HSCNN: CNN-Based Hyperspectral Image Recovery from Spectrally Undersampled Projections

Zhiwei Xiong[1] Zhan Shi[1] Huiqun Li[1] Lizhi Wang[2] Dong Liu[1] Feng Wu[1]
[1]University of Science and Technology of China
[2]Beijing Institute of Technology

Abstract

This paper presents a unified deep learning framework to recover hyperspectral images from spectrally undersampled projections. Specifically, we investigate two kinds of representative projections, RGB and compressive sensing (CS) measurements. These measurements are first upsampled in the spectral dimension through simple interpolation or CS reconstruction, and the proposed method learns an end-to-end mapping from a large number of upsampled/groundtruth hyperspectral image pairs. The mapping is represented as a deep convolutional neural network (CNN) that takes the spectrally upsampled image as input and outputs the enhanced hyperspetral one. We explore different network configurations to achieve high reconstruction fidelity. Experimental results on a variety of test images demonstrate significantly improved performance of the proposed method over the state-of-the-arts.

1. Introduction

Hyperspectral images play an important role in many fields such as remote sensing, medical diagnosis, and space surveillance [22, 14]. Recently, the availability of rich spectral information is demonstrated to be beneficial to various computer vision and graphic applications, such as object tracking, face recognition, appearance modeling, and relighting [23, 18, 16, 8]. However, conventional spectrometers often operate in 1D line or 2D plane scanning, which is time-consuming and not suitable for dynamic scenes [5].

Ideally, it will be great if hyperspectral images can be directly recovered from the corresponding RGB images. However, this problem is severely ill-posed since a large amount of information is lost during the process of spectral integration when RGB sensors capture the light. To make the problem solvable, there are two kinds of approaches. One is using controlled illumination to illuminate a target scene with several narrow-band light sources, and the scene reflectance can be estimated across a number of wavelengths [13]. The other is using dictionary learning to learn a specific hyperspectral prior, and the reconstruction problem is then formulated in a sparse coding manner [1]. Nevertheless, the former approach requires rigorous conditions of the environment (e.g., difficult to be applied outdoors), while the latter one is limited in generalizability and the dictionary learned from domain-specific images may not perform well on drastically different ones.

In the past decade, researchers have also developed quite a few snapshot hyperspectral imagers with distinct optical designs and computational reconstruction algorithms [9, 24, 7]. Take coded aperture snapshot spectral imaging (CASSI) [2] as an example, CASSI recovers a 3D hyperspectral image from a single 2D measurement by leveraging the compressive sensing (CS) theory [11, 6]. Still, the fidelity of computational reconstruction is limited especially for complex scenes, and the reconstruction algorithms are generally of high complexity.

In this paper, we propose a unified deep learning framework to recover hyperspectral images from spectrally undersampled projections. The most representative projection should be RGB images. Specifically, an RGB image is first upsampled in the spectral dimension through simple interpolation, and the proposed method learns an end-to-end mapping from a large number of upsampled/groundtruth hyperspectral image pairs. The mapping is represented as a deep convolutional neural network (CNN) that takes the spectrally upsampled image as input and outputs the enhanced hyperspetral one. We explore different network configurations to find the optimal one that achieves the highest reconstruction fidelity. A main insight here is that, different from spatial super-resolution, RGB to hyperspectral conversion favors CNNs with a small receptive field in the spatial dimension.

Under the same deep learning framework as explored above, we also investigate hyperspectral image recovery from another representative projection, i.e., CS measurements from the CASSI system. Specifically, a CASSI measurement is first "upsampled" to the desired spectral resolution using a simple CS reconstruction algorithm (TwIST

[4]), and the image details missing in the CS reconstruction are then enhanced by deep learning. A main insight here is that, compared with RGB to hyperspectral conversion, CASSI enhancement favors CNNs with a larger receptive field in the spatial dimension. Moreover, the very different residuals learned through the respective networks reflect how information is lost in these two kinds of projections.

Experiments are conducted on a large hyperspectral dataset provided in [1]. Compared with the state-of-the-art sparse coding method in [1], our deep learning method achieves a 48.7% reduction in terms of the normalized root-mean-square-error (NRMSE) for RGB to hyperspectral conversion. It is worth mentioning that, our method using the network trained from non-domain-specific images outperforms the sparse coding method using dictionaries trained from domain-specific images, which validates the superior generalizability of deep learning. For hyperspectral image recovery through CASSI enhancement, the proposed method achieves a 45.5% NRMSE reduction over the direct TwIST reconstruction, and outperforms an advanced yet high-complexity CS reconstruction algorithm (3DNSR [27]) with 1000 times acceleration.

The main contributions of this paper can be summarized as follows:

(1) A unified deep learning framework for hyperspectral image recovery from spectrally undersampled projections, with different network configurations explored.

(2) State-of-the-art results on RGB to hyperspectral conversion, which greatly facilitates hyperspectral image acquisition using ubiquitous RGB cameras.

(3) The first attempt on hyperspectral image recovery through CASSI enhancement, which serves as an efficient alternative to the high-complexity CS reconstruction.

