# LOCATION AWARE SUPER-RESOLUTION FOR SATELLITE DATA FUSION

*Olaoluwa Adigun*

Dept. of Electrical and Computer Engineering
University of Southern California
Los Angeles, CA 90089-2564

*Peder A. Olsen, Ranveer Chandra*

Microsoft Research
Research For Industry
14820 NE 36th St, Redmond, WA 98052

## ABSTRACT

Satellite data fusion involves images with different spatial, temporal, and spectral resolution. These images are taken under different illumination conditions, with different sensors and atmospheric noise. We use classic super-resolution algorithms to synthesize commercial satellite images (Pléiades) from a public satellite source (Sentinel-2). Each super-resolution method is then further improved by adaptive sharpening to the location by use of matrix completion (regression with missing pixels). Finally, we consider ensemble systems and a residual channel attention dual network with stochastic dropout. The resulting systems are visibly less blurry with higher fidelity and yield improved performance.

***Index Terms*—** Super-resolution, matrix completion, cloud removal.

## 1. INTRODUCTION

The Sentinel-2 mission [1] from the European Space Agency has been making planet-wide 13-band multispectral observations with 10m×10m pixel resolution and a 5-day revisit time since it was launched in 2015. Our high-resolution images comes from the Pléiades HR 1A/B satellites [2] that provide 4-band multispectral observations at 2m×2m resolution (panchromatic at 0.5m). These historic images are employed for enhanced super-resolution and to give a steady stream of synthetically generated Pléiades images using Sentinel-2 observations. Fusing public and commercial satellite data is unique in that the commercial satellite data may only provide a *very small number* (1-4) of historical images over a period of several years, while in the public-public setting as in the celebrated STAR-FM method [3] there is a steady stream of images from both satellites.

## 2. CLOUD INPAINTING

We use the matrix completion from [4] for cloud inpainting for Sentinel-2 and Pléiades. This is particularly important for Sentinel-2 as more than 70% of the earth is covered in clouds. The matrix completion decomposition derived on the high-resolution historical data is then used to *adaptively sharpen* the super-resolution estimate.

**Terminology:** The rank is $N_r$, the image size is $N_h \times N_w$, and the respective cardinality for time and spectrum are $N_t$ and $N_c$. Boldfaced variables are matrices with dimension $N_c N_t \times N_h N_w$ unless otherwise noted. $\mathbf{Y}$ is the satellite observation matrix (stacked Sentinel-1 and Sentinel- 2 or Pléiades), $\mathbf{M}$ is a mask of cloud-free pixels (1 for cloud-free, 0 for cloudy), and $\mathbf{X}$ is the cloud-free reconstruction with a low-rank representation $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$, with $\mathbf{U} \in \mathbb{R}^{N_c N_t \times N_r}$, $\mathbf{V} \in \mathbb{R}^{N_h N_w \times N_r}$. $L_t$ and $H_t$ are the respective low-resolution and high-resolution images. The symbol $\uparrow$ represents bilinear upsampling, and $\circ$ represents element wise multiplication.

**Matrix Completion:** We minimize the following loss function over rank $N_r$ matrices $\mathbf{X} = \mathbf{U}\mathbf{V}^\top$

$$F(\mathbf{X}) = \|\mathbf{M} \circ (\mathbf{X} - \mathbf{Y})\|_F^2 + \alpha \sum_{t=1}^{N_t - 1} \|\mathbf{X}_{t+1} - \mathbf{X}_t\|_F^2. \quad (1)$$

Here $\mathbf{V}$ represents abstract land types (lakes, grass lands, etc.), while $\mathbf{U}$ represents the spectro-temporal evolution of each land type. The method in Algorithm 1 was designed specifically to solve (1) efficiently on a GPU. It was used to provide cloud-free images for Sentinel-2 and for Pléiades.

