DEEPCASD: AN END-TO-END APPROACH FOR MULTI-SPECTRAL IMAGE SUPER-RESOLUTION

Bihan Wen[1], *Ulugbek S. Kamilov*[2], *Dehong Liu*[2], *Hassan Mansour*[2], *and Petros T. Boufounos*[2]

[1] Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, IL, USA
[2] Mitsubishi Electric Research Laboratories, Cambridge, MA, USA

ABSTRACT

Multi-spectral (MS) image super-resolution aims to reconstruct super-resolved multi-channel images from their low-resolution images by regularizing the image to be reconstructed. Recently data-driven regularization techniques based on sparse modeling and deep learning have achieved substantial improvements in single image reconstruction problems. Inspired by these data-driven methods, we develop a novel coupled analysis and synthesis dictionary (CASD) model for MS image super-resolution, by exploiting a regularizer that operates within, as well as across, multiple spectral channels using convolutional dictionaries. To learn the CASD model parameters, we propose a deep dictionary learning framework, named DeepCASD, by unfolding and training an end-to-end CASD based reconstruction network over an image data set. Experimental results show that the DeepCASD framework exhibits improved performance on multi-spectral image super-resolution compared to state-of-the-art learning based super-resolution algorithms.

Index Terms— Super-resolution, convolutional dictionary, deep learning, multi-spectral imaging, algorithm unfolding.

1. INTRODUCTION

Multi-spectral (MS) imaging systems measure the response from an area of interest over a wide range of frequency bands including visible optical RGB, infra-red, and short-wave infra-red bands [1]. These multi-band spectra provide rich information for detecting and distinguishing materials, especially for those materials with visually similar colors. Furthermore, the higher atmospheric transmission property of infra-red bands versus optical bands makes the MS imaging more beneficial in hazy or cloudy weather conditions when optical imaging systems do not work well. In recent decades, multi-spectral or even hyper-spectral imaging techniques have been widely used in remote sensing applications ranging from astronomy [2], agriculture [3], and geoscience [4].

However, MS imaging systems encounter a trade-off between spatial and spectral resolution in the hardware implementation [5, 6]. While high spatial resolution of MS images is desirable for many applications, resolution is typically constrained by the limitations on the size, weight, and power of the sensors mounted on-board of airplanes or spacecrafts [4]. To overcome these limitations, image super-resolution (SR) methods have been developed that mitigate hardware shortfall using computational imaging techniques [6–9].

Image SR is generally an ill-posed inverse problem. To solve this problem, a regularizer on the underlying high-resolution image is required, and consequently plays an important role in the final reconstruction. Pan-sharpening methods, for example, are classical image SR methods where the spatial information of the known panchromatic images is fused with the spectral information of color or MS images, to generate high spatial resolution and high spectral resolution images [10]. This fusion process can be improved by low-rank modeling [11, 12], or sparse regularization via dictionary learning on coupled low-resolution and high-resolution image patches [13, 14]. Since the spectral band of panchromatic images does not overlap with infrared bands, pan-sharpening methods inevitably introduce spectral distortion for MS image SR.

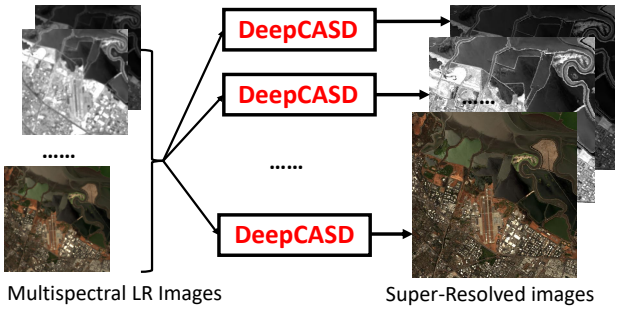

Fig. 1. DeepCASD for Multi-spectral Image Super-Resolution.