2. Related Work

Hyperspectral image acquisition. Conventional spectrometers simply trade temporal resolution for spatial/spectral resolution [5]. For example, pushbroom or whiskbroom based methods capture the spectral information of a slit or a single point of the scene, and spatially scan the whole scene to obtain a full hyperspectral image [3, 19]. Filter wheel or tunable filter based methods integrate multiple color bandpass filters to select one band for each exposure, and multiple exposures are required to capture different spectral information of the scene [12, 20].

RGB to hyperspectral conversion is a low cost way of hyperspectral image acquisition, yet remains severely ill-posed since a large amount of information is lost during the process of spectral integration when RGB sensors capture the light. Solid progresses are only made recently in this direction, through either controlled illumination [13] or sparse coding [1]. Our proposed method can be regarded as a step forward of [1] by upgrading sparse coding to deep learning.

On the other hand, to overcome the limitation of conventional spectrometers and make it possible to capture dynamic scenes, snapshot spectral imagers with computational reconstruction have been developed in the last decade. The three representative techniques are computed tomographic imaging spectrometry (CTIS) [9], prism-mask spectral video imaging system (PMVIS) [7, 17], and CASSI [24, 25, 26]. These snapshot spectral imagers support hyperspectral video acquisition. However, it remains a trade-off between performance and speed of computational reconstruction. Our proposed method offers a new perspective for efficiently improving the reconstruction fidelity.

CNN-based super-resolution. Recovering hyperspectral images from spectrally undersampled projections can be viewed as "spectral super-resolution" and is quite similar to single-image super-resolution that enhances the spatial resolution of a 2D image, yet here the resolution is enhanced in the spectral dimension. Recently, CNN-based super-resolution (SRCNN) [10] has shown remarkable performance on a variety of natural images. Our proposed method is inspired by this seminal work and also the very deep CNNs for super-resolution (VDSR) [15]. However, spectral super-resolution differs from spatial super-resolution. After exploring different network configurations, we find that an adapted VDSR network works well for our tasks. Moreover, RGB to hyperspectral conversion favors a small receptive field in the spatial dimension, while CASSI enhancement favors a larger one. This difference, as also reflected by the residuals learned through the respective networks, indicates that CASSI measurements preserve more spectral details than RGB images, yet at the cost of heavier spatial degradation. In other words, CASSI enhancement is more like spatial super-resolution.

3. Unified Deep Learning Framework

Suppose G is an spectrally undersampled measurement obtained from a 3D hyperspectral image F through a certain projection. To recover F from G, we first upsample G to the desired spectral resolution through simple preprocessing. Denote this upsampled image as G', our goal is to recover from G' an image F' that is as similar as possible to the groundtruth hyperspectral image F.

Inspired by the well-known VDSR network for spatial super-resolution, we wish to learn a mapping $F' = f(G')$ that consists of three procedures: patch extraction, feature mapping, and reconstruction. All these procedures form a CNN. An overview of our deep learning framework (HSCNN in short hereafter) is depicted in Figure 1. The network takes the spectrally upsampled image as input and predicts missing image details, from which the final hyperspectral image is recovered. We will take RGB to hyperspectral conversion as an example to elaborate the network configurations and then extend to CASSI enhancement.

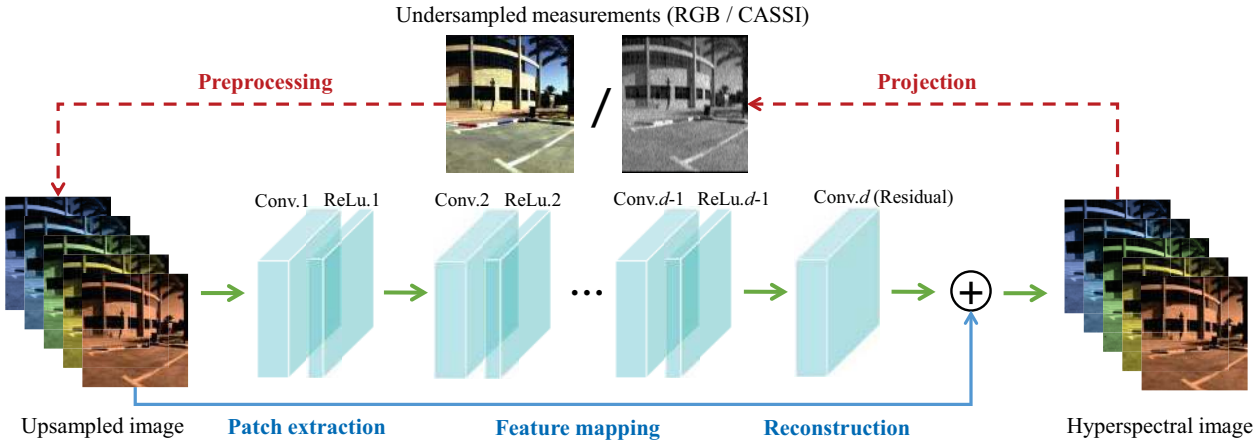


Figure 1. Framework of HSCNN. The network consists of three procedures: patch extraction, feature mapping, and reconstruction. It takes the spectrally upsampled image as input and predicts missing image details, from which the final hyperspectral image is recovered.

