# Super-resolved Chromatic Mapping of Snapshot Mosaic Image Sensors via a Texture Sensitive Residual Network

Mehrdad Shoeiby
DATA61-CSIRO
ACT, Australia
mehrdad.shoeiby@data61.csiro.au

Lars Petersson
Data61-CSIRO
ACT, Australia

Mohammad Ali Armin
Data61-CSIRO
ACT, Australia

Sadegh Aliakbarian
Data61-CSIRO
ACT, Australia

Antonio Robles-Kelly
Deakin University
VIC, Australia

## Abstract

*This paper introduces a novel method to simultaneously super-resolve and colour-predict images acquired by snapshot mosaic sensors. These sensors allow for spectral images to be acquired using low-power, small form factor, solid-state CMOS sensors that can operate at video frame rates without the need for complex optical setups. Despite their desirable traits, their main drawback stems from the fact that the spatial resolution of the imagery acquired by these sensors is low. Moreover, chromatic mapping in snapshot mosaic sensors is not straightforward since the bands delivered by the sensor tend to be narrow and unevenly distributed across the range in which they operate. We tackle this drawback as applied to chromatic mapping by using a residual channel attention network equipped with a texture sensitive block. Our method significantly outperforms the traditional approach of interpolating the image and, afterwards, applying a colour matching function. This work establishes state-of-the-art in this domain while also making available to the research community a dataset containing 296 registered stereo multi-spectral/RGB images pairs.*

## 1. Introduction

Imaging spectroscopy devices can capture an information-rich representation of a scene, often in terms of tens or hundreds of wavelength-indexed bands. Recent advances in imaging spectroscopy have seen the development of real-time snapshot mosaic image sensors, which are compact in size and exhibit comparable frame rates to current trichromatic cameras [44, 6]. Despite of the extensive interest in snapshot mosaic sensors and their potential for multi-spectral imaging, they suffer from
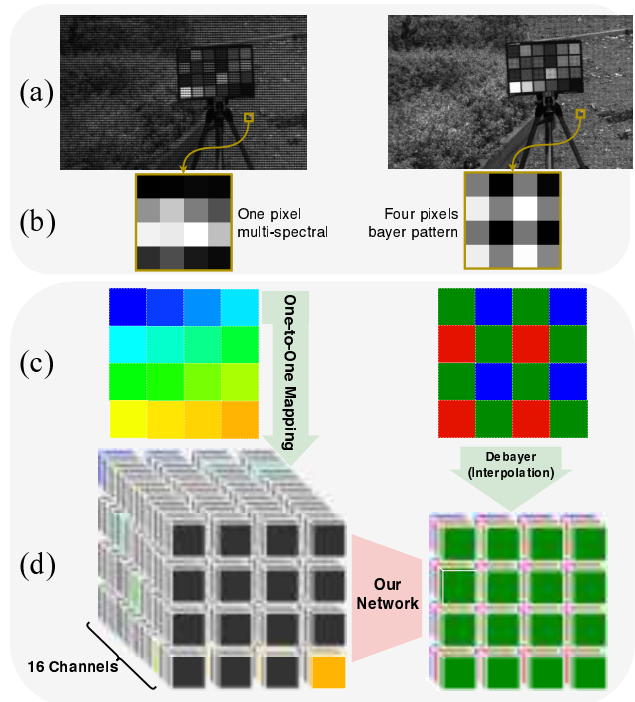


Figure 1. (a) Original mosaic images, with (b) the actual one pixel of the multi-spectral vs 4 pixels of the RGB bayer image encompassing 16 vs 3 channels (wavelengths) per pixel respectively. (c) and (d) demonstrate the data formation for multi-spectral vs RGB images. Each colour is indicative of respective wavelength. The multi-spectral cube is formed by a one-to-one mapping of each wavelength (sub-pixel) to respective channel (one of 16) and zero padding the other 15 sub-pixels. The RGB image is debayered (interpolated) to form the respective R-G-B channels. While a one-to-one mapping (with zero-padding) leads to a large number of redundant zero pixels, as opposed to debayering for RGB images, it results in better super-resolution by taking into account the spatial offset of each pixel.

an inherent trade-off between the spatial and spectral resolution. This is as a result of their architecture, where the raw resolution of the detector is distributed across the number of wavelength-indexed bands in the spectral image produced at output.

As a result, a higher spatial resolution (smaller pixel size) reduces the number of wavelength bins that can fit on that pixel on the image sensor. This creates a constraint for certain applications where smaller/lighter cameras are needed, for instance, on a UAV [10]. A smaller/lighter camera for portability reasons renders a device suffering from lower spatial and spectral resolution. Further, these sensors have promising applications ranging from remote sensing [15, 17] to food monitoring [13], and from astronomy [5] to object detection in autonomous vehicles [37, 29].

Furthermore, in many applications it is useful, or in fact crucial, to obtain an RGB image of the same scene. There is a large body of work in computer vision that can be directly leveraged if we can devise a method that delivers a high-quality, high-resolution RGB image from a multi-spectral sensor. For example, in case of object detection in autonomous vehicles [10, 29], RGB cameras as well as multi-spectral cameras were deployed. The acquired RGB images are usually registered against their multi-spectral counterpart to obtain 3D information of the scene, or to compensate for lower spectral information of the multi-spectral images. However, the low spatial/spectral resolution of multi-spectral cameras could render the registration challenging.

Traditionally, the RGB equivalent of a scene can be extracted from the spectral image using a Colour Matching Function [35], given that the wavelength range of the camera covers, relatively densely, the complete range of the visual spectrum. Given the relatively limited spectral resolution of snapshot mosaic sensors and their uneven spacing over the operating spectral range, the problem of obtaining high resolution RGB images from low spatial and spectral resolution multi-spectral images is an interesting one that needs to be addressed. Therefore, in this paper, we identify a gap in the scientific literature and propose a single unifying method that carries out 1) color-prediction, and 2) super-resolution (SR) from the multi-spectral space to the RGB space simultaneously.