**Matrix Completion Sharpening:** $H_t$ is approximated by $X_t = U_t \mathbf{V}^\top$, where $U_t \in \mathbb{R}^{N_c \times N_r}$ is the spectro-temporal evolution at time t. This becomes interesting when we extrapolate to a time $t$ where $H_t$ is not known. In this case we derive $\mathbf{V}$ from historical high-resolution data $\mathbf{Y}$ and estimate $U_t$ using a super-resolution approximation $\hat{H}_t = \hat{H}_t(L_t) \approx H_t$. Specifically $U_t$ is found by solving[1] $\min_{U_t} \|U_t \mathbf{V}^\top - \hat{H}_t\|_F^2 + \beta \|U_t\|_F^2$. The analytic solution is $\hat{U}_t = \hat{H}_t \mathbf{V}(\mathbf{V}^\top \mathbf{V} + \beta I)^{-1}$ with a reconstruction $\hat{U}_t \mathbf{V}^\top$. The reconstruction $U_t \mathbf{V}^\top$ yields sharper and less blurry images than $\hat{H}_t$, stemming from the high-resolution land types. This can be seen visually in Figure 1. The estimation is done on small patches (25×25 pixels) to balance fidelity against spectral approximation errors. The full stitched image estimate is denoted $\mathcal{S}_L(\hat{H}_t)$ where the dependence of the historical $H_{t-1}, H_{t-2}, \ldots$ is not explicitly shown.

[1] $\beta$ is set to 0.02 only to avoid problems with singularities

**Algorithm 1:** The Cloud Completion Algorithm.

```
Input: Satellite data: $\mathbf{Y}$, cloud-free mask: $\mathbf{M}$, rank
       $N_r$, damping coefficient $\alpha$
Output: Cloud-free data: $\mathbf{X} = \mathbf{U}\mathbf{V}^T$,
        $\mathbf{U} \in \mathbb{R}^{N_c N_t \times N_r}$, $\mathbf{V} \in \mathbb{R}^{N_h N_w \times N_r}$
$\mathbf{D} \in \mathbb{R}^{N_t \times N_t} \leftarrow 0$   // forward difference
$D_{ii} \leftarrow -1, D_{i,i+1} \leftarrow 1, \forall i \in \{1, \ldots, N_t - 1\}$
$\mathbf{Z} \in \mathbb{R}^{N_c N_t \times N_h N_w} \leftarrow 0, \mathbf{\Delta} = \mathbf{D} \otimes \mathbf{I}_{N_c}$
for iter $\leftarrow 1$ to 300 do
    $\mathbf{Y}_Z \leftarrow ((\mathbf{M} \circ \mathbf{Y}) - (1 - \mathbf{M}) \circ \mathbf{Z})$
    $\mathbf{U} \leftarrow (\mathbf{I} + \alpha \mathbf{\Delta}^\top \mathbf{\Delta})^{-1} \mathbf{Y}_Z \mathbf{V} (\mathbf{V}^\top \mathbf{V})^{-1}$
    $\mathbf{V} \leftarrow \mathbf{Y}_Z^\top \mathbf{U} (\mathbf{U}\mathbf{U}^\top + \alpha \mathbf{U}^\top \mathbf{\Delta}^\top \mathbf{\Delta} \mathbf{U})^{-1}$
    $\mathbf{Z} \leftarrow (1 - \mathbf{M}) \circ (\mathbf{U}\mathbf{V}^\top)$.
// $\mathbf{U}$: evolution, $\mathbf{V}$: land types
$\mathbf{X} \leftarrow \mathbf{U}\mathbf{V}^\top$
```

## 3. SUPER-RESOLUTION ALGORITHMS

In this section we briefly survey five super-resolution algorithms used in this paper starting from the first neural network based method to state-of-the-art methods. The *super-resolution convolutional neural network* (SRCNN) [5] is a 2-4 layer deep convolutional neural network that maps pre-upsampled images $L_{t\uparrow}$ to a super-resolution estimate $\mathcal{S}_{\mathtt{SRCNN}}(L_{t\uparrow})$. The *residual dense network* (RDN) [6, 7] is made up of several residual-in-residual dense blocks (RRDB) each consisting of several residual dense blocks. RDN is deep, compact and also used for image transformation, making it a great fit for satellite fusion. The RDN super-resolution system is denoted $\mathcal{S}_{\mathtt{RDN}}(L_t)$. Next, the *super-resolution generative adversarial network* (SRGAN) [8] uses a GAN architecture where the generator transforms low-resolution images to high-resolution images while the discriminator measures similarity between the target and generated images using a perceptual loss. The batch normalization layers of SRGAN tend to introduce unpleasant artifacts and limit generalization that the Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) [9] solved by replacing batch normalization layers with residual-in-residual dense blocks (RRDB).The Residual Channel Attention Network (RCAN) [10] proposed a multiplicative channel attention mechanism that adaptively weights the channel feature, and they use short and long skip connections to train very deep networks.