3.1. RGB to Hyperspectral Conversion

Denote $F(x, y, \lambda)$ the 3D hyperspectral image to be recovered, where $1 \leq x \leq W$, $1 \leq y \leq H$ index the spatial coordinates, and $1 \leq \lambda \leq \Omega$ indexes the spectral coordinate. The corresponding RGB image can be obtained as

$$G(x, y, k) = \sum_{\lambda=1}^{\Omega} \omega_k(\lambda) F(x, y, \lambda) \quad (1)$$

where $k = 1, 2, 3$ corresponds to blue, green, and red channels respectively, and $\omega_k(\lambda)$ denotes the spectral response of the RGB sensor corresponding to each channel. Here we use the CIE color matching functions for hyperspectral to RGB projection. To recover F from G, we first upsample G to G' through a simple spectral interpolation algorithm [21]. The image details missing in G' are then recovered through HSCNN, yielding the final hyperspectral image F'.

Network structure. Our network has d layers. The first layer is responsible for extracting patches from the input image. Since the hyperspectral image is 3D, the first layer has 64 filters of size $3 \times 3 \times \Omega$ and operates on 3×3 spatial region across Ω spectral bands. This layer produces 64 feature maps. Layers except the first and the last are of the same type as in VDSR: 64 filters of size $3 \times 3 \times 64$, each reproducing 64 feature maps from those obtained by the previous layer. The last layer consists of Ω filters of size $3 \times 3 \times 64$, which reconstructs the hyperspectral image with Ω spectral bands from 64 feature maps. The details of the HSCNN structure are listed in Table 1.

Given a training set $\{G_i, F_i\}_{i=1}^{N}$, the goal of our network training is to learn a mapping $F' = f(G')$ that minimizes the mean squared error $||F - F'||^2$ averaged over the whole training set. Also, residual learning, high learning rates, and adjustable gradient clipping are inherited from VDSR, which are demonstrated effective to optimize the network training. To investigate the influence of different network configurations on the reconstruction performance, we use a large hyperspectral dataset provided in [1], from which 100 images are used for training and 10 representative images for testing. These hyperspectral images have a uniform resolution of $1392 \times 1300 \times 31$.

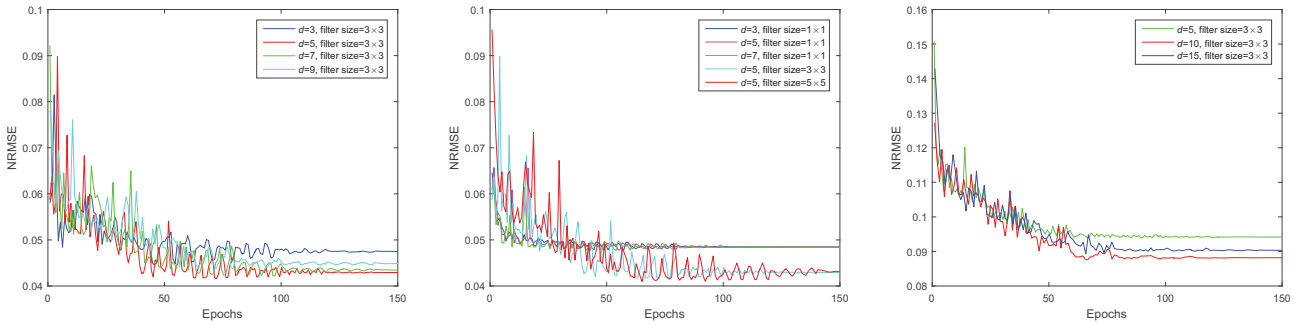
Table 1. Network structures in terms of filter size/number.

Procedure	VDSR	HSCNN
Patch extraction	3×3 / 64	$3 \times 3 \times \Omega$ / 64
Feature mapping	$3 \times 3 \times 64$ / 64	$3 \times 3 \times 64$ / 64
Reconstruction	$3 \times 3 \times 64$ / 1	$3 \times 3 \times 64$ / Ω

Network depth. The network depth is an important property of the deep learning framework. As demonstrated by VDSR for spatial super-resolution, a deeper network gives a better performance, partially owing to a larger receptive field (i.e., more neighboring pixels involved for reconstruction). When 3×3 filters are used, a d-layer network has a receptive field of size $(2d+1) \times (2d+1)$. It is observed that a 20-layer network with a 41×41 receptive field works quite well for spatial super-resolution. However, the above conclusion may not hold true for spectral super-resolution. Actually, the information contained in a 3D receptive field is much enriched compared with a 2D one. For example, a hyperspectral cube of size $7 \times 7 \times 31$ has nearly the same number of pixels as a 2D image patch of size 41×41. Therefore, the network depth fit for HSCNN is largely dependent on the target spectral resolution.

