

# Deeply Learned Filter Response Functions for Hyperspectral Reconstruction

Hongzhi Liu

July 6, 2018

## Abstract

*Recently, Hyperspectral reconstruction from RGB imaging has achieved significant progress via sparse coding and deep learning. However, a largely ignored fact is that existing RGB cameras are tuned to mimic human trichromatic perception, thus their spectral responses are not necessarily optimal for hyperspectral reconstruction. Today, I read a thesis written by Shijie Nie who is from National Institute of Informatics. His team introduce two types of designed filters that are a three-chip setup without spatial mosaicing and a single-chip setup with a Bayer-style  $2 \times 2$  filter array. Numerical simulations verify the advantages of deeply learned spectral responses compared to existing RGB cameras. Thus they are able to realize the deeply learned spectral response functions by using modern film filter production technologies and thus construct data inspired multispectral cameras for snapshot hyperspectral imaging.*

## 1. Overview of Tracking Model

Hyperspectral imaging captures detailed light distribution along the wavelength axis. It is shown to be beneficial for remote sensing, medical diagnosis, industrial detection and so on. Most existing devices to capture hyperspectral images are scanning based, that is, either to drive a line slit along one spatial dimension or to continuously change narrow bandpass filters in front of a grayscale camera. The key drawback is that scanning is slow, which prevents their application to dynamic scenes. Thus scanning-free, snapshot hyperspectral devices have been developed but these devices are extremely limited in spatial resolution.

A computational hyperspectral reconstruction method from a single RGB image is promising in overcoming the drawbacks of the aforementioned devices, as evidenced in recent research on RGB-to-Spectrum reconstruction [1]. Existing RGB cameras are designed to mimic human color perception [5], thus their spectral response functions are not necessarily optimal for computer vision tasks, *i.e.* hyperspectral reconstruction.

In this paper, Shijie Nie and his team aim to directly learn

optimized spectral response functions in the infinite space of non-negative and smooth functions [8]. They then manufacture their learned filter, based on the data-driven approach to construct a multispectral camera for snapshot hyperspectral imaging. Based on observation that camera spectral filters act in effect, they are able to optimize them by using deep learning techniques. The team realize the deeply learned filters by using interference film production technologies and construct a snapshot hyperspectral imaging system.

## 2. Design-Realization-Application Framework

In the paper mentioned above, Nie advocates a novel design-realization-application framework for data-inspired and task-oriented spectral imaging hardware development, which is illustrated in Fig. 1. As for the data-driven design stage, his team construct an end-to-end network by appending a tailored convolution layer onto the spectral reconstruction network. Since they properly incorporate the nonnegativity and smoothness constraints, the convolution layer acts in effect as the filter spectral response functions that they aim to design. It encodes an input hyperspectral image into the most appropriate hidden feature map, such that the succeeding reconstruction network can recover the original input hyperspectral image as faithfully as possible. In this sense, their end-to-end network is similar to the autoencoder-decoder.

In the realization stage, Nie's team try to physically realize the deeply learned response functions by using film filter production technologies. Besides, in the online application stage, they use the customized multispectral camera to capture images and directly import them into the already trained reconstruction network to generate hyperspectral images. Therefore, their reconstruction will in principle share all the benefits arising from deep learning as mentioned in [4]. Although their framework is presented for spectral reconstruction, we believe it can also be generalized to many other data-driven hardware design tasks.

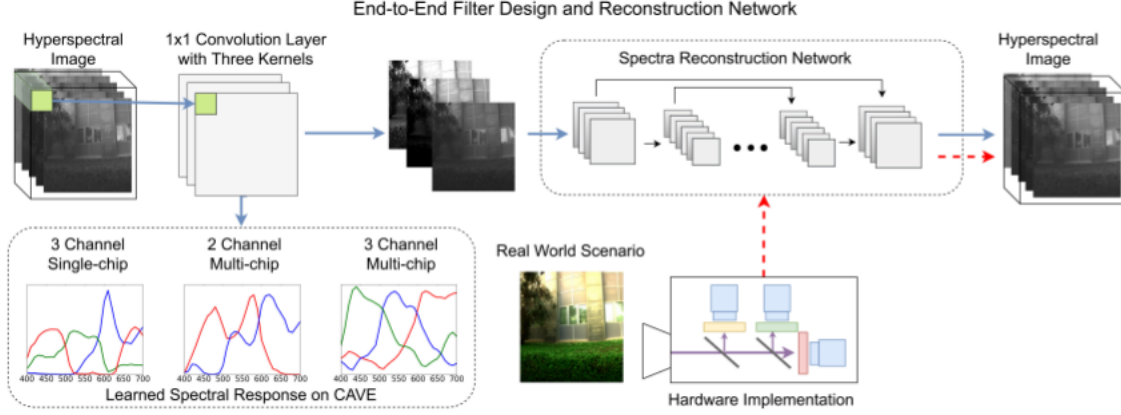


Figure 1. Their proposed design-realization-application framework for data-inspired spectral imaging hardware. The design stage (marked in blue arrow) is data-driven. It includes an end-to-end network to simultaneously learn the filter response and the spectral reconstruction mapping. The learned spectral response function on CAFE dataset is also shown. In the realization stage (marked in red arrow), the learned response functions are realized by using film filter production technologies, and a data-inspired multispectral camera is constructed. In the online application stage, the captured multispectral image is imported into the already trained spectral reconstruction network to generate hyperspectral images. This framework is illustrated using the multi-chip setup with three channels.

### 3. Filter Design and Spectral Reconstruction

Nie gives the details on the end-to-end network for simultaneous filter response design and spectral reconstruction.

