presented here, we have used imresize command from python scikit-image[5] toolbox with bicubic upsampling as a baseline. This is important, since, for purposes of benchmarking, only those teams that improved upon the baseline, *i.e.* upsampling using a bicubic kernel, have been included herein.

In Tables 1 and 2, we show the performance, per-team for each of the two tracks. In the tables, we have written in bold font the two winning teams and, in parenthesis, we show the standard deviation for each of the mean metrics

[5] `https://scikit-image.org/`

used in the experiments. Note that the IVL_Prime team is consistently the best performer followed by VIDAR. This is somewhat expected since the IVL_Prime team used both sets of down-sampled imagery, which gives a better training and provides more information at testing. In Table 3, we summarise the details with respect to the platform used and runtime in seconds reported by each of the participating teams. Note that IVRL_Prime is also the fastest of the methods under consideration, followed by VIDAR for Track 1 and Grit for Track 2.

Table 3. Reported runtimes and information provided in submitted factsheets by the teams in the challenge.

Factsheet details	**IVRL_Prime**	**VIDAR**	**Grit**	**Spectral_SR**
Runtime $[s]$	Track 1: 0.3 Track 1: 0.5	Track 1: 1 Track 2: 2.5	Track 1: 1.2 Track 2: 1.2	Track 1: 0.23
Training time $[hours]$	0.8/epoch	12	40	0.62
Language	Python	Python	Python	Python
Platform	PyTorch	PyTorch	PyTorch	Keras with Tensorflow back end
CPU/GPU (at runtime)	GTX Titan X	NVIDIA 1080Ti	NVIDIA Titan X	NVIDIA Titan X
Ensemble	-	flip/rotation	flip/rotation	-
Number of Parameters	Stage-I: 385K Stage-II: 275K	Track 1: 0.6M Track 2: 1.3M	52M	128K
Memory requirements	3700 MB GPU for training, and 800 MB GPU for inference on input	16 GB with no parallelization	12 GB GPU	647 MB GPU

6 Conclusions

In this paper, we have presented the PIRM 2018 spectral image super-resolution challenge, elaborating upon its structure, the participating teams and the results of the testing imagery submitted for benchmarking. The challenge is one of the first of its kind to date, focusing on both, single-image and colour-guided spectral SR. The challenge provides a means to a spectral image super-resolution benchmark that is directly applicable to on-sensor spectral filter arrays that have recently come to market and that are akin to Bayer arrays widely used in trichromatic cameras. Moreover, it leverages modern machine learning techniques making use of a novel dataset.

Acknowledgments

The PIRM2018 challenge was sponsored by CSIRO's DATA61, Deakin University, ETH Zurich, HUAWEI, and MediaTek.

Teams and members

PIRM2018 spectral SR team

Members: Mehrdad Shoeiby, Antonio Robles-Kelly, and Radu Timofte
Contact email: *mehrdad.shoeiby@data61.csiro.au*
antonio.robles-kelly@deakin.edu.au
radu.timofte@vision.ee.ethz.ch

IVRL_Prime

Members: Ruofan Zhou, Fayez Lahoud, and Sabine Süsstrunk
Contact email: *ruofan.zhou@epfl.ch*

VIDAR

Members: Zhiwei Xiong, Zhan Shi, Chang Chen, Dong Liu, Zheng-Jun Zha, and Feng Wu
Contact email: *zwxiong@ustc.edu.cn*

Grit

Members: Kaixuan Wei, Tao Zhang, Lizhi Wang, and Ying Fu
Contact email: *kaixuan_wei@outlook.com*

Spectral_SR

Members: Koushik Nagasubramanian, Asheesh K. Singh, Arti Singh, Soumik Sarkar, and Baskar Ganapathysubramanian
Contact email: *koushikn@iastate.edu*

References

1. Eren, P.E., Sezan, M.I., Tekalp, A.M.: Robust, object-based high-resolution image reconstruction from low-resolution video. IEEE Transactions on Image Processing **6**(10) (1997) 1446–1451
2. Bishop, T.E., Zanetti, S., Favaro, P.: Light field superresolution. In: Computational Photography (ICCP), 2009 IEEE International Conference on, IEEE (2009) 1–9
3. Farsiu, S., Robinson, D., Elad, M., Milanfar, P.: Robust shift and add approach to superresolution. In: Applications of Digital Image Processing XXVI. Volume 5203., International Society for Optics and Photonics (2003) 121–131
4. Li, T.: Single image super-resolution: A historical review. In: ObEN Research Seminar. (2018)
5. Tsai, R.: Multiframe image restoration and registration. Advance Computer Visual and Image Processing **1** (1984) 317–339
6. Kim, S., Bose, N.K., Valenzuela, H.: Recursive reconstruction of high resolution image from noisy undersampled multiframes. IEEE Transactions on Acoustics, Speech, and Signal Processing **38**(6) (1990) 1013–1027
7. Bose, N., Kim, H., Valenzuela, H.: Recursive total least squares algorithm for image reconstruction from noisy, undersampled frames. Multidimensional Systems and Signal Processing **4**(3) (1993) 253–268

8. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38**(2) (2016) 295–307
9. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE transactions on image processing **19**(11) (2010) 2861–2873
10. Timofte, R., De Smet, V., Van Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Asian Conference on Computer Vision, Springer (2014) 111–126
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. (2016) 1646–1654
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
13. Kappeler, A., Yoo, S., Dai, Q., Katsaggelos, A.K.: Video super-resolution with convolutional neural networks. IEEE Transactions on Computational Imaging **2**(2) (2016) 109–122
14. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L., Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., et al.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on, IEEE (2017) 1110–1121
15. Timofte, R., Gu, S., Wu, J., Van Gool, L., Zhang, L., Yang, M.H., Haris, M., Shakhnarovis, G., Ukita, N., Shijia, H., et al.: Ntire 2018 challenge on single image super-resolution: Methods and results. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on, IEEE (2018)
16. Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: 2018 PIRM challenge on perceptual image super-resolution. In: European Conference on Computer Vision Workshops (ECCVW). (2018)
17. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: The IEEE conference on computer vision and pattern recognition (CVPR) workshops. Volume 1. (2017) 4
18. Fan, Y., Shi, H., Yu, J., Liu, D., Han, W., Yu, H., Wang, Z., Wang, X., Huang, T.S.: Balanced two-stage residual networks for image super-resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE (2017) 1157–1164
19. Bei, Y., Damian, A., Hu, S., Menon, S., Ravi, N., Rudin, C.: New techniques for preserving global structure and denoising with low information loss in single-image super-resolution. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. Volume 4. (2018)
20. Ahn, N., Kang, B., Sohn, K.A.: Image super-resolution via progressive cascading residual network. progressive **24** (2018) 0–771
21. Haris, M., Shakhnarovich, G., Ukita, N.: Deep backprojection networks for super-resolution. In: Conference on Computer Vision and Pattern Recognition. (2018)
22. Loncan, L., Almeida, L.B., Bioucas-Dias, J.M., Briottet, X., Chanussot, J., Dobigeon, N., Fabre, S., Liao, W., Licciardi, G.A., Simoes, M., et al.: Hyperspectral pansharpening: A review. arXiv preprint arXiv:1504.04531 (2015)
23. Lanaras, C., Baltsavias, E., Schindler, K.: Hyperspectral super-resolution by coupled spectral unmixing. In: Proceedings of the IEEE International Conference on Computer Vision. (2015) 3586–3594
24. Kawakami, R., Matsushita, Y., Wright, J., Ben-Ezra, M., Tai, Y.W., Ikeuchi, K.: High-resolution hyperspectral imaging via matrix factorization. In: Computer Vi-

sion and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 2329–2336

25. Li, Y., Hu, J., Zhao, X., Xie, W., Li, J.: Hyperspectral image super-resolution using deep convolutional neural network. Neurocomputing **266** (2017) 29–41
26. Yasuma, F., Mitsunaga, T., Iso, D., Nayar, S.K.: Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. IEEE transactions on image processing **19**(9) (2010) 2241–2253
27. Chakrabarti, A., Zickler, T.: Statistics of real-world hyperspectral images. In: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011) 193–200
28. Foster, D.H., Nascimento, S.M., Amano, K.: Information limits on neural identification of colored surfaces in natural scenes. Visual neuroscience **21**(3) (2004) 331–336
29. Shoeiby, M., Robles-Kelly, A., Wei, R., Timofte, R.: PIRM2018 challenge on spectral image super-resolution: Dataset and study. In: European Conference on Computer Vision Workshops (ECCVW). (2018)
30. Arad, B., Ben-Shahar, O., Timofte, R.: Ntire 2018 challenge on spectral reconstruction from rgb images. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. (June 2018)
31. Chang, C.I.: Spectral information divergence for hyperspectral image analysis. In: Geoscience and Remote Sensing Symposium, 1999. IGARSS'99 Proceedings. IEEE 1999 International. Volume 1., IEEE (1999) 509–511
32. Wang, Z., Bovik, A.C.: Mean squared error: Love it or leave it? a new look at signal fidelity measures. IEEE signal processing magazine **26**(1) (2009) 98–117
33. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4) (2004) 600–612
34. Lahoud, F., Zhou, R., Süsstrunk, S.: Multi-modal spectral image super-resolution. In: Proceedings of the European Conference on Computer Vision (ECCV). (2018)
35. Achanta, R., Arvanitopoulos, N., Süsstrunk, S.: Extreme image completion. In: Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on, Ieee (2017) 1333–1337
36. 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. (2016) 1646–1654
37. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
38. Shi, Z., Chen, C., Xiong, Z., Liu, D., Zha, Z.J., Wu, F.: Deep residual attention network for spectral image super-resolution. In: European Conference on Computer Vision Workshops (ECCVW). (2018)
39. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: The IEEE conference on computer vision and pattern recognition (CVPR) workshops. Volume 1. (2017) 4
40. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2017) 4539–4547
41. Liu, Z., Li, J., Shen, Z., Huang, G., Yan, S., Zhang, C.: Learning efficient convolutional networks through network slimming. In: Computer Vision (ICCV), 2017 IEEE International Conference on, IEEE (2017) 2755–2763

42. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI. Volume 4. (2017) 12
43. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. arXiv preprint arXiv:1709.01507 **7** (2017)
44. 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. (2016) 770–778
45. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision. (2015) 1026–1034