Recent advances in data-driven methods have led to substantial improvements in many imaging tasks, including image SR [7–9, 15–17]. Meanwhile, convolutional dictionaries also have shown to be more effective than spatial patch dictionaries by enforcing the neighboring patch dependency [18–20]. However, to the best of our knowledge, such methods have not been exploited in multi-spectral image fusion.

In this paper, we introduce a data-driven Coupled Analysis and Synthesis Dictionary (CASD) model that relates the multi-channel low resolution (LR) measurements with the high resolution (HR) images through shared feature maps and coupled convolutional dictionaries. To learn the CASD model, we propose a deep-learning framework, named DeepCASD, which leverages training data more effectively with an end-to-end approach. Fig. 1 shows the MS image SR process using DeepCASD. Compared to the existing solutions, our proposed method exhibits several advantages. First, it allows couplings within and across multiple channels with a shared feature map. Second, it provides an end-to-end learning framework for image SR. Third, it provides greater flexibility in terms of the total number of spectral channels and image resolution compared to traditional pan-sharpening and image SR methods. We evaluate the proposed DeepCASD framework on synthesized MS images using AVIRIS [1] hyper-spectral images, and compare the results with competing methods quantitatively and perceptually.

This work was completed while B. Wen was an intern with Mitsubishi Electric Research Laboratories.

2. BACKGROUND AND RELATED WORK

In image super-resolution, an HR image $\mathbf{x} \in \mathbb{R}^p$ is typically reconstructed by solving the following regularized imaging problem

$$\text{(P0)} \quad \hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \mathcal{R}(\mathbf{x}),$$

where $\mathbf{A}$ is the sensing matrix. Typically $\mathbf{A}$ involves a system-dependent low-pass filtering process and a down-sampling process when $\mathbf{y}$ is an LR image of size smaller than that of $\mathbf{x}$. Alternatively, if we up-scale $\mathbf{y}$ to the same size of $\mathbf{x}$ using standard techniques like bicubic interpolation, the sensing matrix $\mathbf{A}$ can be approximated as the $p \times p$ identity matrix [7–9]. In the remainder of this paper, we take the latter option by assuming that bicubic interpolation has already been applied to the low resolution images such that $\mathbf{y} \in \mathbb{R}^p$.

As SR is ill-posed, an effective image regularizer $\mathcal{R}(\mathbf{x})$ is the key to successful reconstruction. Previous SR methods imposed various forms of $\mathcal{R}(\mathbf{x})$, exploiting image properties such as total variation [21] or sparsity [22]. With the introduction of dictionary learning [13, 18], data-driven methods have attracted great attention in image SR [7–9, 14]. Furthermore, recent advances in unfolded dictionary learning [20, 23, 24] demonstrated promising performance in single image super-resolution. In particular, the Shrinkage Field (SF) network [20] provides state-of-the-art imaging performance and efficient training on a single channel image $\mathbf{x}$ by solving

$$\text{(P1)} \quad \min_{\mathbf{x},\mathbf{u}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \frac{\lambda}{2} \|\mathbf{u} - \mathbf{W}\,\mathbf{x}\|_2^2 + \Phi_\theta(\mathbf{u}),$$

where $\mathbf{W} = \left[\mathbf{W}_1^T | ... | \mathbf{W}_M^T\right]^T$ is a convolutional analysis dictionary with M convolutional filters, such that

$$\mathbf{W}\mathbf{x} \triangleq \left[(\mathbf{w}_1 * \mathbf{x})^T | ... | (\mathbf{w}_M * \mathbf{x})^T\right]^T,$$

where $*$ denotes the convolution operator, and each $\mathbf{w}_m, m \in \{1, \ldots, M\}$, is a filter in $\mathbb{R}^q$. Here $\mathbf{u} \in \mathbb{R}^{Mp}$ is the feature map, with $\Phi_\theta(\mathbf{u})$ being its regularizer. A parametric function $\Phi_\theta(\cdot)$ is used, which leads to a shrinkage function composed of a linear combination of Gaussian radial basis functions (RBF) [20, 23]. The shrinkage parameter set θ and the convolutional dictionary $\mathbf{W}$ are learned using a training dataset by maximizing the reconstruction Peak Signal-to-Noise Ratio (PSNR).