Figure 2(a) shows the performance curves in terms of the average NRMSE[1] on the test set under different network depths $d = 3, 5, 7, 9$. As can be seen, the highest fidelity reconstruction is obtained under $d = 5$, equivalent to a re-

[1]The NRMSE between F and F' is defined as $e(F, F') = \frac{1}{\Omega} \sum_{\lambda=1}^{\Omega} \sqrt{\frac{1}{WH} \sum_{x=1,y=1}^{W,H} |\frac{F'(x,y,\lambda) - F(x,y,\lambda)}{F(x,y,\lambda)}|^2}$



(a) (b) (c)

Figure 2. Exploration of optimal network configurations. (a) Performance curves for RGB to hyperspectral conversion under $d = 3, 5, 7, 9$ with spatial filter size of 3×3. (b) Performance curves for RGB to hyperspectral conversion under $d = 3, 5, 7$ with spatial filter size of 1×1, as well as those under $d = 5$ but with spatial filter size of 3×3 and 5×5. (c) Performance curves for CASSI enhancement under $d = 5, 10, 15$ with spatial filter size of 3×3.

ceptive field of $11 \times 11 \times 31$. It suggests that, different from spatial super-resolution, RGB to hyperspectral conversion favors a small receptive field in the spatial dimension. Besides the reason mentioned above that the spectral dimension compensates for the amount of contextual information, another underlying reason is that spatial structures are elegantly preserved in the RGB image. In other words, the information loss mainly occurs in the spectral dimension.

Filter size. As investigated in the seminal work of SRCNN, the network performance would improve as the network width (i.e., the number of filters) increases, yet at the cost of running time. This also applies to the filter size in the spatial dimension. Therefore, we use a default setting of 64 filters with spatial size of 3×3 for the balance of performance and speed. Interestingly, due to the additional spectral dimension, the spatial filter size can now be as small as 1×1. In this case, the network has a constant receptive field of $1 \times 1 \times 31$ regardless of the network depth, which is quite different from the VDSR-like network as explored above. Therefore, we also investigate the network performance with spatial filter size of 1×1.

Figure 2(b) shows the performance curves under different network depths $d = 3, 5, 7$ with spatial filter size of 1×1, and we find that the fidelity of reconstruction is almost independent of the network depth. For a more comprehensive evaluation, we also compare the performance curves under the same network depth $d = 5$ but with different spatial filter size of 3×3 and 5×5. It can be seen that, the 3×3 filter outperforms the 1×1 filter by a large margin and is competitive with the 5×5 filter. Therefore, we finally set the spatial filter size as 3×3 considering both performance and speed.

3.2. CASSI Enhancement

Relying on the CS theory, CASSI has made a significant breakthrough towards snapshot hyperspectral imaging. A typical CASSI system consists of an objective lens, a coded aperture, a relay lens, a dispersive prism, and a detector. The 3D hyperspectral image F is optically encoded onto a 2D detector, and the CASSI measurement can be written as

$$G(x, y) = \sum_{\lambda=1}^{\Omega} \omega(\lambda) S(x, y - \phi(\lambda)) F(x, y - \phi(\lambda), \lambda) \quad (2)$$

where $S(x, y)$ denotes the transmission function of the coded aperture, $\phi(\lambda)$ the wavelength-dependent dispersion function of the prism, and $\omega(\lambda)$ the spectral response function of the detector. (Please refer to [24] for a detailed formulation of the CASSI measurement.) To recover a full hyperspectral image F from G, different CS reconstruction algorithms can be employed. However, advanced algorithms such as dictionary learning are of high complexity. Here we propose an alternative solution. That is, we first use a simple TwIST algorithm [4] to upsample G to G', and the image details missing in G' are then enhanced through HSCNN, yielding the final hyperspectral image F'.

For CASSI enhancement, we retain the network configurations as explored in RGB to hyperspectral conversion, except the network depth d that is sensitive to the property of measurements. Figure 2(c) shows the performance curves under different network depths $d = 5, 10, 15$. As can be seen, the highest fidelity reconstruction is obtained under $d = 10$, equivalent to a receptive field of $21 \times 21 \times 31$. It suggests that, compared with RGB to hyperspectral conversion, CASSI enhancement favors a larger receptive field in the spatial dimension. The underlying reason is that, owing to the CS-based spectral encoding, the CASSI measurement preserves more spectral details than the RGB one, yet at the cost of heavier spatial degradation. Therefore, the network for CASSI enhancement prefers a larger spatial dependency, which is more like spatial super-resolution. This will be evidenced by the experimental results below.

4. Experimental Results

We conduct experiments on a large hyperspectral dataset provided in [1], which contains a number of subsets with diverse image content. The original hyperspectral images have a uniform resolution of $1392 \times 1300 \times 31$, spanning the spectral range of 400nm to 700nm with 10nm intervals. Images used for training and testing are specified below for a fair comparison with the state-of-the-arts. Training parameters are set following the VDSR network. Batch size is set to 128. Momentum and weight decay parameters are set to 0.9 and 0.0001, respectively. All networks are trained over 150 epochs. Learning rate is initially set to 0.1 and then decreased by a factor of 10 every 20 epochs. Training takes about 20 hours on a Tesla K80 GPU.

