Deep unfolding for hypersharpening using a high-frequency injection module

Jamila Mifdal[1], Marc Tomás-Cruz[2], Alessandro Sebastianelli[1], Bartomeu Coll[2], Joan Duran[2]

[1] European Space Agency (ESA)

[2] Universitat de les Illes Balears (UIB) & Institute of Applied Computing Community Code (IAC3)

Funded by the European Union NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

Introduction

This paper proposes a deep unfolding model-driven approach for hypersharpening, which involves merging a hyperspectral (HS) image and a panchromatic (PAN) image. The key contributions of this work can be summarized as follows: firstly, a novel model-based method is introduced for hypersharpening by incorporating a constraint for injecting high-frequency details from the panchromatic image into the fused image. Secondly, the model is formulated as an energy functional optimization problem, solved using a forward-backward splitting algorithm. The solution is further unfolded into a deep learning framework.

Mathematical Model

The observation models [5, 4] relating $\mathbf{H}$ and $\mathbf{P}$ with $\mathbf{U}$ are generally given by

$$\begin{aligned} \mathbf{H} &= \mathbf{DBU} + \boldsymbol{\eta}_h, \\ \mathbf{P} &= \mathbf{US} + \boldsymbol{\eta}_p, \end{aligned} \tag{1}$$

where $\mathbf{B} \in \mathbb{R}^{N\times N}$ is the low-pass filter modeling the point spread function of the HS sensors, $\mathbf{D} \in \mathbb{R}^{n\times N}$ is the l-fold downsampling operator, $\mathbf{S} \in \mathbb{R}^{C\times 1}$ is the spectral response of the PAN sensor, and $\boldsymbol{\eta}_h$ and $\boldsymbol{\eta}_p$ are assumed to be additive, white Gaussian noise.

Then, $\mathbf{U}$ can be estimated by solving the following minimization problem:

$$\min_{\mathbf{U}} \frac{1}{2}\|\mathbf{DBU} - \mathbf{H}\|^2 + \frac{\gamma}{2}\|\mathbf{US} - \mathbf{P}\|^2 + \mu R(\mathbf{U}), \tag{2}$$

On one hand, the high frequencies of the fused image can be estimated as $\mathbf{U} - \tilde{\mathbf{H}}$, where $\tilde{\mathbf{H}} \in \mathbb{R}^{N\times C}$ is the result of upscaling $\mathbf{H}$ by bicubic interpolation. On the other hand, the high frequencies of the scene are given by $\mathbf{P} - \tilde{\mathbf{P}}$, where $\tilde{\mathbf{P}} \in \mathbb{R}^{N\times 1}$ contains the low frequencies of the PAN data.

$$U_{ij} - \tilde{H}_{ij} = \frac{\tilde{H}_{ij}}{\tilde{P}_i}(P_i - \tilde{P}_i), \tag{3}$$

We can assume that $\mathbf{U}$ can be linearly represented by $\mathbf{P}$ and an unknown matrix $\mathbf{V} \in \mathbb{R}^{N\times(r-1)}$, where $r = \text{rank}(\mathbf{U}) > 1$, i.e.,

$$\mathbf{U} = \mathbf{PX} + \mathbf{VY} \tag{4}$$

Putting it all together, the proposed fusion model is

$$\begin{aligned} \min_{\mathbf{V}} \ \mu R(\mathbf{V}) + F(\mathbf{V}) = \min_{\mathbf{V}} \ \mu R(\mathbf{V}) + \frac{1}{2}\|\mathbf{DB}\left(\mathbf{PX} + \mathbf{VY}\right) - \mathbf{H}\|^2 \\ + \frac{\lambda}{2}\|\tilde{\mathbf{P}} \circ \left(\mathbf{PX} + \mathbf{VY}\right) - \mathbf{P} \circ \tilde{\mathbf{H}}\|^2, \end{aligned} \tag{5}$$