3. PROPOSED METHOD

3.1. Proposed CASD model

For MS image SR, we introduce an end-to-end fusion network, named DeepCASD, with an overview chart shown in Fig. 1. The structure of each DeepCASD block, as illustrated in Fig. 2, is composed of both single-channel and multi-channel SR modules, to fully operate within and across different image channels. We adopt the SF network [20] as our single-channel module due to its demonstrated performance and efficiency. Given L spectral channels, the multi-channel CASD model assumes that each HR image $\mathbf{x}_l$, for $l \in \{1, .., L\}$, and all of the multi-channel LR measurements $\mathbf{y} = \left[\mathbf{y}_1^T | ... | \mathbf{y}_L^T\right]^T \in \mathbb{R}^{Lp}$ are approximately sparse under convolutional synthesis and analysis dictionaries, respectively, with a shared coefficient map $\mathbf{u} \in \mathbb{R}^{Mp}$.

We super-resolve the HR image $\mathbf{x}_l$ in each channel, by solving the following CASD imaging problem

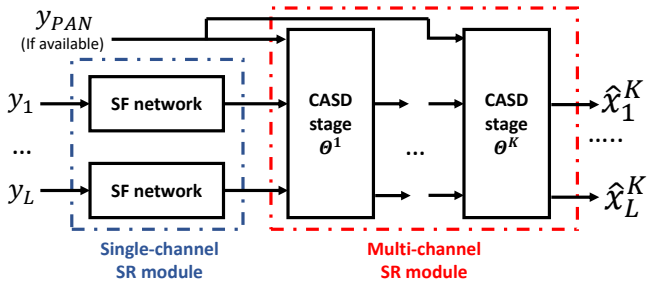

Fig. 2. The outline of the DeepCASD for MS image SR.

$$\text{(P2)} \quad \min_{\mathbf{x}_l,\mathbf{u}_l} \frac{1}{2} \|\mathbf{y}_l - \mathbf{x}_l\|_2^2 + \frac{\lambda_2}{2} \|\mathbf{B}_l\mathbf{y} - \mathbf{u}_l\|_2^2 + \frac{\lambda_1}{2} \|\mathbf{x}_l - \mathbf{D}_l\mathbf{u}_l\|_2^2 + \Phi_{\theta_l}(\mathbf{u}_l)$$

where $\mathbf{D}_l = \left[\mathbf{D}_{l,1} | ... | \mathbf{D}_{l,M}\right] \in \mathbb{R}^{p \times Mp}$ is the convolutional synthesis dictionary [18, 19, 25] for $\mathbf{x}_l$, while the convolutional analysis dictionary [20, 24, 26] $\mathbf{B}_l$ for $\mathbf{y}$ in (P2) is defined as

$$\mathbf{B}_l = \begin{bmatrix} \mathbf{W}_{1,1}^l & \cdots & \mathbf{W}_{1,L}^l \\ \vdots & \ddots & \vdots \\ \mathbf{W}_{M,1}^l & \cdots & \mathbf{W}_{M,L}^l \end{bmatrix} \in \mathbb{R}^{Mp \times Lp}. \tag{1}$$

In (P2), the terms $\|\mathbf{B}_l\mathbf{y} - \mathbf{u}_l\|_2^2$ and $\|\mathbf{x}_l - \mathbf{D}_l\mathbf{u}_l\|_2^2$ denote the modeling errors for $\mathbf{x}_l$ under the analysis dictionary, and the synthesis dictionary, respectively. Comparing to the single analysis model used in (P1), the CASD model in (P2) further exploits the correlation between the multi-channel LR measurements and the HR image using analysis dictionaries, rather than only the HR image structure.

To learn the CASD, one can directly solve (P2) using alternating minimization, which has been widely used in previous work on coupled dictionary learning [7, 9]. Recent works on deep learning by unfolding the synthesis sparse coding problem [27, 28] demonstrated superior performance in many imaging applications, comparing to conventional iterative methods. Inspired by this work, instead of directly optimizing the loss function in (P2), we propose an end-to-end learning framework, namely DeepCASD, by unfolding the CASD learning for image fusion.