4.1. RGB to Hyperspectral Conversion

For RGB to hyperspectral conversion, HSCNN is compared with the baseline interpolation method [21] and the state-of-the-art sparse coding method [1]. The sparse coding method reports the NRMSE results on a "complete set" containing 100 images and a few subsets with 59 domain-specific images (belonging to the 100 images). These results are obtained by 1) selecting one image in a set for testing and the rest images in the set for training the dictionary, and 2) repeating the above step until each image in the set has been selected for testing and the average NRMSE is calculated. In this way, the sparse coding method can achieve better results in the domain-specific subsets compared with the complete set. Contrastively, HSCNN is performed in a more generalizable way. We use a total of 200 images including the complete set, and first train a CNN model with the 141 images excluding the domain-specific subsets. This model is tested on the 59 domain-specific images. We then train another model with the 159 images excluding the non-domain-specific images in the complete set, and the obtained model is tested on these 41 images. In this way, images for training and testing are strictly non-overlapped, and HSCNN gets rid of the domain-specific restriction as imposed in sparse coding.

Table 2 gives the numerical results of different methods on the specified test sets. As can be seen, HSCNN achieve a 48.7% reduction in NRMSE compared with sparse coding on the complete set. The results on domain-specific subsets are also consistently improved, except the indoor subset containing only 3 images. It demonstrates that, in most cases, the model learned by HSCNN can be well generalized. Figure 3 shows the visual results of five selected bands from a test hyperspectral image, together with the residuals learned by HSCNN. As can be seen, the interpolation results well maintain the spatial structures but suffer from large spectral deviation from the groundtruth, yet this deviation is compensated in our reconstructed results. Figure 4 further shows the spectral signatures of two selected spatial points from the test image. While interpolation can only recover a smooth signature that is heavily deviated from the groundtruth, HSCNN gives high fidelity reconstruction.

4.2. CASSI Enhancement

For CASSI enhancement, HSCNN is compared with the baseline TwIST algorithm [4] and the state-of-the-art 3DNSR algorithm [27]. TwIST is a simple CS reconstruction algorithm using the total variation prior, while 3DNSR is an advanced yet high-complexity algorithm using self-learned dictionaries. We generate the CASSI measurement and set the parameters of TwIST and 3DNSR following [27]. Table 3 reports the numerical results on the 59 domain-specific images using the other 141 images for training. (Note we only generate the 3DNSR results in the central 256×256 spatial region of each image, and NRMSE is calculated in this region for all methods). HSCNN achieves a 45.5% NRMSE reduction over TwIST and outperforms 3DNSR by a large margin. The running time per image on an Intel Core i7-6700K CPU is included in Table 3. Since 3DNSR also uses the TwIST results for initialization, we exclude the TwIST reconstruction time for both 3DNSR and HSCNN. HSCNN is nearly 1000 times faster than 3DNSR, as the former is one-time feed forward while the latter needs hundreds of iterations. Although it requires additional time for training, HSCNN could serve as an efficient alternative to the high-complexity CS reconstruction.

Figure 5 shows the visual results of five selected bands from a test hyperspectral image, together with the residuals learned by HSCNN. Different from RGB to hyperspectral conversion, the TwIST results well maintain the spectral consistency with the groundtruth but suffer from heavy spatial corruption. This can be clearly observed from the very different residuals learned by the respective networks. It reflects that information losses in these two kinds of projections (RGB and CASSI) are distinct, and also explains why CASSI enhancement favors a larger receptive field in the spatial dimension. Figure 6 further shows the spectral signatures of two selected spatial points from the test image. While all methods faithfully recover the trend of the signature, HSCNN gives the highest fidelity reconstruction.

Figure 7 shows the close-up results of one selected band from the test image, from which the reconstruction of spatial structures by different methods can be better observed. The TwIST result suffers from severe spatial artifacts, e.g., textures on the wall and the ground are blurred. The 3DNSR result is improved to a certain extent yet still has some degradation. Contrastively, the HSCNN result best preserves the spatial structures.

5. Conclusion

This paper presents a unified deep learning framework for hyperspectral image recovery from spectrally undersam-