The reason for this gap in the literature is the fact that these devices, with relatively recent technology, have just come in to the market. This leads to a limited exposure to researchers and a lack of rich, publicly available, datasets. Thus, we not only present a method that can simultaneously super-resolve and colour-predict spectral images acquired by snapshot mosaic sensors, but also introduce a novel stereo registered multi-spectral/RGB dataset. Further, our method is quite general in nature, being applicable not only to mosaic snapshot sensor imagery but also to spectral images delivered by other kinds of cameras.

## Contributions

- We propose a data representation method which exploits the mosaic structure of the images acquired by a snapshot sensor directly as opposed to demosaicing images to perform SR and colour prediction sequentially.

- We introduce a novel algorithm, to our knowledge the first in the literature, to carry out SR and colour-prediction simultaneously from mosaic images, establishing state-of-the-art in the field.

- We introduce a novel dataset containing 296 registered stereo snapshot mosaic-RGB image pairs.

## 2. Related work

Due to the lack of suitable datasets, color-prediction is not a problem that has been extensively studied for multi-spectral images in general. Most of the work is carried out with simulated multi-spectral images [27, 28, 26, 18]. The images were simulated exploiting high resolution hyperspectral images. In addition, the focus of these works, while producing RGB images from simulated multi-spectral images, is to mitigate the structural artifacts introduced by different demosaicing methods. They achieve this, with demonstrated good results, via using forms of interpolation (eg, linear, polynomial, low pass filtering in the frequency domain) to insert additional pixels between the observed spatial/spectral ones. None of the works above attempt to predict RGB from spectrally under-sampled data. The camera we are using is an off-the-shelf commercial camera with narrow FWHM $\approx 15nm$. In addition, it covers the blue and red spectra only partially (see Figure 3). Furthermore, many demosaicing methods such as [27, 28] are dependent on a given mosaic pattern as part of their approach. Also, [27, 28, 26, 18] and most of the demosaicing algorithms rely on a wavelength channel being more densely sampled than the others, using that as a guide image. A $4 \times 4$ pattern, with each of the 16 pixels/wavelengths appearing only once (similar to our camera), would be considered by [18] as severely under-sampled and as demonstrated experimentally, leads to poor results [18]. Note that a multi-spectral camera that covers a broader spectral range by using a larger number of narrow wavelength bins would be rendered very bulky and expensive, while the camera used here[1] has dimensions of $26mm \times 26mm \times 26mm$.

Image super-resolution (SR) is a problem that has been studied extensively for RGB images. While these algorithms are not perfect for multi-spectral images, they could be exploited to design efficient multi-spectral SR methods. Early approaches to SR were often based upon the rationale that images with higher spatial information have a fre-

---

[1]Ximea model MQ022HG-IM-SM4x4 470-620

quency domain response whose higher frequency components contribute more compared to images with lower spatial information. Hence, such methods [41] utilise the shift and aliasing properties of the Fourier transform to obtain a high-resolution representation of the image. Kim *et al*. [21] further extended the concept in [41] to take into account noise and spatial blurring present in the input image. In a related development, in [8], Tikhonov regularization was exploited to carry out SR in the frequency domain.

Modern single-image methods, often based upon learning, also known as example-based single image SR aim at learning the relationship between low resolution (LR) and high resolution (HR) images by training with LR and HR image pairs. Dong *et al*. [11] 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 [45] or anchored neighborhood regression [39]. Kim *et al*. [19] go deeper with a network based on VGG-net [34]. The network in [19] is comprised of 20 layers so as to exploit the image context across larger image regions. More recently, thanks to some of the recent benchmarks on example-based single image SR [38, 40, 7], several algorithms were introduced for super-resolving images [25, 12, 4, 3, 16]. These algorithms can be directly used on multi-spectral images, however, as applied to snapshot mosaic sensors, they do not take into account the spectral correlation of different channels nor the spatial offset of each pixel.

Despite the fact that modern multi-spectral cameras are more adversely affected by resolution constraints than regular RGB cameras, there are not many works specifically on CNN based multi-spectral SR. Example-based learning methods are limited mainly due to the lack of multi-spectral SR benchmarking platforms and difficulty accessing suitable SR spectral datasets. For example, [24], which focuses on hyperspectral SR and not multi-spectral SR, is among one of the few example-based spectral SR methods. The only directly related multi-spectral SR methods [23, 31], to the best of our knowledge, were recently introduced through the PIRM2018 spectral SR challenge [32, 33]. The work proposed by Lahoud *et al*. [23], used an image completion technique followed by 12 convolutional layers to super-resolve images. The second work, [31] by Shi *et al*., proposes a deep residual network with channel attention (RCAN) to super-resolve images. The former method [23], unlike the latter [31] involves some image pre-processing and is not an end-to-end CNN implementation. The RCAN network exploited in [31], has also exhibited state-of-the-art performance in the context of RGB image SR [46]. Both of these works take into account spectral correlation, and do not consider the spatial offsets of each wavelength channel.

While all the above exploit demosaiced (debayered or interpolated) images as their LR/HR pair, a very recent work by Fu *et al*. [14] exploits the mosaic RGB images directly to super-resolve hyperspectral images using a variational method. The work was preceded by the SR work by Zhou *et al*. [47] who presented a deep residual network for RGB SR that uses mosaic images. They highlighted the fact that demosaicing, which involves some sort of interpolation (such as bicubic), introduces artifacts that can deteriorate SR performance.