3.2. End-to-end DeepCASD learning

The multi-channel SR module in the proposed DeepCASD contains K CASD stages. In each stage, the trainable parameter set is $\{\mathbf{B}_l, \theta_l, \mathbf{D}_l\}$, which is used to generate the feature map $\mathbf{u}_l$, and thus super-resolve each $\mathbf{y}_l$. Given $\mathbf{u}_l$ and the dictionary $\mathbf{D}_l$, the solution $\hat{\mathbf{x}}_l$ to (P2) is given by

$$\begin{aligned} \hat{\mathbf{x}}_l &= \lambda_1 / (1 + \lambda_1)\, \mathbf{D}_l\mathbf{u}_l + 1 / (1 + \lambda_1)\, \mathbf{y}_l \\ &= \tilde{\mathbf{D}}_l\, \mathbf{u}_l + \lambda' \mathbf{y}_l \end{aligned} \tag{2}$$

where $\lambda' = 1/(1+\lambda_1)$, and $\lambda_1/(1+\lambda_1)$ is absorbed into $\tilde{\mathbf{D}}_l$ during the learning. To obtain the feature map $\mathbf{u}_l$, directly solving (P2) involves gradient calculation using the trainable $\mathbf{D}_l$ and the output $\hat{\mathbf{x}}_l$ in the end-to-end training, which leads to the recurrent neural network structure. We restrict ourselves to construct feed-forward DeepCASD network for efficient implementation. Thus each $\mathbf{u}_l$ is estimated by solving the following analysis (i.e. transform) model

sparse coding problem

$$\begin{aligned}\hat{\mathbf{u}}_l &= \arg\min_{\mathbf{u}_l} \|\mathbf{u}_l - \mathbf{B}_l\mathbf{y}\|_2^2 + \Phi_{\theta_l}(\mathbf{u}_l) \\ &= \eta_{\theta_l}(\mathbf{B}_l\mathbf{y}) \quad \forall l = 1, ..., L \end{aligned} \tag{3}$$

where $\eta_{\theta_l}(\cdot)$ denotes the corresponding shrinkage function of the feature regularizer $\Phi_{\theta_l}(\cdot)$. For example, when $\Phi_\theta(\mathbf{u}) := \theta \|\mathbf{u}\|_0$ with the ℓ_0 "norm" that counts the number of non-zero elements of $\mathbf{u}$, the corresponding $\eta_\theta(\cdot)$ becomes the hard-thresholding function of $\mathbf{u}$, with threshold value θ. Here we use the trainable Gaussian RBF to be the shrinkage function $\eta_{\theta_l}(\cdot)$, which shows good generalization of non-linear shrinkage functions in the SF network [20].

To analyze the cascaded structure in the proposed DeepCASD, we denote the l-channel super-resolved image as $\hat{\mathbf{x}}_l^k$ at the k-th stage, $k = 1, ..., K$. The trainable set at the k-th stage for the $\hat{\mathbf{x}}_l^k$ reconstruction is denoted as $\{\mathbf{B}_l^k, \theta_l^k, \mathbf{D}_l^k\}$. Fig. 2 illustrates how the K cascaded CASD stages construct the multi-channel SR module: the k-th stage multi-channel output $\hat{\mathbf{x}}^k \in \mathbb{R}^{Lp}$ is passed to the next stage as its LR input. It can be expressed as a function of the dictionaries $\mathbf{D}_l^k$ and $\mathbf{B}_l^k$, the shrinkage function parameter set θ_l^k, the previous stage output $\hat{\mathbf{x}}^{k-1} = \{\hat{\mathbf{x}}_l{}^{k-1}\}_{l=1}^L$, and l-th channel LR image $\mathbf{y}_l$ recursively as follow