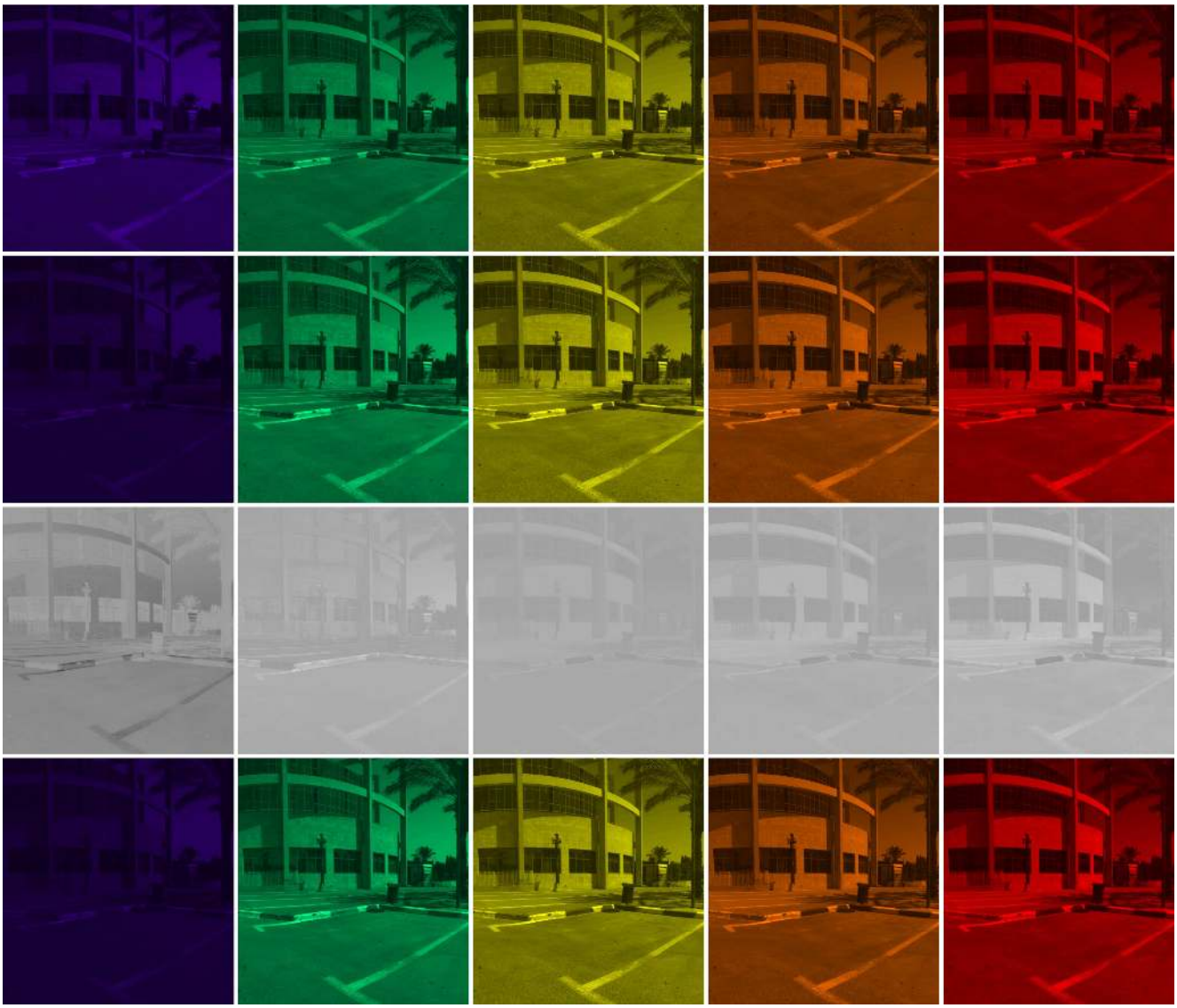
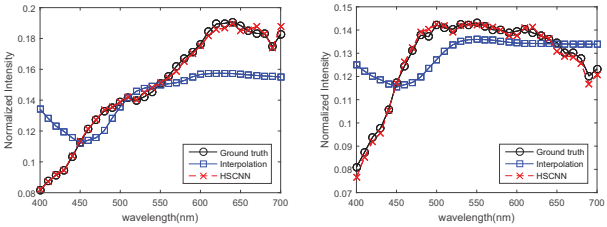


Figure 3. Visual results of five selected bands for RGB to hyperspectral conversion. From top to bottom: interpolation, HSCNN, residuals learned by HSCNN (properly scaled for better visualization), and groundtruth. From left to right: 430nm, 500nm, 570nm, 600nm, 630nm.

Table 2. NRMSE results of RGB to hyperspectral conversion.

Data Set	Interpolation	Sparse coding	HSCNN
Complete set	0.2085	0.0756	**0.0388**
Park subset	0.1843	0.0590	**0.0371**
Indoor subset	0.1632	**0.0510**	0.0638
Urban subset	0.2150	0.0620	**0.0388**
Rural subset	0.2551	0.0350	**0.0331**
Plant-life subset	0.2074	0.0470	**0.0445**
Subset average	0.2105	0.0572	**0.0398**



(a) (b)

Figure 4. Signature of two selected spatial points in Figure 3.

pled projections. Two representative tasks, RGB to hyperspectral conversion and CASSI enhancement, are investigated and encouraging results are reported. We believe the proposed method will greatly facilitate hyperspectral image acquisition, either by using ubiquitous RGB cameras, or by improving computational reconstruction for snapshot hyperspectral imagers.