Minimitzation Method

To solve (5) we use the forward-backward splitting method [2, 1] to compute the solution. The basic idea is to combine an explicit step of descent in the smooth function F with an implicit step of descent in μR:

$$\mathbf{V}^{k+1} = \text{prox}_{\tau\mu}\left(\mathbf{V}^k - \tau\nabla F(\mathbf{V}^k)\right), \tag{6}$$

where

$$\begin{aligned} \nabla F(\mathbf{V}) = \mathbf{B}^\top\mathbf{D}^\top\left(\mathbf{DB}\left(\mathbf{PX} + \mathbf{VY}\right) - \mathbf{H}\right)\mathbf{Y}^\top \\ + \lambda\left[\tilde{\mathbf{P}} \circ \left(\tilde{\mathbf{P}} \circ \left(\mathbf{PX} + \mathbf{VY}\right) - \mathbf{P} \circ \tilde{\mathbf{H}}\right)\right]\mathbf{Y}^\top. \end{aligned} \tag{7}$$

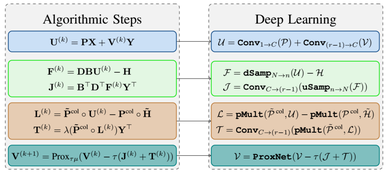

Figure 1. Relationship between the steps of the optimization algorithm and the modules of the deep unfolded network. The operator **ProxNet** stands for the proximal operator and it is replaced by a ResNet [3] as suggested in [7] and $\mathbf{uSamp}_{n_{in}\to n_{out}}$ and $\mathbf{dSamp}_{n_{in}\to n_{out}}$ are the learned upsampling and downsampling operators.

Deep Learning Model

The right side of Figure 1 shows the corresponding operations in the DL framework of the hypersharpening algorithm. The four blocs highlighted in Figure 1 are the main components of the complete network illustrated in Figure 2 where we could see three main stages. Each stage is composed of the blocks detailed in Figure 1 and at each epoch all three stages are executed, then, the estimated result is fed to the following loss function.

$$L = \|\hat{\mathbf{U}}^{(k)} - \mathbf{U}\|^2 + \alpha\|\mathbf{F}^{(k)}\|^2 + \beta\|\mathbf{L}^{(k)}\|^2, \tag{8}$$

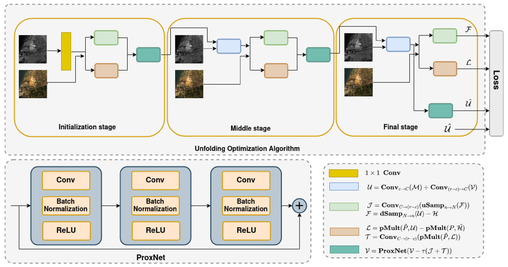

Figure 2. The unfolded network of the hypersharpening algorihtm using a high-frequency injection module. The network is composed of three stages, each one of these stages follows the unfolding steps detailed in Figure 1.

Implementation details for PRISMA dataset

Our model is trained in a PyTorch framework, using an Nvidia A100 GPU, during 4500 epochs for the PRISMA dataset. We use an Adam optimizer with a learning rate of 10^{-3} and a batch size of 8 images. The trade-off parameters α and β are fixed at 10^{-3}.

The PAN data contains one single band at a spatial resolution of 5m. We selected and downloaded 20 large-scale scenes of PRISMA images throughout the PRISMA mission's portal (https://prisma.asi.it). The downloaded HS images have an original size of $1000 \times 1000 \times 240$, each one of the scenes was cropped into non-overlapping tiles of $128 \times 128 \times 240$. Given that the SWIR bands are not covered by the spectral response of the PAN sensor, only the first 66 bands were considered which resulted in tiles of $128 \times 128 \times 66$, from each tile a new HS and PAN images were generated following the Wald protocol [6] and using the spectral and spatial responses provided by PRISMA mission engineers. The chosen downsampling factor for the PRISMA dataset is 12, thus, from each tile of $128 \times 128 \times 66$, an HS image of $11 \times 11 \times 66$ and PAN image of the size 128×128 were considered for the fusion process which is called hyper-sharpening.