#### 3.1. Spectral Reconstruction Network

For the sake of generality, Nie constructs a spectral reconstruction network by adapting the well-known U-net architecture. They use modules formed as 2D convolution-BatchNorm-Relu. The network takes images of size  $256 \times 3$  as input and finally produces the corresponding spectral images of size  $256 \times 256 \times 31$ . Let  $C_k$  denote a convolutional block including one convolutional layer with  $k$  filters, one leakyReLU activation layer, one BatchNorm-normalization layer. The convolutional layer in each  $C_k$  has  $3 \times 3$  sized kernels with stride 2. The downsampling factor is 2, with proper zero padding to edges. The  $\alpha$  parameter in the leakyReLU layer is set to 0.2.  $CD_k$  denotes the same block as  $C_k$ , except that the convolution layer is replaced by the deconvolution layer. It upsamples the input by a factor of 2 as well. A dropout layer with 50% dropout rate is added after each block.

#### 3.2. Filter Spectral Response Design

As shown in Fig. 1, one key novelty mentioned in paper is in drawing the connection between camera color imaging formulation and a convolutional layer. This allows them to optimize the spectral imaging parameters by using existing network training algorithms and tools. For simplicity, they will assume that the CCD/CMOS sensor has an ideal flat response temporarily and will address this factor when constructing a real system.

Given the spectral radiance  $L(x, y, \lambda)$  at position  $(x, y)$ , the recorded intensity by a linear sensor coupled with a color filter is given as Eq. 1:

$$I_c(x, y) = \int_{\lambda} S_c(\lambda) L(x, y, \lambda) d\lambda. \quad (1)$$

where  $\lambda$  is the wavelength and  $S_c(\lambda)$  is the spectral response function of the color filter. In most commercial cameras, there are red-green-blue trichromatic filters, *i.e.*  $c \in \{R, G, B\}$ , so as to mimic the human color perception. In practice, the above equation could be discretely approximated as Eq. 2:

$$I_c(x, y) = \sum_{n=1}^N S_c(\lambda_n) L(x, y, \lambda_n). \quad (2)$$

where the filter response is in the form of a vector  $S_c = [S_c(1), S_c(2), \dots, S_c(\lambda_N)]$  at sampled wavelengths and  $N$  is number of spectral channels. With the appended layer, the team train this end-to-end network with the  $N$ -channel hyperspectral images as input.

#### 3.3. Nonnegative and Smooth Response

There are various regularizers in convolutional neural network, which were originally designed to penalize the layer parameters during training. Interestingly, their non-negativity and smoothness constraints on the spectral response functions could be easily enforced by borrowing those regularizers.

To achieve nonnegative responses, Nie's team enforce a nonnegative regularizer on the kernel in their filter design convolution layer, such that  $S_c(\lambda) \geq 0$ . As for the

Table 1. Average and Variance of RMSE of reconstruction on the hyperspectral databases [9, 3].

	CAVE [9]	Harvard Natural [3]	Mixed [3]
Theirs	$4.48 \pm 2.97$	$7.57 \pm 4.59$	$8.88 \pm 4.25$
[2]	$8.84 \pm 7.23$	$14.89 \pm 13.23$	$9.74 \pm 7.45$
[7]	$14.91 \pm 11.09$	$9.06 \pm 9.69$	$15.61 \pm 8.76$
[6]	$7.92 \pm 3.33$	$8.72 \pm 7.40$	$9.50 \pm 6.32$

smoothness constraint, they use the L2 norm regularizer, which is commonly used to avoid over-fitting in deep network training. Specifically, they introduce a regularization term  $\eta \sqrt{\sum_{n=1}^M (S_c(\lambda_n))^2}$ , where  $\eta$  controls the smoothness. Throughout the experiment,  $\eta$  is set to 0.02.

#### 4. Experiment Results of the Method

Nie and his team conduct experiments on synthetic data to demonstrate the effectiveness of our method. We evaluate our method on the dataset comprising of both natural and indoor scenes. They run their proposed algorithms on an NVIDIA GTX 1080 GPU. Our server is equipped with an Intel(R) Core(TM) i7-6800K CPU @ 3.40GHz and 128GB of memory. The training time for the CAVE, Harvard Natural and Mixed [3] datasets take 1.84, 8.88 and 8.52 hours, respectively. The average time to reconstruct spectra from an individual image takes about 5.83 seconds.

The average and variance of the RMSE are shown in Tab. 1, which was compared with three baseline methods that are [2], [7] and [6]. The RGB inputs of three baseline methods are generated from the spectral response function of Cannon 600D. This table shows that the RMSE of Nie’s method outperforms the alternative methods in spectral reconstruction in all three datasets.

#### References

- [1] A. Alvarez-Gila, J. V. D. Weijer, and E. Garrote. Adversarial networks for spatial context-aware spectral image reconstruction from RGB. In *ICCVW*, 2018. 1
- [2] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural RGB images. In *ECCV*, 2016. 3
- [3] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. In *CVPR*, 2011. 3
- [4] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler. Learned spectral super-resolution. *arXiv preprint arXiv:1703.09470*, 2017. 1
- [5] J. Gu, J. Jiang, S. Susstrunk, and D. Liu. What is the space of spectral sensitivity functions for digital color cameras? In *WACV*, 2013. 1
- [6] Y. Jia, Y. Zheng, L. Gu, A. Subpaasa, A. Lam, Y. Sato, and I. Sato. From RGB to spectrum for natural scenes via manifold-based mapping. In *ICCV*, 2017. 3
- [7] R. M. H. Nguyen, D. K. Prasad, and M. S. Brown. Training-based spectral reconstruction from a single RGB image. In *ECCV*, 2014. 3
- [8] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato. Deeply learned filter response functions for hyperspectral reconstruction. In *CVPR*, 2018. 1
- [9] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE TIP*, 2010. 3