## 3. Proposed Method

As mentioned earlier, the method presented here is quite general in nature. For the sake of generality we view the problem at hand as that of super-resolving and chromatically mapping images with missing or unevenly distributed wavelength bands. To this end, we propose to investigate the structure of the RCAN network in [31] as a baseline, for the combined task of image SR and color-prediction. Moreover, we propose an additional texture network as a means to re-introduce lost information about these bands in the scene. In addition, we notice that the concept of using mosaic images can be extended to multi-spectral images as well, while to the best of our knowledge there is no work reported on using mosaic multi-spectral SR. As a result of this treatment, we can also capitalise on the on-sensor spatial arrangement of the wavelength indexed channels on the mosaic images to improve the SR and colour-prediction performance of our proposed network instead of using demosaiced imagery as input.

### 3.1. Texture Sensitive Residual Channel Attention Network (TSRCAN)

As depicted in Figure 2, our network consists of an RCAN network, and a texture network structure. The RCAN network [46, 31] encompasses three main parts, the head, the body, and the tail. The head of the network carries out feature extraction via two convolutional layers. The body is comprised of $g$ number of sequential residual groups as the body of the network. Each residual group contains $b$ number of residual channel attention networks, each constituting a residual block which incorporates within it a channel attention network (CA). The tail of the RCAN network is the reconstruction part which consists of only one convolutional layer to produce an output with the desired dimension of RGB images. The RCAN network, on its own, given the $LR_{MS}$ and $HR_{RGB}$ pairs, can do a relatively modest job to super-resolve and colour-predict the input multi-spectral images. The RCAN network can be shown as

$$SR'_{RGB} = RCAN(I_{LR_{MS}}). \qquad (1)$$

However, our network expands the RCAN by introducing a texture sensitive network (TN) on the output of the RCAN. Its structure constitutes two convolutional layers and a pool-
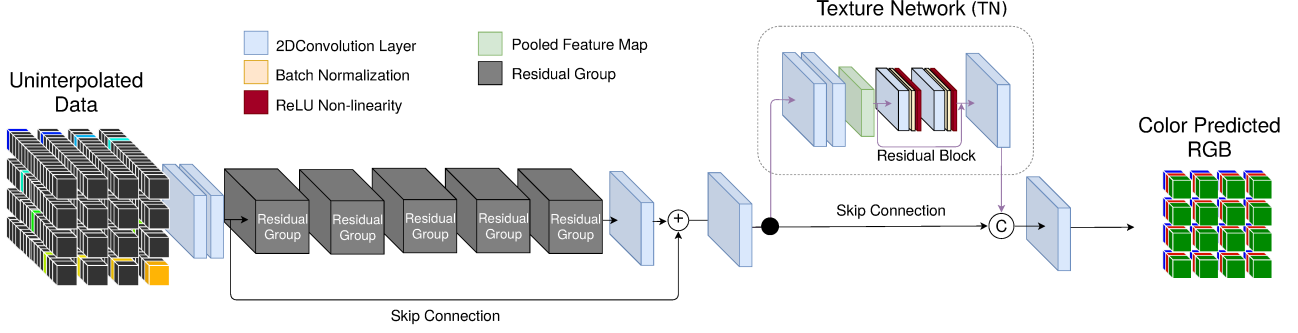
Figure 2. Illustration of the proposed network. The residual group depicted in this figure corresponds to the network of [31]. Channel-wise concatenation is denoted by $C$. The input of the network is uninterpolated data (see Figure 1), and the output is the pseudo-colour image.

ing layer, followed by a residual block. The residual block comprises two stacks of convolution, batch normalization, and ReLU gating. The whole representation is then up-sampled through a deconvolution layer with $k$ number of channels. Each filter in the deconvolution layer represents a texture in the image. The deconvolution layer produces an output of size $k \times w \times h$, where $w$, and $h$ are the width and height of the desired RGB images ($HR_{RGB}$).

The transformation imposed by the TN network is

$$I_{texture} = TN(SR'_{RGB}) \qquad (2)$$

where $I_{texture}$ has the dimensions $k \times w \times h$. $I_{texture}$ is then concatenated with the output of the RCAN network ($SR'_{RGB}$), producing a tensor of dimension $(k+3) \times w \times h$. The concatenation operation ($\oplus$) is expressed as

$$I'_{texture} = I_{texture} \oplus SR'_{RGB}. \qquad (3)$$

Through a convolutional layer, $I'_{texture}$ with $k+3$ channels is reduced to the 3 channels required to produce RGB images. Lets express the operation of this convolutional layer by the transformer $CONV(.)$. The relationship between the input image $I_{LR_{MS}}$, and the output $SR_{RGB}$ image can be expressed using the transfer function $TSRCAN(.)$ as

$$\begin{aligned} SR_{RGB} &= TSRCAN(I_{LR_{MS}}) \\ &= CONV( \\ &\quad TN(RCAN(I_{LR_{MS}})) \oplus RCAN(I_{LR_{MS}}) \\ &\quad ) \end{aligned} \qquad (4)$$

## 3.2. Loss functions

In the CNN based SR literature, a simple loss function such as $L_1$[46] or $L_2$[31, 23] is usually utilized to train models. An $L_1$ function is less sensitive to outliers compared to an $L_2$ function, and our dataset, which uses registered stereo images, is prone to outliers due to inherent registration artifacts. Hence, we choose the $SmoothL1$ [2] PyTorch[1] implementation which is a more stable implementation compared to vanilla $L_1$. For simplicity of notation, we refer to
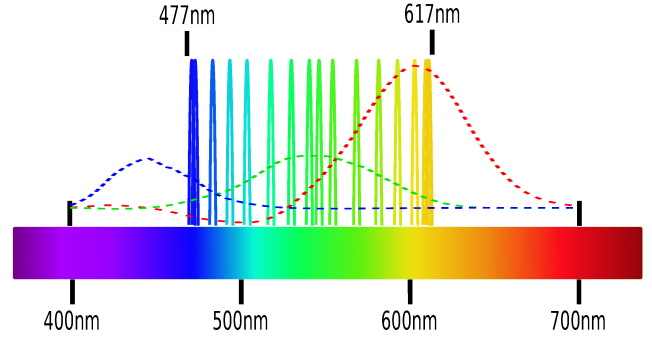


