UNROLLED PROJECTED GRADIENT DESCENT FOR MULTI-SPECTRAL IMAGE FUSION

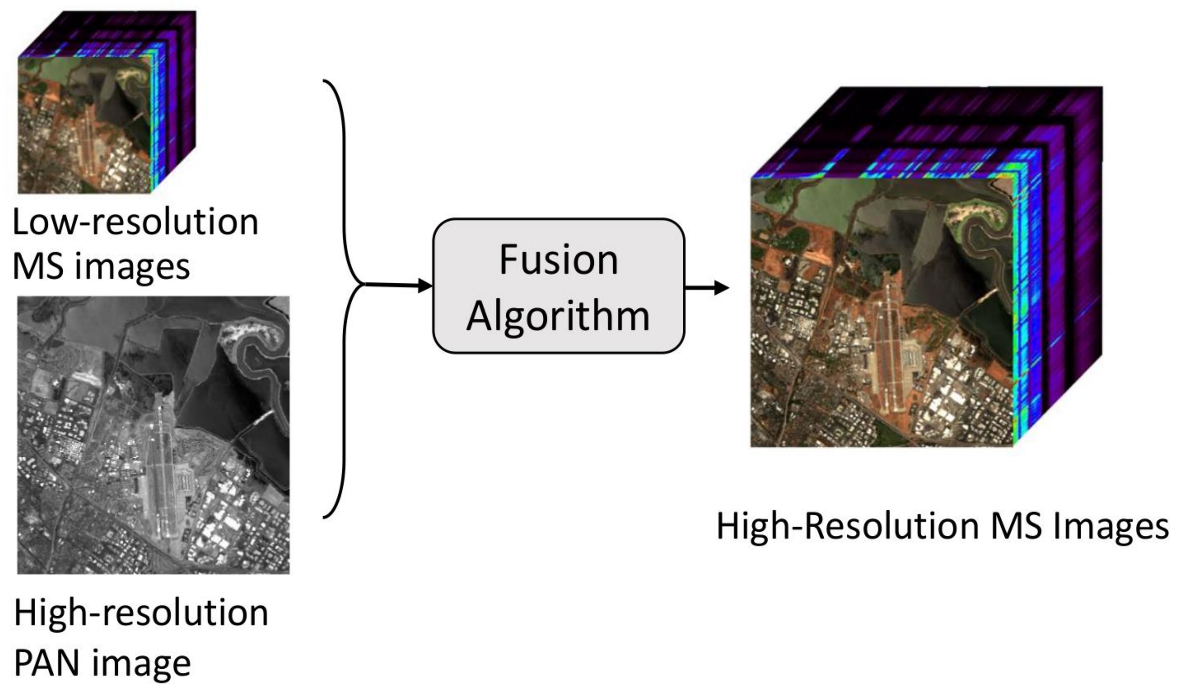
Suhas Lohit[1], Dehong Liu[2], Hassan Mansour[2], and Petros Boufounos[2]

[1]Arizona State University, Tempe, AZ, USA

[2]Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, USA

Introduction

➤ Multi-spectral (MS) image fusion



- There is a fundamental trade-off between spectral resolution and spatial resolution.
- Our goal is to achieve high resolution in both spectral and spatial domains by fusing the information in the images.
- We propose a hybrid model/data-driven approach inspired by well studied signal processing algorithms.

➤ Projected gradient descent (PGD) for linear inverse problems

- The given MS image $\mathbf{y}$ can be modeled as measurements of unknown high-resolution MS $\mathbf{x}$ through a blurring and down-sampling operator $\mathbf{A}$

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

- The fusion problem can be posed a constrained optimization problem, where the constraint set is the set of high-resolution images.

$$\mathbf{x}^* = \operatorname{argmin} \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 \text{ s.t. } \mathbf{x} \in \mathcal{C}$$

- The optimization problem can be solved using projected gradient descent.