Table 1 shows results for five baseline super-resolution systems for the satellite fusion task. The matrix completion sharpening improves the performance across the board. Although the improvements look modest from the metrics, the results are less blurry and the fidelity is significantly improved as seen in Figure 1. The simplest super-resolution SRCNN performs the best. We ascribe this to the difficulty of training deep models in the presence of the atmospheric and illumination noise and the sensor mismatch. We discuss methods to reverse the situation next.

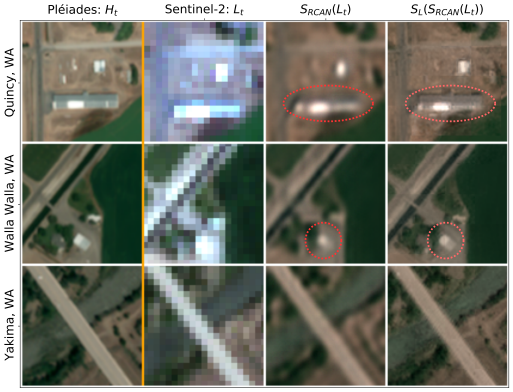


**Fig. 1**: RCAN with and without matrix completion sharpening: The dotted lines highlights features that are significantly sharper and less blurry after matrix completion sharpening.

| | Super-resolution system | | | Matrix compl. sharpened | | |
|---|---|---|---|---|---|---|
| System | MAE | PSNR | SSIM | MAE | PSNR | SSIM |
| LinReg | 0.0556 | 23.454 | 0.768 | 0.0532 | 23.852 | 0.784 |
| SRCNN | 0.0467 | 25.354 | 0.810 | **0.0455** | **25.618** | 0.814 |
| RDN | 0.0481 | 25.081 | 0.816 | 0.0478 | 25.263 | 0.818 |
| RCAN | 0.0483 | 25.067 | 0.810 | 0.0464 | 25.440 | **0.819** |
| SRGAN | 0.0732 | 21.011 | 0.738 | 0.0533 | 24.122 | 0.802 |
| ESRGAN | 0.0617 | 23.171 | 0.771 | 0.0522 | 24.578 | 0.796 |

**Table 1**: Baselines with Matrix Completion Sharpening. LinReg is bilinear upsampling followed by linear regression.

## 4. TRAINING IMPROVEMENTS

To improve the performance of RDN and RCAN we employed both the layer-wise adaptive large batch optimization technique called LAMB [11] training as well as stochastic depth [12]. LAMB is an improved optimization technique that gives stable convergence for larger learning rates and deeper networks, thus enabling RDN and RCAN to surpass the performance of SRCNN as seen in Table 2. For SRCNN LAMB yields no gains. The combination of these two allowed us to train deeper systems. For the stochastic depth we used a survival probability of 1 for the first block and a linear descent down to 1/2 for the last block. For RDN we applied stochastic depth to the residual-in-residual blocks (RRDB) while for RCAN stochastic depth was applied to the residual channel attention blocks (RCAB) layers inside each Residual Group (RG) as shown in Figure 2d.

**Residual Channel Attention dual network with Stochastic Dropout:** We propose a new network that takes high dimensional input $\hat{H}_t$ from the matrix completion sharpening and low-resolution input $L_t$ as shown in Figure 3. The low-resolution branch connects to the high-resolution branch at 3 intermediate points. The communication is made possible by the connector network in Figure 2c which is similar to the RCAB Figure 2a, but with a reduction of the number of filters. As seen in Table 2 the system outperforms all the others. We also show that the performance can be improved by using SRCNN as an ensemble mechanism that takes a stack of the best SRCNN, RDN and RCAN systems as input. We used SRCNN as it takes high-resolution input images. Table 2 shows the improved results.

## 5. CONCLUSION

We introduced matrix completion sharpening and showed that it can be applied to improve existing super-resolution methods and we also showed the importance of the LAMB optimization and stochastic depth for improving accuracy on the task. Finally, we introduced the RCAN-dual-SD network that further improves on the matrix completion sharpening. While these enhancements yield visible improvements, the methods still suffer from propagation of errors from the cloud-detector and the improvements diminish with the number of bands and historical images.

| | Super-resolution system | | | Matrix compl. sharpened | | |
|---|---|---|---|---|---|---|
| System | MAE | PSNR | SSIM | MAE | PSNR | SSIM |
| RDN+LAMB | 0.0456 | 25.566 | 0.826 | 0.0444 | 25.863 | 0.825 |
| +SD | 0.0448 | 25.718 | 0.826 | 0.0503 | 24.941 | 0.806 |
| RCAN+LAMB | 0.0473 | 25.267 | 0.816 | 0.0423 | 26.110 | 0.827 |
| +SD | 0.0461 | 25.630 | 0.817 | 0.0437 | 25.995 | 0.825 |
| RCAN-dual-SD | | | | **0.0420** | **26.454** | **0.843** |
| ensemble | 0.0425 | 26.248 | 0.830 | **0.0400** | **26.765** | **0.846** |

**Table 2**: LAMB training and stochastic depth.

## 6. REFERENCES

[1] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al., “Sentinel-2: ESA’s optical high-resolution mission for GMES operational services,” *Remote sensing of Environment*, vol. 120, pp. 25–36, 2012.