$$\begin{aligned}\hat{\mathbf{x}}_l^k &= f(\mathbf{D}_l^k, \mathbf{B}_l^k, \theta_l^k, \hat{\mathbf{x}}^{k-1}, \mathbf{y}_l) \\ &= \mathbf{D}_l^k \, \eta_{\theta_l^k}(\mathbf{B}_l^k \hat{\mathbf{x}}^{k-1}) + \lambda' \mathbf{y}_l \qquad \forall\, l, k \end{aligned} \tag{4}$$

where $\hat{\mathbf{x}}^0$ denotes the input at the first stage. Figure 3 illustrates the layer-level structure of the k-th stage CASD network for $\hat{\mathbf{x}}_l^k$ reconstruction. Note that $\mathbf{W}_{m,:}^k$ denotes the m-th row of $\mathbf{B}_l^k$, which fuses the multi-channel input $\hat{\mathbf{x}}^{k-1}$ before applying the shrinkage, i.e., $\mathbf{v}_m^k = \mathbf{W}_{m,:}^k \hat{\mathbf{x}}^{k-1} \in \mathbb{R}^p$.

We train the DeepCASD over a training set which contains N pairs of multi-channel HR images $\mathbf{X} = [\mathbf{x}_{(1)}| ... |\mathbf{x}_{(N)}]$ and their interpolated LR measurements $\mathbf{Y} = [\mathbf{y}_{(1)}| ... |\mathbf{y}_{(N)}]$. The negative average reconstruction PSNR over all images and channels is employed as the cost function $\mathbb{L}$ at the final output:

$$\begin{aligned}\mathbb{L}(\mathbf{X}, \hat{\mathbf{X}}^K) &\triangleq -PSNR \\ &= -\frac{20}{NL} \sum_{i=1}^{N} \sum_{l=1}^{L} \log_{10} \frac{B\sqrt{P}}{\|\mathbf{x}_{(i),l} - \hat{\mathbf{x}}_{(i),l}^K\|_2}. \end{aligned} \tag{5}$$

Here B is the maximum image pixel value (e.g., $B = 255$ for 8-bit image), and $\hat{\mathbf{X}}^K$ denotes the super-resolved multi-channel images using DeepCASD. Let the set of all trainable parameters in the K-stage DeepCASD be $\Theta = \{\Theta^k\}_{k=1}^K \triangleq \left\{\{\mathbf{B}_l^k, \theta_l^k, \mathbf{D}_l^k\}_{l=1}^L\right\}_{k=1}^K$. The *joint* DeepCASD training problem is formulated as

$$\text{(P3)} \quad \hat{\Theta} = \arg\min_{\Theta} \mathbb{L}(\mathbf{X}, \hat{\mathbf{X}}^K(\Theta, \mathbf{Y}))$$

Problem (P3) can be solved using error back-propagation. Alternatively, as each DeepCASD stage itself is a stand-alone image fusion network [20], each Θ_k can be trained separately by solving the following *stage-wise* DeepCASD training problem

$$\text{(P4)} \quad \hat{\Theta}_{\text{stg}}^k = \arg\min_{\Theta^k} \mathbb{L}(\mathbf{X}, \hat{\mathbf{X}}^k(\Theta^k, \hat{\mathbf{X}}^{k-1}, \mathbf{Y})) \;\; \forall k.$$

In practice, as (P3) is highly non-convex, it is more efficient to use the stage-wise $\{\Theta_{\text{stg}}^k\}_{k=1}^K$ learned using (P4) as the initialization in the joint training for (P3). Once the DeepCASD network training is completed, the multi-channel SR is conducted by applying (4) recursively with the trained $\hat{\Theta}$.

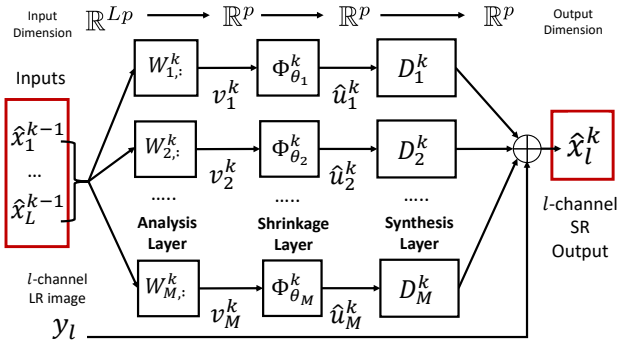

Fig. 3. The layer-level structure of the k-th stage CASD for super-resolving the l-th channel $\hat{\mathbf{x}}_l^k$, and the input / output data dimensions.