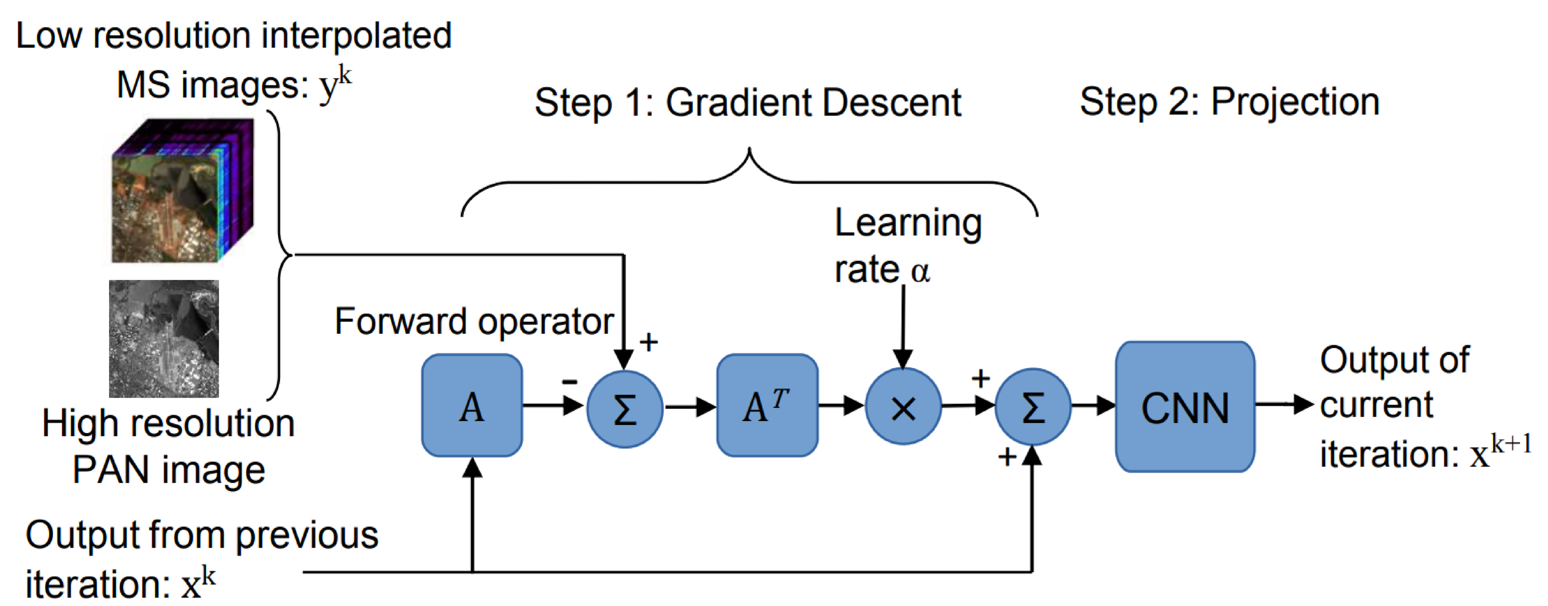
$$\mathbf{w}^{k+1} = \mathbf{x}^k + \alpha \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^k)$$
$$\mathbf{x}^{k+1} = \Pi_{\mathcal{C}}(\mathbf{w}^{k+1})$$

- However, usually both the constraint set and the projection operator cannot be expressed analytically.

Proposed Method

➤ Unrolled PGD

- We propose to unroll the PGD iterations for a predetermined number of iterations and use a trained neural network as the projection operator.
- The convolutional neural network (CNN) is trained to map to the space of high resolution images using Euclidean loss.
- Note that the input images are first up-sampled using bicubic interpolation.



▪ Using $\mathbf{A} = \mathbf{I}$

- As $\mathbf{A}$ is unknown, we can make a simplifying assumption that $\mathbf{A}$ is equal to identity.
- The linear inversion problem reduces to the problem of denoising and the PGD algorithm can be solved in one step. This makes the algorithm equivalent to [1].