Figure 3. Representation of the visual wavelength against the 16 wavelength bins (sub-pixels) of the multi-spectral camera, and the ideal CMF. The wavelengths for the 16 channels of the multi-spectral camera are $477.2nm$, $478.2nm$, $489.5nm$, $500.3nm$, $510.9nm$, $523.2nm$, $537.9nm$, $548.9nm$, $553.0$, $562.5$, $577.3nm$, $590.5nm$, $599.9nm$, $612.9nm$, $615.9nm$, $617.5nm$

the $SmoothL1$ function as $L_{1s}$, which can be expressed as

$$L_{1s}(\Theta) = \frac{1}{N} \sum_{i=1}^{M} Z^i \qquad (5)$$

where

$$Z^i = \begin{cases} 0.5 \times (DIF)^2 & if\,|DIF| < 1 \\ |DIF| - 0.5 & otherwise, \end{cases}$$

and $DIF = HR^i_{RGB} - LR^i_{MS}$.

In addition, we know that (through experiments) RCAN by itself is capable of learning SR and also the colour relationship between input and output images. Hence, the output of RCAN and input of the TN, is also expected to be similar to $I_{HR_{RGB}}$. Therefore, we choose to minimise the cost function

$$L^c_{1s} = L^c_{1s}(SR'_{RGB} - I_{HR_{RGB}}) + L^c_{1s}(SR_{RGB} - I_{HR_{RGB}}). \qquad (6)$$

Figure 3 displays the wavelength range of our multi-spectral images vs the wavelength range of the visual light.

It also shows the relative amplitude of a CMF for three channels. It is obvious that our multi-spectral images have incomplete blue and red channels, which is one of the main drivers behind this work. We believe that the above loss functions, along with our proposed network, can predict the missing channels and hence improve the colour-prediction performance.

### 3.3. Implementation Details

Now we specify the implementation details of our proposed TSRCAN. The RCAN part of our network has $g = 5$ residual groups. Each residual group contains $b = 3$ residual channel attention blocks (RCAB). The channel attention, similar to [46], has a 64 channel input and 64 weighted channel output with a reduction factor of 16. The kernel size of all our convolutional layers are set to $3 \times 3$. Convolutional layers in shallow feature extraction and the body have $c = 64$ filters, except at the tail of the RCAN where channels are reduced to 3.

The TN structure constitutes a convolutional layer, followed by batch normalization, ReLU, maxpooling, and a residual block similar to that of RCAN but without a channel attention mechanism. In fact, the texture network is identical to the first few layers of the Resnet-18 structure, and we only remove the last layers up to the first residual block. This is followed by a convolutional layer with $k = 256$ channels to achieve a tensor with the size $256 \times 576 \times 1152$. After concatenating this tensor with $SR'_{RGB}$, the last layer, a convolutional layer with 3 filters produces the desired output dimensions of $3 \times 576 \times 1152$.

### 3.4. Zero padded, uninterpolated data

As explained in the introduction, interpolation of mosaic images gives rise to artifacts. For example, SR CNN based methods such as VDSR [19], and SRCNN [20] that first interpolate the input LR images up to the scale of the HR images suffer from these artifacts via losing information and decreasing computational efficiency [46]. Hence, inspired by the procedures in [14, 47], where authors super-resolved hyperspectral [14], and RGB images [47], using RGB bayer patterns, we choose not to interpolate the multi-spectral image. Instead, we use the mosaic pattern in the manner presented in Figure 1. The mosaic multi-spectral pattern in Figure 1(c) represents 1 multi-spectral pixel which constitutes 16 sub-pixels of 16 wavelength channels. To transform the mosaic multi-spectral input to a format that is suitable to be consumed by the network, and to avoid interpolation, we take the following approach. We generate an image with size $16 \times 576 \times 1152$, that is a multi-spectral image with height and width of the mosaic image, but with 16 channels. For each channel the value of respective sub-pixel is used and another 15 sub-pixels are added and set to zero. This process, for 1 multi-spectral pixel alone for ease of il-

lustration, is shown in Figure 1(c-d).

## 4. Experiments

### 4.1. Dataset description[2]

We carry out our experiments using our 296 registered stereo pair multi-spectral/RGB images which were collected from a diverse range of environments. During acquisition time, no gamma correction was applied to the images. In addition, since the stereo pairs were captured using cameras with different image mosaic sensors, the exposure time was optimised for each camera individually for optimum image quality. One is an RGB camera and the other is a multi-spectral camera covering the visible wavelength range $(477 - 617nm)$. The RGB camera has a CMOS image sensor with a $2 \times 2$ mosaic (bayer) pattern delivering the three RGB channels whereas the mosaic sensor of the multi-spectral camera has a $4 \times 4$ pattern delivering 16 wavelength bands. Hence, the resolution of the RGB images in each axis is twice that of the spectral images. Figure 1 illustrates this resolution relationship between the two filter arrays on both cameras. The original images were interpolated and converted to grayscale for registering using PWC-Net [36], the state-of-the-art optical flow algorithm. Original multi-spectral and the registered RGB were then cropped to the size $576 \times 1152$ to minimise optical flow artifacts on the border of the images which also led to training acceleration. We split our 296 image pairs to 250 image pairs for training, 25 for validation and 21 for testing. For each image pair, the multi-spectral image with lower spectral and spatial resolution is referred to as $LR_{MS}$ and the registered RGB image with higher spatial resolution is referred to as $HR_{RGB}$.