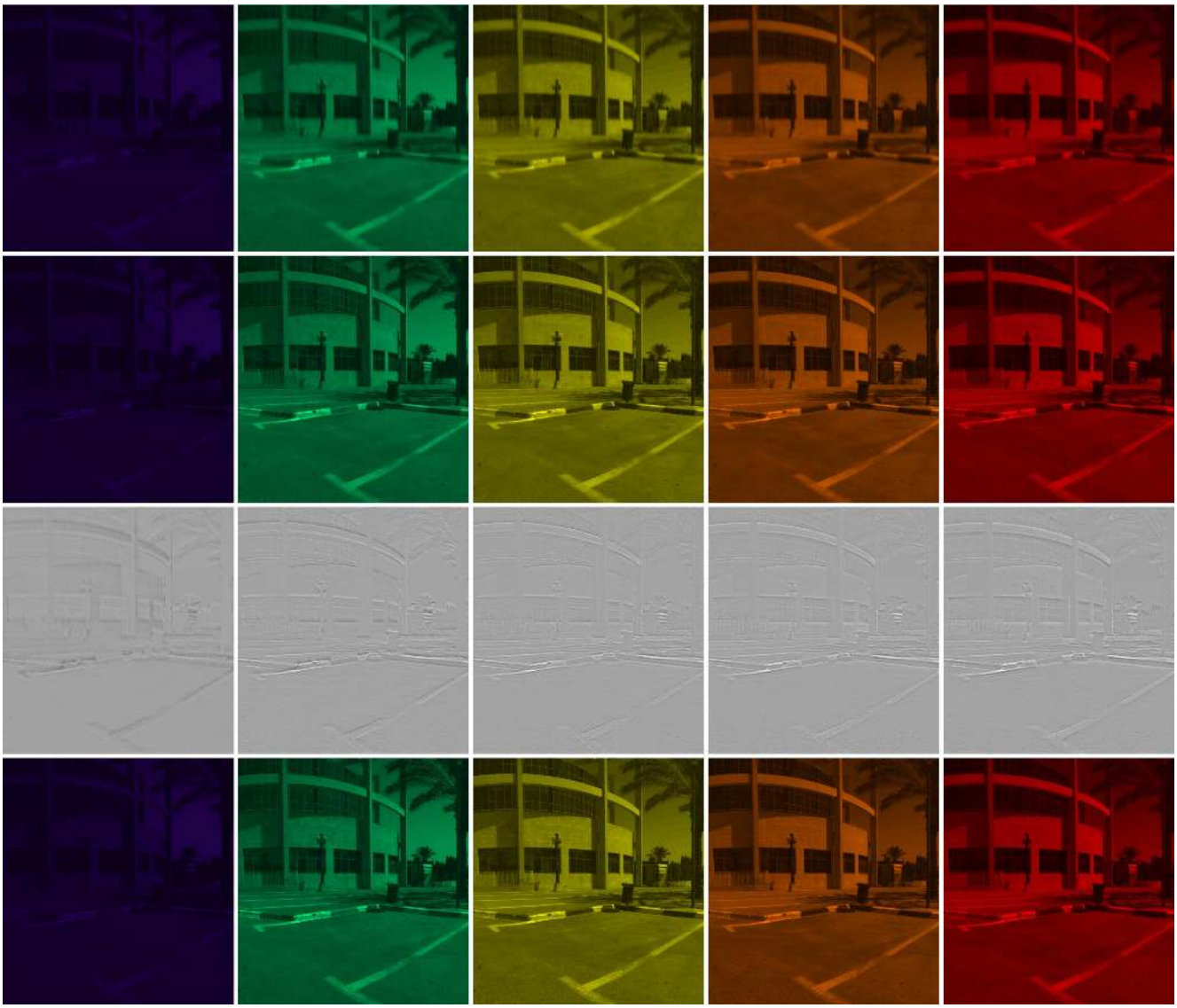
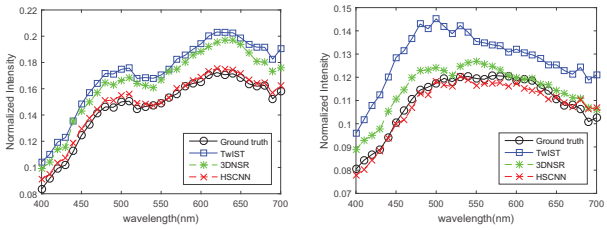


Figure 5. Visual results of five selected bands for CASSI enhancement. From top to bottom: TwIST, HSCNN, residuals learned by HSCNN (properly scaled for better visualization), and groundtruth. From left to right: 430nm, 500nm, 570nm, 600nm, 630nm.

Table 3. NRMSE results of CASSI enhancement.

Data Set	TwIST	3DNSR	HSCNN
Park subset	0.1601	0.1421	**0.0873**
Indoor subset	0.0941	0.0734	**0.0580**
Urban subset	0.1220	0.1061	**0.0688**
Rural subset	0.1381	0.1257	**0.0733**
Plant-life subset	0.1688	0.1468	**0.0994**
Subset average	0.1317	0.1150	**0.0740**
Time average (s)	*142*	*2896*	***2.56***



(a) (b)

Figure 6. Signature of two selected spatial points in Figure 5.