▪ Learning $\mathbf{A}$ jointly with the CNN

- We also propose jointly learning the blurring operator $\mathbf{A}$, jointly with the projection operator.
- We assume that the blurring is uniform over the entire image. Thus, we can model $\mathbf{A}$ using a single convolutional kernel $\mathbf{K}_\mathbf{A}$ of size S×S (we use S = 9 in our experiments).
- We also enforce additional constraints on $\mathbf{K}_\mathbf{A}$ so as to make it a valid blurring operator.

$$\mathbf{K}_\mathbf{A} = \mathbf{K}_\mathbf{B} + \mathbf{K}_\mathbf{I}, \text{ s.t. } \sum_{i=1}^{S}\sum_{j=1}^{S} \mathbf{K}_\mathbf{A}(i,j) = 1$$
$$\mathbf{K}_\mathbf{A}(i,j) \geq 0, \forall i,j \in \{1,\ldots,S\}$$

where the coefficients of kernel $\mathbf{K}_\mathbf{B}$ are learned and $\mathbf{K}_\mathbf{I}$ is the identity filter.

Experimental Results

➤ Datasets, training and testing protocol

- We consider synthesized MS images of 17 discrete channels, including panchromatic, RGB, infra-red, and short-wave infra-red channels using publicly available NASA AVIRIS database.
- The CNN in each stage consists of 4 convolutional layers. The networks are trained to perform block-wise image fusion with a block size of 32×32.
- The training data set contains 138 pairs of high-resolution aerial multi-spectral images and their corresponding low-resolution measurements. The networks are trained to super-resolve from 128×128 to 256×256.
- The test set consists of 4 aerial MS images. At test time, the networks super-resolve from 256×256 to 512×512.

➤ Results

- We compare with two learning based baselines – Shrinkage Fields [2] and DeepCASD [3]. Our results yield higher PSNRs and are sharper in quality.