Table 1. PSNR values (in dB) for MS image $\times 2$ SR, averaged over 16 channels, using bicubic interpolation, dictionary learning (DL) [9], Shrinkage Field (SF) [20], and the proposed DeepCASD method. The best PSNR value in each row is marked in bold.

MS Images	Bicubic	DL [9]	SF [20]	DeepCASD
Moffett	32.27	33.81	34.25	**34.57**
Cambria Fire	35.49	36.55	37.09	**37.22**
Cuprite	32.36	33.60	34.49	**34.68**
Los Angeles	27.97	29.62	30.34	**30.46**
Average	32.02	33.41	34.04	**34.23**

4. NUMERICAL EXPERIMENTS

4.1. Experimental Setup

We evaluate and compare the performance of the proposed DeepCASD on the SR problem over remote sensing MS images. We synthesize MS images of 17 discrete channels, including panchromatic, RGB, infra-red, and short-wave infra-red channels, using AVIRIS hyper-spectral image data sets [1]. Each high-resolution MS channel is generated as a weighted sum of multiple hyper-spectral channels covering adjacent frequency bands. The corresponding low-resolution MS channels are then generated by down-sampling the high-resolution MS image through a low-pass filter. The parameters of the DeepCASD network are first trained using a set of MS images. The training set contains 138 pairs of high-resolution MS images and their corresponding low-resolution measurements, across 16 channels. Each HR image in a single channel is of size 256×256. The LR images are first up-scaled to the HR image size by bicubic interpolation [7–9]. As the HR panchromatic image is typically available in remote sensing applications, we pass it through a skip link directly to each multi-channel SR stage (i.e., there are $L = 17$ input channels and $L = 16$ output channels in each CASD stage) in training and testing (see Fig. 2).

We use 3 single-channel SR stages, followed by 1 multi-channel SR stage in the DeepCASD network, and perform a for $2\times$-SR experiment. We set the number of feature channels $M = 8$, and the size of convolutional filters to be 3×3. We observed improved SR performance using DeepCASD with more SR stages. However, more training data are required for training deeper networks, otherwise DeepCASD may suffer from overfitting which causes degraded SR quality. For the training process, we observe that joint training