Acknowledgments

We acknowledge funding from the NSFC Grant 61671419, the CAS Pioneer Hundred Talents Program, and the Fundamental Research Funds for the Central Universities under Grant WK3490000001.

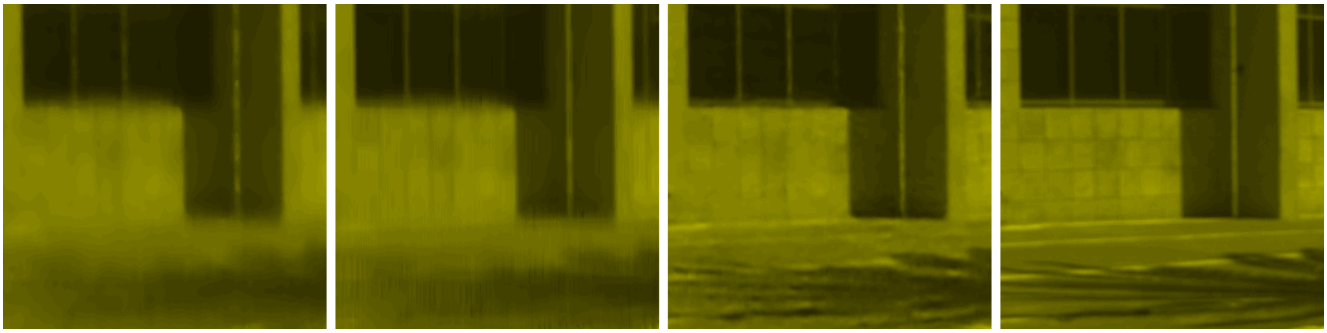


Figure 7. Close-up results (256 × 256 central region) of one selected band (570nm) for CASSI enhancement. From left to right: TwIST, 3DNSR, HSCNN, and groundtruth.

References

[1] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *ECCV*, 2016. 1, 2, 3, 5

[2] G. Arce, D. Brady, L. Carin, H. Arguello, and D. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Process. Mag.*, 31(1):105–115, 2014. 1

[3] R. W. Basedow, D. C. Carmer, and M. E. Anderson. Hydice system: Implementation and performance. In *Proc. SPIE*, 1995. 2

[4] J. Bioucas-Dias and M. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.*, 16(12):2992 – 3004, 2007. 2, 4, 5

[5] D. Brady. *Optical Imaging and Spectroscopy*. John Wiley and Sons Inc., 2008. 1, 2

[6] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52(2):489–509, 2006. 1

[7] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin. A prism-mask system for multispectral video acquisition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(12):2423 – 35, 2011. 1, 2

[8] T. F. Chen, G. V. Baranoski, B. W. Kimmel, and E. Miranda. Hyperspectral modeling of skin appearance. *ACM Trans. Graph.*, 34(3):31:1–31:14, 2015. 1

[9] M. Descour and E. Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Appl. Opt.*, 34(22):4817–4826, 1995. 1, 2

[10] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 38(2):295–307, 2016. 2

[11] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006. 1

[12] N. Gat. Imaging spectroscopy using tunable filters: a review. In *Proc. SPIE*, 2000. 2

[13] M. Goel, E. Whitmire, A. Mariakakis, T. S. Saponas, N. Joshi, D. Morris, B. Guenter, M. Gavriliu, G. Borriello, and S. N. Patel. Hypercam: hyperspectral imaging for ubiquitous computing applications. In *UbiComp*, 2015. 1, 2

[14] E. K. Hege, D. O'Connell, W. Johnson, S. Basty, and E. L. Dereniak. Hyperspectral imaging for astronomy and space surviellance. In *Proc. SPIE*, 2004. 1

[15] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 2

[16] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Trans. Graph.*, 31(4):38:1–38:11, 2012. 1

[17] C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin. Acquisition of high spatial and spectral resolution video with a hybrid camera system. *Int. J. Comput. Vision*, 110(2):141–155, 2014. 2

[18] Z. Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(12):1552–1560, 2003. 1

[19] W. M. Porter and H. T. Enmark. A system overview of the airborne visible/infrared imaging spectrometer (aviris). In *Proc. SPIE*, 1987. 2

[20] Y. Schechner and S. Nayar. Generalized mosaicing: wide field of view multispectral imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(10):1334–1348, 2002. 2

[21] B. Smits. An rgb-to-spectrum conversion for reflectances. *J. Graph. Tools*, 4(4):11–22, 1999. 3, 5

[22] J. Solomon and B. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–1152, 1985. 1

[23] H. Van Nguyen, A. Banerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. In *CVPRW*, 2010. 1

[24] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44 – B51, 2008. 1, 2, 4

[25] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 54(4):848–858, 2015. 2

[26] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *CVPR*, 2015. 2

[27] L. Wang, Z. Xiong, G. Shi, F. Wu, and W. Zeng. Adaptive nonlocal sparse representation for dual-camera compressive hyperspectral imaging. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016. 2, 5