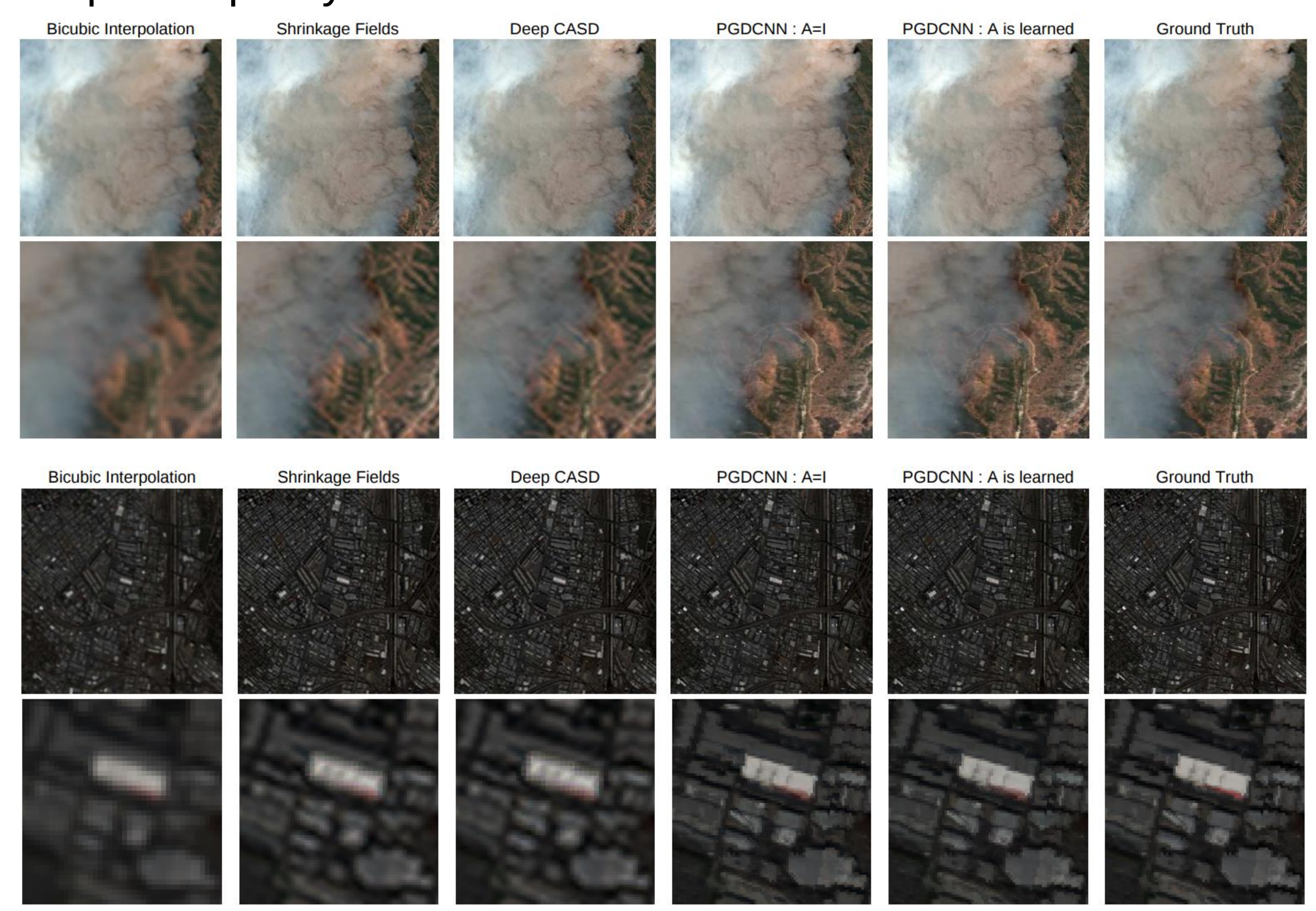


Image Name	Bicubic	Shrinkage Fields	DeepCASD	Unrolled PGD: $\mathbf{A} = \mathbf{I}$ (reduces to [1]), Number of Layers: 4	12	20	Unrolled PGD: $\mathbf{A}$ is learned, Number of Iterations: 1	3	5
Moffett	32.24 0.4788	34.21 0.6981	34.53 0.7185	37.44 0.9710	38.29 0.9768	37.46 0.9729	37.59 0.9706	38.52 0.9778	38.17 0.9776
Cambria Fire	35.32 0.5887	37.51 0.7941	37.62 0.7987	37.83 0.9734	38.91 0.9734	38.71 0.9696	37.99 0.9775	38.91 0.9765	39.33 0.9771
Cuprite	32.44 0.5060	34.33 0.7437	34.52 0.7616	36.88 0.9750	37.56 0.9842	36.82 0.9823	37.95 0.9794	38.56 0.9837	39.02 0.9840
Los Angeles	27.96 0.4888	30.39 0.7628	30.50 0.7761	36.27 0.9702	37.38 0.9755	37.28 0.9760	36.42 0.9712	37.79 0.9777	37.77 0.9790
Mean	31.99 0.5156	34.11 0.7497	34.29 0.7637	37.11 0.9760	38.03 0.9775	37.57 0.9752	37.49 0.9746	38.45 0.9789	38.57 0.9794

[1] Yancong Wei, Qiangqiang Yuan, Huanfeng Shen, and Liangpei Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," IEEE GRSL, 2017

[2] Uwe Schmidt and Stefan Roth, "Shrinkage fields for effective image restoration," in Proc. IEEE CVPR, 2014

[3] Bihan Wen, Ulugbek Kamilov, Dehong Liu, Hassan Mansour, and Petros T Boufounos, "DeepCASD: An end-to-end approach for multi-spectural image super-resolution," in Proc. IEEE ICASSP, 2018

*SL performed this work while at MERL.