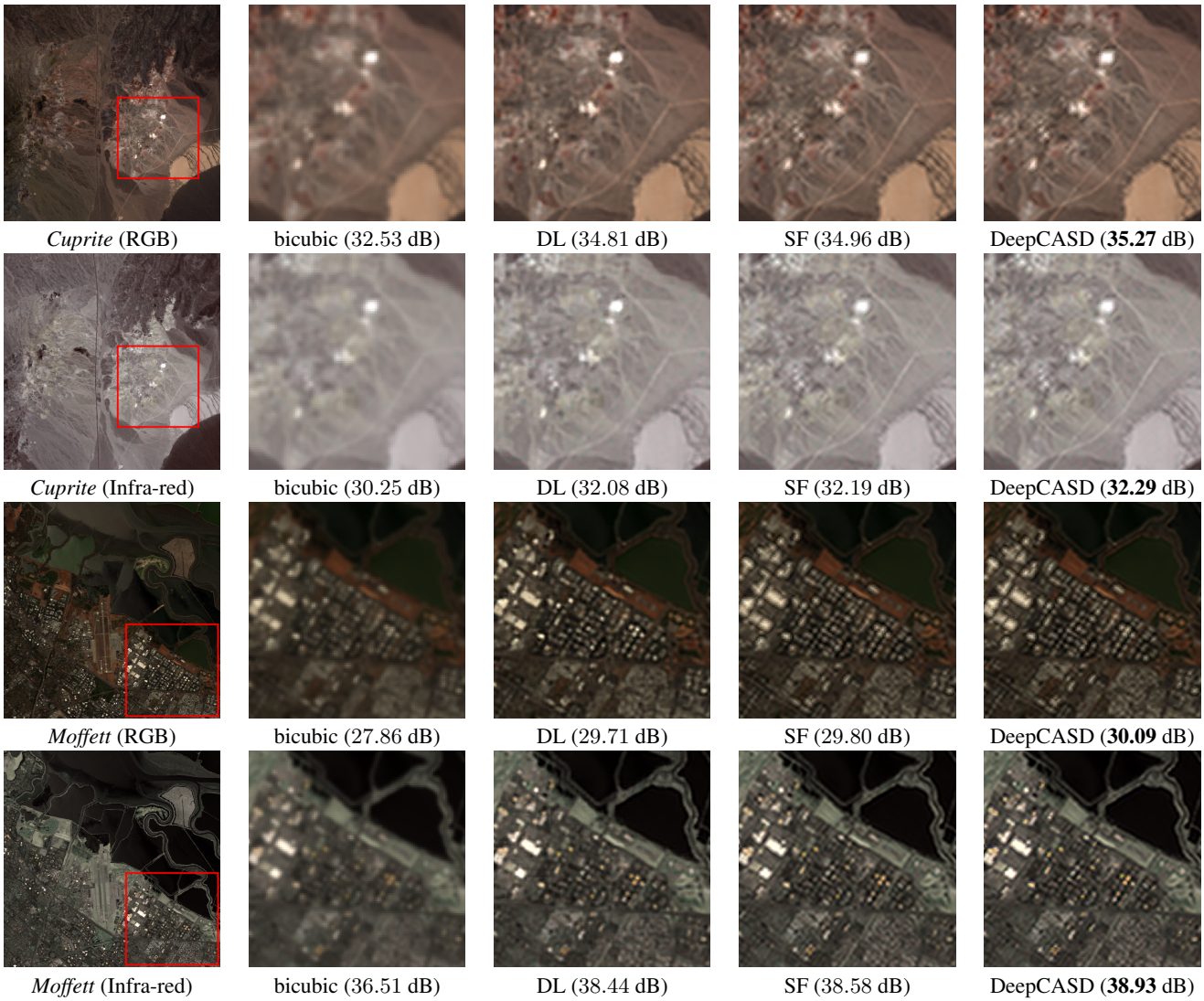

Fig. 4. The zoom-in of example regions, and their $\times 2$ SR results of RGB and Infra-red channels of MS images, using different SR methods.

may achieve slightly better performance in SR compared to stage-wise training, but is more time-consuming. This agrees with the observations reported in [20]. We choose stage-wise training for its efficiency and resilience to degradation in SR performance, compared to joint training. Once the training is complete, we super-resolve the MS images of other regions in the AVIRIS dataset, which contain diverse geometric properties. We use the reconstructed image PSNR as the quality metric.

4.2. Results

We compare the SR performance of MS images using the proposed method, to several other popular SR methods including up-scaling by bicubic interpolation, the state-of-the-art dictionary learning (DL) based SR [9], and the deep SF network [20]. For fair comparison, we use 3-stage SF network, the same number of stages used in the single-channel module of our DeepCASD framework.

Figure 4 shows example SR results of the true color (RGB) and false color of three infra-red channels of MS images. Comparing to the results obtained by DL [9] and SF [20] methods, DeepCASD is capable of discovering more details, especially by better resolving the important local regions such as roads and buildings.

To analyze the performance of MS image SR results quantitatively, we list the reconstructed MS image PSNRs in Table 1 for four different testing areas in California. Each PSNR value is averaged over 16 MS channels, obtained using the aforementioned methods. It is clear that the proposed DeepCASD scheme outperforms all competing methods for all testing MS images. The average PSNR improvement of DeepCASD results over bicubic interpolation, dictionary learning (DL) based SR [9], and the SF network [20] are 2.2 dB, 0.8 dB, and 0.2 dB, respectively.