[2] M Alain Gleyzes, Lionel Perret, and Philippe Kubik, “Pleiades system architecture and main performances,” *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 39, no. 1, pp. 537–542, 2012.

[3] Khaled Hazaymeh and Quazi K Hassan, “Spatiotemporal image-fusion model for enhancing the temporal resolution of Landsat-8 surface reflectance images using MODIS images,” *Journal of Applied Remote Sensing*, vol. 9, no. 1, pp. 096095, 2015.

[4] Mingmin Zhao, Peder A Olsen, and Ranveer Chandra, “Seeing through clouds in satellite images,” *arXiv preprint arXiv:2106.08408*, 2021.

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Image super-resolution using deep convolutional networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[6] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, “Residual dense network for image super-resolution,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481.

[7] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu, “Residual dense network for image restoration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

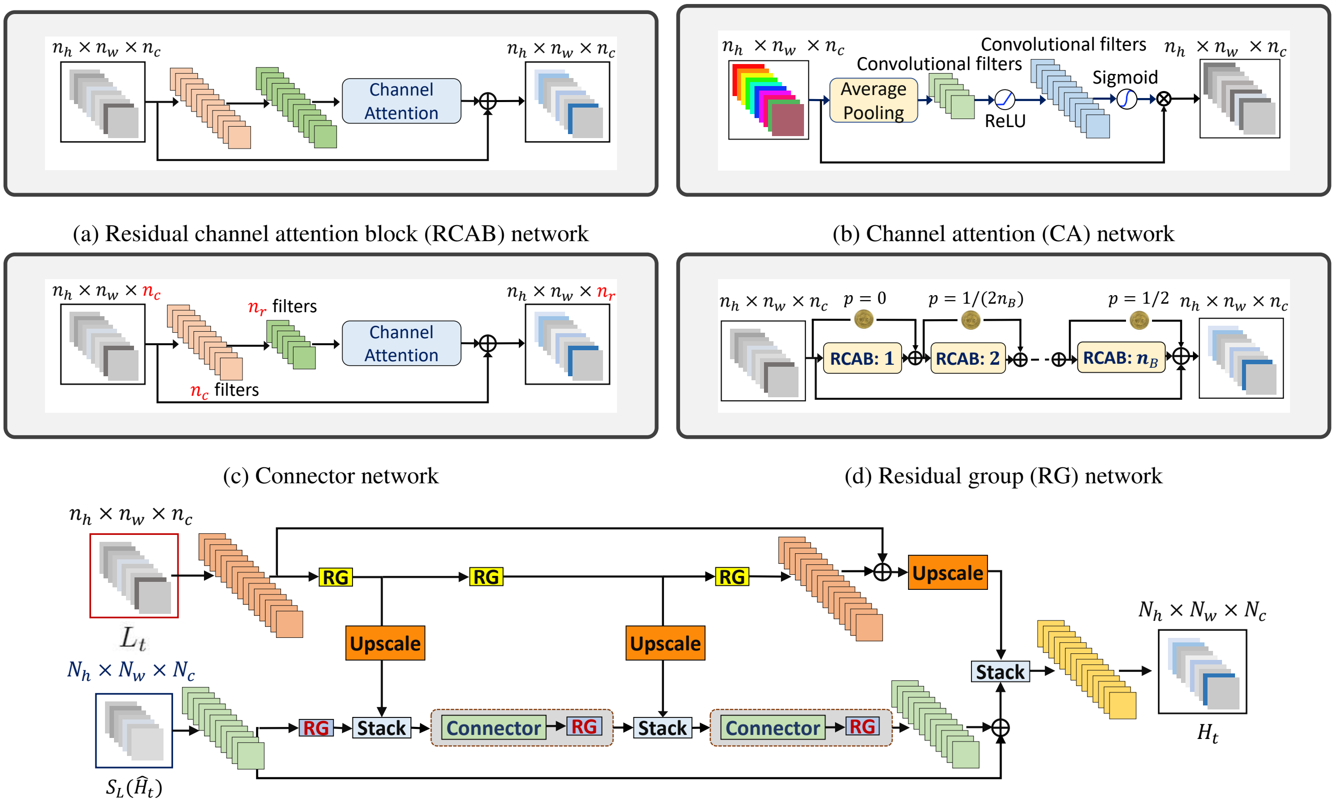
[8] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[9] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy, “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0–0.