Quantitative Results

	ERGAS ↓	PSNR ↑	SSIM ↑	DD ↓	SAM ↓
PCA	376.06	15.70	0.3990	0.1333	35.41
Brovey	92.86	27.68	0.9180	0.0309	4.69
Bicubic	225.81	23.40	0.8303	0.0420	4.64
GS	92.14	27.75	0.9181	0.0308	4.78
GSA	186.01	23.98	0.8706	0.0399	4.67
IHS	101.00	26.78	0.8882	0.0346	7.20
SFIM	208.91	23.95	0.8801	0.0392	4.67
DiCNN	41.44	33.38	0.9520	0.0145	3.70
MSDCNN	43.10	33.07	0.9496	0.0153	4.06
GPPNN	253.18	20.99	0.8453	0.0700	7.92
MHFnet	45.29	32.69	0.9402	0.0157	4.13
Ours	**15.31**	**42.17**	**0.9900**	**0.0078**	**1.59**

Table 1. Average of the quality measures over all images of the PRISMA dataset. The methods are divided in classical, pure DL and unfolding methods. The best results are in bold and the second best ones are underlined.

Qualitative Results

The visualization results of the model on the Prisma dataset are depicted in Figure 3. These visual outputs are complemented by quantitative findings presented in Table 1. The qualitative analysis further supports the quantitative results, establishing a comprehensive understanding of the model's performance.

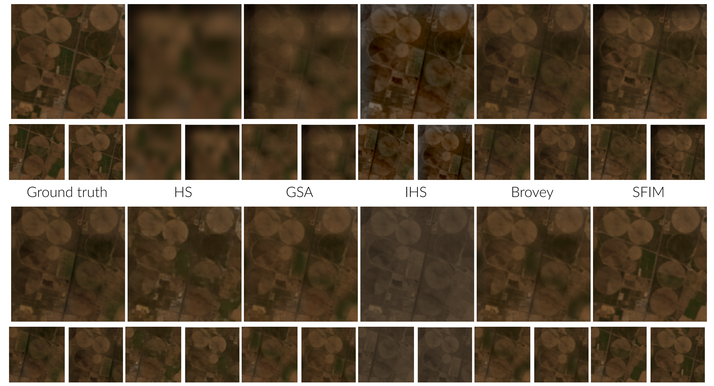

Figure 3. Visual comparison of the fusion approaches on an image of the PRISMA dataset. We display the 35th, 45th and the 57th bands in place of the RGB channels. The proposed deep unfolding network successfully combines the geometry of the PAN image with the spectral information of the HS data, while all other results are affected by blur, color artifacts and spatial distortions. Our method is also able to recover both large structures, such as the circular ground contours, and the finest ones, such as roads and small building structures.

Acknowledgements

This work is part of the MaLiSat project TED2021-132644B-I00 funded by MCIN/AEI/10.13039/501100011033 and by the European Union "NextGenerationEU"/PRTR. The authors were also supported by the Conselleria de Fons Europeus, Universitat i Cultura del Govern de les Illes Balears under grant AP_2021_023.

References

[1] Antonin Chambolle and Thomas Pock. "An introduction to continuous optimization for imaging". In: *Acta Numerica* 25 (2016), pp. 161–319.

[2] Patrick L Combettes and Valérie R Wajs. "Signal recovery by proximal forward-backward splitting". In: *Multiscale modeling & simulation* 4.4 (2005), pp. 1168–1200.

[3] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[4] Jamila Mifdal et al. "Variational Fusion of Hyperspectral Data by Non-Local Filtering". In: *Mathematics* 9.11 (2021), p. 1265.

[5] Rafael Molina, Aggelos K Katsaggelos, and Javier Mateos. "Bayesian and regularization methods for hyperparameter estimation in image restoration". In: *IEEE Transactions on Image Processing* 8.2 (1999), pp. 231–246.

[6] Lucien Wald, Thierry Ranchin, and Marc Mangolini. "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images". In: *Photogrammetric engineering and remote sensing* 63.6 (1997), pp. 691–699.

[7] Qi Xie et al. "Multispectral and hyperspectral image fusion by MS/HS fusion net". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 1585–1594.