### 4.2. Analysis of the effect of occlusions

We train our CNN using the $LR_{MS}$ and its registered $HR_{RGB}$ pair. However, with every registration, there are some artifacts including wrong registration and occlusions [42]. We hypothesise that if these artifacts are abundant, they could affect the training process. To check if errors of the above nature could affect the training process, we take the following approach. We calculate optical flow from the multi-spectral image to the RGB image and vice-versa using the [36] algorithm. A straightforward way to detect erroneous flow and occlusions is to calculate the euclidean distance between the two optical flows and remove the pixels with errors larger than a threshold [42]. Thereby, we removed pixels with errors larger than 3 pixels, and created a mask for each image. We multiplied this mask with $LR_{MS}$, $HR_{RGB}$, and the output of respective model ($SR_{RGB}$). Af-

---

ter carrying out several experiments, we did not see a strong correlation between removing the occlusions and improvements in the results. Hence, we believe that either occlusions do not have a significant adverse effect on the training process (which can be attributed to good registration of the images) or substantially more advanced occlusion detection techniques are required to remove their effect. To avoid this topic turning into a research subject of its own, in this work, we decide to acknowledge but not address their effect further in this paper.

## 4.3. Settings

**Evaluation metrics:** The 21 test images were super-resolved and colour-predicted and then evaluated using Pixel Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Spectral Information Divergence (SID) per channel. As a reference metric, as is customary with SR evaluations, we also present the results for bicubic up-sampled images where the images were colour-predicted using an ideal CMF function [30]. We compare the results from our TSRCAN network to the conventional method and the baseline RCAN network.

**Training settings:** During training, we performed data augmentation on our batches of 10 images of our 250 training images, which included random cropping with size $120 \times 120$, random rotation by $0°$, $90°$, $180°$, $270°$ with $p = 0.25$, and random horizontal flip with $p = 0.5$. The batch size is fixed at 10. Our model is trained by ADAM optimizer [22] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The initial learning rate is set to $10^{-4}$ and then halved every 2500 epochs. To implement our models we used PyTorch [1], and in particular, to implement our $L_1$ functions, the $SmoothL_1$ function [2] was used as the main building block. To test our algorithms, we select the models with the best performance on the validation dataset, and present the test results for those models.

## 4.4. Baseline RCAN Network vs the conventional method

Table 1 presents the numerical results of our ablation studies. The first two rows compares the performance of the RCAN network (a deep learning approach) vs the traditional method of bicubic upsampling followed by a CMF transfer. There are a number of different CMF available due to the fact that the combination of light wavelengths to produce a given perceived color is not unique [35], and there is a degree of subjectivity in drawing a given CMF function [35]. Hence, to calculate the metrics presented here, only for the conventional method, we normalize all the three R-G-B channels of the ground truth image and the results obtained using the conventional method between 0 and 255

| Method | PSNR (dB) | SSIM | SID | | |
|---|---|---|---|---|---|
| | | | Blue | Green | Red |
| Bicubic + CMF | 22.375 (3.40) | 0.779 (0.102) | 7.29e-05 (3.39e-4) | 5.72e-05 (1.79e-4) | 7.43e-05 (1.96e-4) |
| RCAN* | 24.78 (4.62) | 0.814 (0.093) | 5.53e-05 (4.38e-05) | 4.18e-05 (3.34e-05) | 4.38e-05 (3.90e-05) |
| RCAN (Ours) | 24.90 (3.50) | 0.847 (0.0812) | **4.90e-05** (3.64e-05) | **3.63e-05** (3.25e-05) | 4.147e-05 (4.38e-05) |
| TSRCAN (Ours) | **26.02** (4.59) | **0.855** (0.095) | 5.74e-05 (6.29e-05) | 3.99e-05 (3.83e-05) | **3.53e-05** (2.81e-05) |

Table 1. Mean and standard deviation (in parenthesis) of PSNR, SSIM, and SID obtained using different models. RCAN* denotes RCAN trained with data without zero padding.

before calculating the metrics. In spite of this, considerable improvements can be observed with the trained baseline RCAN method compared to the conventional method.

## 4.5. Effect of using zero padded data



Figure 4. Qualitative comparison of the results for training RCAN with zero-padded uninterpolated data (left) with training RCAN with uninterpolated data without zero-padding [32] (right) which is denoted as RCAN* in Table 1.