[10] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, “Image super-resolution using very deep residual channel attention networks,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[11] Yang You, Jing Li, Sashank Reddi, Jonathan Hseu, Sanjiv Kumar, Srinadh Bhojanapalli, Xiaodan Song, James Demmel, Kurt Keutzer, and Cho-Jui Hsieh, “Large batch optimization for deep learning: Training BERT in 76 minutes,” *arXiv preprint arXiv:1904.00962*, 2019.

[12] Gao Huang, Yu Sun, Zhuang Liu, Daniel Sedra, and Kilian Q Weinberger, “Deep networks with stochastic depth,” in *European conference on computer vision*. Springer, 2016, pp. 646–661.



(a) Residual channel attention block (RCAB) network

(b) Channel attention (CA) network

(c) Connector network

(d) Residual group (RG) network

**Fig. 2**: Full architecture of residual channel attention dual network with stochastic dropout.

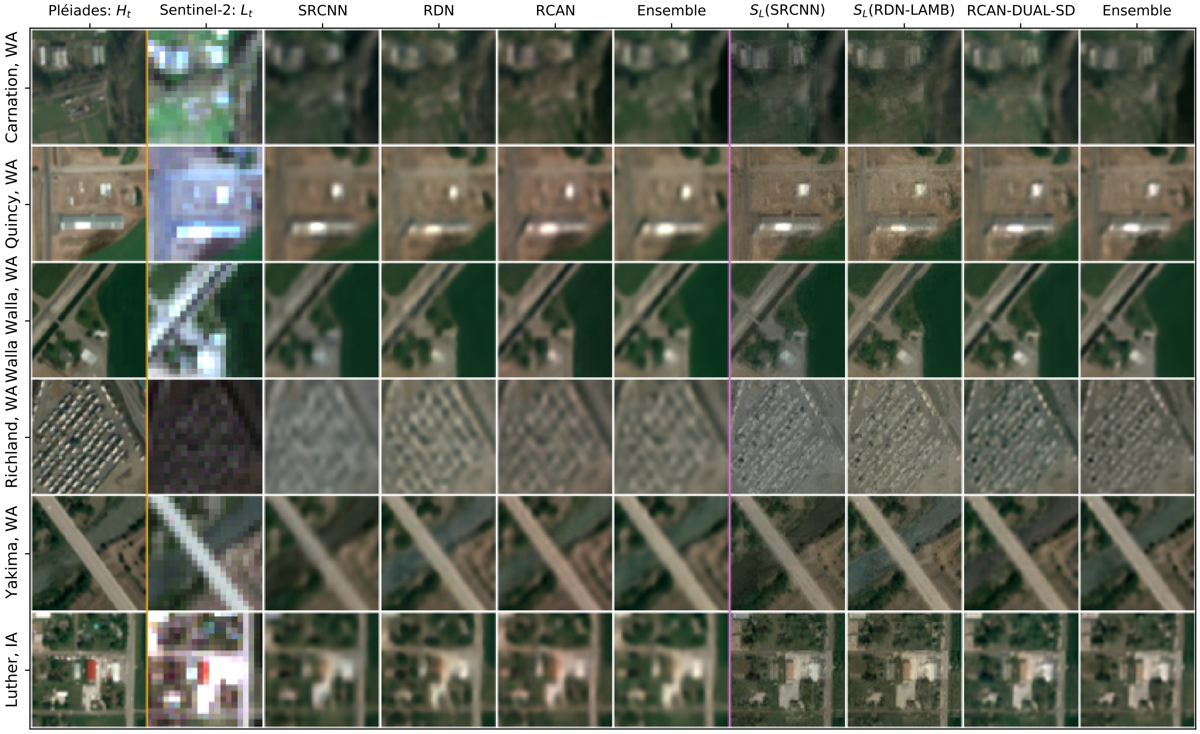


**Fig. 3**: Example of several super-resolution methods from 6 locations. The standard super-resolution methods are notably blurrier compared to the ground truth in column 1. The ensembles combines the 3 preceding columns using SRCNN.