5. CONCLUSION

We propose a novel data-driven method using deep coupled analysis and synthesis dictionary (DeepCASD) framework for multi-spectral image super-resolution. Our method allows couplings of convolutional dictionaries within and across multiple image channels while leveraging high-dimensional data in an effective way within an end-to-end training process. Experiments on MS image super-resolution using synthesized data show that the proposed DeepCASD framework consistently outperforms competing methods.

6. REFERENCES

[1] NASA, "AVIRIS data repository," available at `https://aviris.jpl.nasa.gov/`.

[2] E. Hege, D. O'Connell, W. Johnson, S. Basty, and E. Dereniak, "Hyperspectral imaging for astronomy and space surviellance," in *Imaging Spectrometry IX*, 2004, vol. 5159, pp. 380–391.

[3] Y. Chen, K. Chao, and M. Kim, "Machine vision technology for agricultural applications," *Computers and electronics in Agriculture*, vol. 36, no. 2, pp. 173–191, 2002.

[4] J. Solomon and B. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1152, 1985.

[5] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 2337–2352, 2016.

[6] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3586–3594.

[7] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[8] D. Liu, Z. Wang, B. Wen, J. Yang, W. Han, and T. Huang, "Robust single image super-resolution via deep networks with sparse prior," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3194–3207, 2016.

[9] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.

[10] C. Laben and B. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," Jan. 4 2000, US Patent 6,011,875.

[11] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4729–4743, 2014.

[12] Z. Zha, X. Liu, X. Huang, H. Shi, Y. Xu, Q. Wang, L. Tang, and X. Zhang, "Analyzing the group sparsity based on the rank minimization methods," in *Multimedia and Expo (ICME), 2017 IEEE International Conference on*. IEEE, 2017, pp. 883–888.

[13] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[14] D. Liu and P. Boufounos, "Dictionary learning based pan-sharpening," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2397–2400.

[15] B. Wen, S. Ravishankar, and Y. Bresler, "FRIST - flipping and rotation invariant sparsifying transform learning and applications," *Inverse Problems*, vol. 33, no. 7, pp. 074007, 2017.

[16] B. Wen, Y. Li, and Y. Bresler, "When sparsity meets low-rankness: Transform learning with non-local low-rank constraint for image restoration," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017.

[17] A. Teodoro, J. Bioucas-Dias, and M. Figueiredo, "Sharpening hyperspectral images using plug-and-play priors," in *International Conference on Latent Variable Analysis and Signal Separation*. Springer, 2017, pp. 392–402.

[18] B. Wohlberg, "Efficient algorithms for convolutional sparse representations," *IEEE Transactions on Image Processing (TIP)*, vol. 25, no. 1, pp. 301–315, 2016.

[19] K. Degraux, U. S Kamilov, P. Boufounos, and D. Liu, "Online convolutional dictionary learning for multimodal imaging," *arXiv preprint arXiv:1706.04256*, 2017.

[20] U. Schmidt and S. Roth, "Shrinkage fields for effective image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2774–2781.

[21] A. Marquina and S. Osher, "Image super-resolution by tv-regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.

[22] H. Demirel and G. Anbarjafari, "Image resolution enhancement by using discrete and stationary wavelet decomposition," *IEEE transactions on image processing*, vol. 20, no. 5, pp. 1458–1460, 2011.

[23] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.

[24] U. Kamilov, H. Mansour, and D. Liu, "Learning convolutional proximal filters," in *Workshop on Signal Process. with Adaptive Sparse Structured Representations (SPARS)*, 2017.

[25] J. Yang, K. Yu, and T. Huang, "Supervised translation-invariant sparse coding," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 3517–3524.

[26] L. Pfister and Y. Bresler, "Learning sparsifying filter banks," in *Proc. SPIE Wavelets & Sparsity XVI*. 2015, vol. 9597, SPIE.

[27] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[28] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2010, pp. 399–406.