To assess the effect of zero padded uninterpolated data we need to compare it with a baseline. We chose to train the RCAN network not only with the uninterpolated data, but also with the data format the does not take into account the spatial location of each pixel. In this way, the multi-spectral mosaic pixel in Figure 1(c) with dimension $4 \times 4 \times 16$ translates to $1 \times 1 \times 16$. In other words, we remove the zero padded sub-pixels in Figure 1(d). In fact, this is the approach that was taken in the algorithms introduced in [32]. In Table 1, the results obtained using this method is presented as RCAN*. Comparing the results of RCAN with RCAN*, it can be seen that with zero-padded uninterpolated data, while displaying an improvement across all metrics, the improvements in PSNR is minimal. However, looking at Figure 4, it is obvious that the quality of the images produced by RCAN are superior to that of RCAN*. We can reaffirm the notion that, while PSNR remain a descent measure of image quality, it does not provide an accurate measure of perceptual quality of the image [43] compared to SSIM and SID [9].
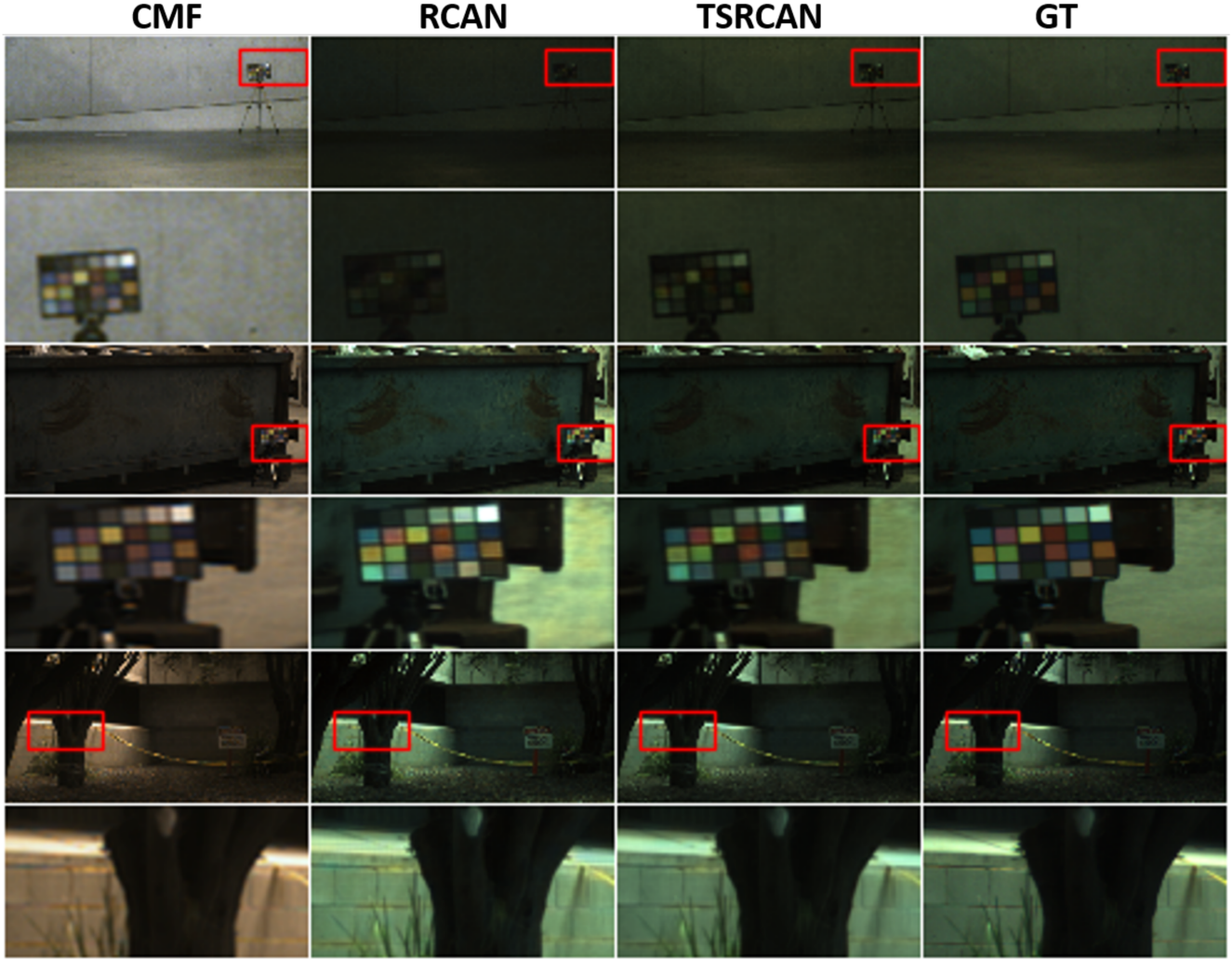
Figure 5. Visual representation of the performance of three methods: the conventional method (white-balanced), RCAN, and TSRCAN along with the ground truth with diverse lighting environments. For better visualization, we show a zoomed in area of the images in the even rows.
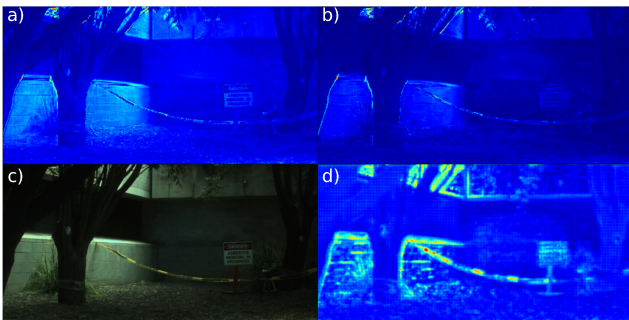


Figure 6. Effect of the TN. For an image from the test set (a) shows the error map at the output of RCAN, (b) shows the error map at the output of TSRCAN, (c) depicts the output image from TSR-CAN, and (d) illustrates the activation map at the output of TN.

## 4.6. Effect of Texture Network (TN)

We trained TSRCAN from scratch. A considerable improvement of $1.1dB$ in PSNR performance is observed compared to the baseline RCAN network performance. SSIM also displays some improvements. Regarding the SID metric, there are some improvements in the red channel compared to the baseline RCAN network but not in the green or the blue channels. Also, the qualitative results in Figure 5 evidence the superiority of our network relative to the baselines. To explain the overall improvement, we look into the activation map at the output of the TN. Figure 6 presents, for an image from the test set, (a) the error map at the output of RCAN, (b) the error map at the output of TSRCAN, (c) output image of TSRCAN, and (d) the activation map at the output of the TN. It seems generally, in
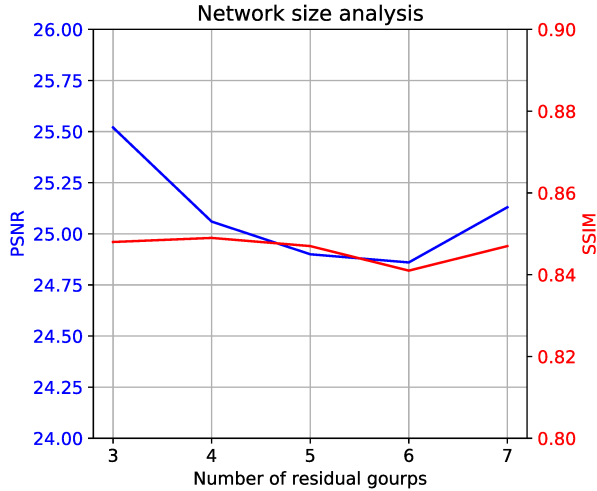
Figure 7. Effect of the number of residual groups on the performance of the network in terms of PSNR and SSIM on the test textures/pixels that RCAN produces a larger error, the activation map is more "active", leading to less error for TSR-CAN in the same areas/pixels.

## 4.7. Effect of network size

Figure 7 displays PSNR and SSIM results for RCAN networks with different numbers of residual groups. There is no clear correlation between the number of residual groups and the performance of the network. For example, as the number of residual groups increases from 5 to 6 and 7, the PSNR improves only negligibly by $0.27dB$. Also, SSIM results deteriorate slightly and then bounce back slightly. For 4, and 3 residual groups, the improvement in PSNR is 0.16dB, and 0.62dB respectively. In addition, these results do not correlate with the SSIM results in which the baseline we use (with 5 residual groups) exhibits approximately the mean of all the RCAN versions. Our choice of using five residual groups was due to the fact that the most recent work on multi-spectral SR, based on RCAN, used 5 residual groups, and we chose to investigate that model as our baseline [31]. These observations, indeed, highlight the effectiveness of the texture network module in our network (TSRCAN) which exhibited a 1.1dB improvement in PSNR compared to the baseline RCAN.

## 4.8. Spectral Information Divergence (SID)

A close observation of SID per channel results in Table 1 shows that the blue channel generally exhibits a larger SID error, followed by the red channel, and with the green channel having the lowest SID. This is true except for the results obtained using the conventional method, where we white-balanced the channels to calculate the metrics. These results correlate with the fact that the multi-spectral camera has a large portion of the blue wavelength range missing (see Figure 3) compared to the red wavelengths. The

green channel exhibits the lowest SID because the multi-spectral camera covers the whole green wavelength range, although sparsely. The reddish appearance in the white-point balanced images transferred using the conventional method can also be explained by the fact that the camera has a larger portion of the blue wavelengths missing compared with the blue wavelengths. Therefore, leading to an exaggerated contribution of the red channel compared to the blue channel.

## 4.9. Effect of poor lighting conditions

In Figure 5, we include scenarios with diverse lighting conditions. Specifically, the first row of the figure includes an indoor image with poor lighting which implies an incomplete wavelength spectrum. To elaborate, the camera channels depicted in Figure 3 can be thought of as sampling wavelengths which already do not cover the complete visible wavelength range. On top of this, poor lighting conditions results in fewer efficient samples. For example, with the indoor image in the first row, most of the wavelength spectrum is likely to be emitted from fluorescent lights (which is known for producing a poor, nonuniform, sparse spectrum) and some leakage of outdoor light which results in a poor spectrum. Hence, there are much fewer wavelength samples and the network is having a harder time predicting the unknown wavelengths, and hence the R-G-B values of multi-spectral pixels. Our algorithm produces somewhat decent results for this difficult scenario, which means that the network is doing a good job in learning to predict the unknown wavelengths. However, the performance can be improved by expanding the dataset with images taken in a controlled laboratory environment to include more examples of poorly lit conditions. Given that the dataset contains 296 image pairs to train and test our networks, expanding this data set is a perfectly feasible task. In fact, this is a future work that we are planning to carry out.

## 5. Conclusion

In this paper we proposed a novel deep learning approach that addresses the ill-posed problem of producing RGB images from spectrally and spatially under-sampled multi-spectral images, which significantly outperforms, quantitatively and qualitatively, conventional methods as well as the state-of-the-art RCAN network. Moreover, the method is quite general in nature and can be applied to multi-spectral images of low spatial and spectral resolution with unevenly spaced or missing channels. Our approach uses a texture sensitive block to enable the network so as to re-introduce information from missing wavelength bands that may be still implicitly available in the texture of the image. In addition, we have introduced a novel dataset consisting of 296 registered stereo multi-spectral/RGB image pairs.

# References

[1] Pytorch v0.4.0. https://pytorch.org/docs/0.4.0/. Last accessed: 2019-03-22.

[2] Smooth1 loss function. https://pytorch.org/docs/stable/nn.htmlsmoothl1loss. Last accessed: 2019-03-22.

[3] N. Ahn, B. Kang, and K.-A. Sohn. Image super-resolution via progressive cascading residual network. *progressive*, 24:0–771, 2018.

[4] Y. Bei, A. Damian, S. Hu, S. Menon, N. Ravi, and C. Rudin. 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.

[5] J. F. Bell, D. Wellington, C. Hardgrove, A. Godber, M. S. Rice, J. R. Johnson, and A. Fraeman. Multispectral imaging of mars from the mars science laboratory mastcam instruments: Spectral properties and mineralogic implications along the gale crater traverse. In *AAS/Division for Planetary Sciences Meeting Abstracts*, volume 48, 2016.

[6] M. Bigas, E. Cabruja, J. Forest, and J. Salvi. Review of cmos image sensors. *Microelectronics journal*, 37(5):433–451, 2006.

[7] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor. 2018 PIRM challenge on perceptual image super-resolution. In *European Conference on Computer Vision Workshops (ECCVW)*, 2018.

[8] N. Bose, H. Kim, and H. Valenzuela. Recursive total least squares algorithm for image reconstruction from noisy, undersampled frames. *Multidimensional Systems and Signal Processing*, 4(3):253–268, 1993.

[9] C.-I. Chang. Spectral information divergence for hyperspectral image analysis. In *Geoscience and Remote Sensing Symposium, 1999. IGARSS'99 Proceedings. IEEE 1999 International*, volume 1, pages 509–511. IEEE, 1999.

[10] D. Doering, M. Vizzotto, C. Bredemeier, C. da Costa, R. Henriques, E. Pignaton, and C. Pereira. Mde-based development of a multispectral camera for precision agriculture. *IFAC-PapersOnLine*, 49(30):24–29, 2016.

[11] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[12] Y. Fan, H. Shi, J. Yu, D. Liu, W. Han, H. Yu, Z. Wang, X. Wang, and T. S. Huang. Balanced two-stage residual networks for image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1157–1164. IEEE, 2017.

[13] Y.-Z. Feng and D.-W. Sun. Application of hyperspectral imaging in food safety inspection and control: a review. *Critical reviews in food science and nutrition*, 52(11):1039–1058, 2012.

[14] Y. Fu, Y. Zheng, H. Huang, I. Sato, and Y. Sato. Hyperspectral image super-resolution with a mosaic rgb image. *IEEE Transactions on Image Processing*, 27(11):5539–5552, 2018.

[15] A. F. Goetz. Three decades of hyperspectral remote sensing of the earth: A personal view. *Remote Sensing of Environment*, 113:S5–S16, 2009.

[16] M. Haris, G. Shakhnarovich, and N. Ukita. Deep backprojection networks for super-resolution. In *Conference on Computer Vision and Pattern Recognition*, 2018.

[17] M. Hasan, X. Jia, A. Robles-Kelly, J. Zhou, and M. R. Pickering. Multi-spectral remote sensing image registration via spatial relationship analysis on sift keypoints. In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 1011–1014. IEEE, 2010.

[18] S. P. Jaiswal, L. Fang, V. Jakhetiya, J. Pang, K. Mueller, and O. C. Au. Adaptive multispectral demosaicking based on frequency-domain analysis of spectral correlation. *IEEE Transactions on Image Processing*, 26(2):953–968, 2016.

[19] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[20] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.

[21] S. Kim, N. K. Bose, and H. Valenzuela. Recursive reconstruction of high resolution image from noisy undersampled multiframes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(6):1013–1027, 1990.

[22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[23] F. Lahoud, R. Zhou, and S. Süsstrunk. Multi-modal spectral image super-resolution. In *European Conference on Computer Vision*, pages 35–50. Springer, 2018.

[24] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 266:29–41, 2017.

[25] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. In *The IEEE conference on computer vision and pattern recognition (CVPR) workshops*, volume 1, page 4, 2017.

[26] L. Miao, H. Qi, R. Ramanath, and W. E. Snyder. Binary tree-based generic demosaicking algorithm for multispectral filter arrays. *IEEE Transactions on Image Processing*, 15(11):3550–3558, 2006.

[27] Y. Monno, S. Kikuchi, M. Tanaka, and M. Okutomi. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10):3048–3059, 2015.

[28] Y. Monno, M. Tanaka, and M. Okutomi. N-to-srgb mapping for single-sensor multispectral imaging. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 33–40, 2015.

[29] M. Najafi, S. T. Namin, and L. Petersson. Classification of natural scene multi spectral images using a new enhanced crf. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3704–3711. IEEE, 2013.

[30] A. Robles-Kelly and C. P. Huynh. *Imaging spectroscopy for scene analysis*. Springer Science & Business Media, 2012.

[31] Z. Shi, C. Chen, Z. Xiong, D. Liu, Z.-J. Zha, and F. Wu. Deep residual attention network for spectral image super-resolution. In *European Conference on Computer Vision Workshops (ECCVW)*, 2018.

[32] M. Shoeiby, A. Robles-Kelly, R. Timofte, R. Zhou, F. Lahoud, S. Susstrunk, Z. Xiong, Z. Shi, C. Chen, D. Liu, Z.-J. Zha, F. Wu, K. Wei, T. Zhang, L. Wang, Y. Fu, Z. Zhong, K. Nagasubramanian, A. K. Singh, A. Singh, S. Sarkar, and G. Baskar. PIRM2018 challenge on spectral image super-resolution: Methods and results. In *European Conference on Computer Vision Workshops (ECCVW)*, 2018.

[33] M. Shoeiby, A. Robles-Kelly, R. Wei, and R. Timofte. Pirm2018 challenge on spectral image super-resolution: Dataset and study. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.

[34] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[35] T. Smith and J. Guild. The cie colorimetric standards and their use. *Transactions of the optical society*, 33(3):73, 1931.

[36] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz. Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8934–8943, 2018.

[37] K. Takumi, K. Watanabe, Q. Ha, A. Tejero-De-Pablos, Y. Ushiku, and T. Harada. Multispectral object detection for autonomous vehicles. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, pages 35–43. ACM, 2017.

[38] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, L. Zhang, B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, 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*, pages 1110–1121. IEEE, 2017.

[39] R. Timofte, V. De Smet, and L. Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Asian Conference on Computer Vision*, pages 111–126. Springer, 2014.

[40] R. Timofte, S. Gu, J. Wu, L. Van Gool, L. Zhang, M.-H. Yang, M. Haris, G. Shakhnarovis, N. Ukita, H. Shijia, 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.

[41] R. Tsai. Multiframe image restoration and registration. *Advance Computer Visual and Image Processing*, 1:317–339, 1984.

[42] Y. Wang, Y. Yang, Z. Yang, L. Zhao, P. Wang, and W. Xu. Occlusion aware unsupervised learning of optical flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4884–4893, 2018.

[43] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117, 2009.

[44] D. Wu and D.-W. Sun. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—part i: Fundamentals. *Innovative Food Science & Emerging Technologies*, 19:1–14, 2013.

[45] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.

[46] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision*, pages 294–310. Springer, 2018.

[47] R. Zhou, R. Achanta, and S. Süsstrunk. Deep residual network for joint demosaicing and super-resolution. In *Color and Imaging Conference*, volume 2018, pages 75–80. Society for Imaging Science and Technology, 2018.