

28-30 SEPTEMBER 2022

## ONBOARD SEQUENCE COMPRESSION THROUGH MULTI-FRAME SUPER-RESOLUTION

Grigorios Tsagkatakis<sup>(1)</sup>, Simon Vellas<sup>(2)</sup>, Sam Bammens<sup>(3)</sup>, Vladimir Zelenevskiy<sup>(4)</sup>, David Evans<sup>(3)</sup>,  
Mathieu Bernou<sup>(2)</sup>, Panagiotis Tsakalides<sup>(1)</sup>

<sup>(1)</sup> *Foundation for Research and Technology Hellas (FORTH)*

*N. Plastira 100, Vassilika Vouton, 71309, Crete, Greece*

*GT: [greg@ics.forth.gr](mailto:greg@ics.forth.gr), PT: [tsakalid@ics.forth.gr](mailto:tsakalid@ics.forth.gr)*

<sup>(2)</sup> *OHB Hellas*

*Ifaistoi 3, Marousi 15124, Greece*

*mailing address continued*

*Email: [simon.vellas@ohb-hellas.gr](mailto:simon.vellas@ohb-hellas.gr), BM: [Mathieu.Bernou@ohb-hellas.gr](mailto:Mathieu.Bernou@ohb-hellas.gr)*

<sup>(3)</sup> *Europeans Space Agency (ESA)*

*Robert-Bosch-Straße 564293 Darmstadt, German*

*SM: [Sam.Bammens@esa.int](mailto:Sam.Bammens@esa.int), DE: [David.Evans@esa.int](mailto:David.Evans@esa.int)*

<sup>(4)</sup> *Telespazio Germany GmbH*

*Europapl. 5, 64293 Darmstadt, Germany*

*Email: [Vladimir.Zelenevskiy@esa.int](mailto:Vladimir.Zelenevskiy@esa.int)*

### INTRODUCTION

Technology miniaturization is a key driver for the next generation of EO platforms utilizing CubeSats, which promises a radical departure from the established paradigm in terms of design and development cost and time, as well as reduced complexity and higher flexibility deployment processes.

CubeSats have the potential to revolutionize Earth Observation by exploring different operating points in terms of spatial, temporal, and spectral resolution. While physical limitations, mostly in terms of component size and weight, prevent them from obtaining high-quality imagery, technology miniaturization has allowed such platforms to support impressive capabilities and flexibility in terms of hardware and software.

One very promising aspect of these capabilities is that even low-cost imaging sensors can acquire relatively low spatial quality but high temporal sampling frequency image sequences, i.e., videos. Furthermore, the advent of flash memory has allowed a dramatic increase in onboard memory. Lastly, system designers can leverage flexible software/hardware designs, for deploying state-of-the-art machine learning systems like deep neural networks offering promising capabilities. On the other side, the increase in space-to-ground communications capabilities has been moving at a much slower pace, leading to scenarios where observations are discarded due to bandwidth limitations.

In this work, we study how one can leverage the capabilities of CubeSats, and the ESA OPS-SAT platform, in particular, to effectively fuse sequences of low-resolution images into a higher resolution one and transmit the generated images instead of the unprocessed ones. By generating a single high-resolution image, different operating points can be explored in terms of the number of frames and spatial resolution. We consider the analysis of scenarios where this process leads to compression of the observations, while at the same time, providing higher quality images.

We formulate the problem as an instance of Multi-Frame Super Resolution (MFSR) where we explore the ability of the onboard HD camera, to acquire multiple frames, i.e., short video or burst sequences, from the same (approximately) ground location and utilize these frames for increasing spatial resolution, by effectively fusing multiple Low Resolution (LR) images into a single High-Resolution (HR) one. To achieve this objective, we consider deep neural networks operating on registered image sequences which can be deployed onboard the satellite.

It is worth noting that one could consider applying this paradigm on the ground, and indeed, the ESA Kelvin competition exploits the ability of PROBA-V to acquire images with different resolutions, to do just that. Despite the potential, this

approach mandates the transfer from the S/C to the G/S and finally to the end-user, a large amount of redundant, low-quality content. This can introduce a bottleneck to space-to-ground communication channels.

## STATE-OF-THE-ART

The Multi-Frame Super-Resolution (MFSR) paradigm extends this paradigm by considering sub-pixel between subsequent frames and utilizing this information to increase the spatial resolution.

State-of-the-art approaches for MFSR applied on remote sensing imagery are based on different deep learning model architectures, which introduce a non-linear fusion of the LR images, and are designed to operate using either unregistered or registered frame sequences. A major milestone in this field was the release of the ESA's "PROBA-V Super Resolution" Kelvin Competition [1] and the associated Proba-V satellite data [2], which took place in 2020. Most of the proposed algorithms have been published after the release of the dataset and thus the dataset serves as the main reference for MFSR on satellite data. A representative list of existing methods for satellite-based MFSR is given next.

Bicubic interpolation is applied to every frame or a selected (reference) frame independently. The method is generic, fast, and can serve as a non-trainable baseline approach.

EDSR [8] is a state-of-the-art method for single-image super-resolution that achieves excellent performance for cases of natural image super-resolution. EDSR architecture is based on the SRResNet architecture and consists of multiple residual blocks. In our case, we consider an architecture of 16 residual blocks with 64 channels. This method will act as the machine learning-based baseline.

DeepSUM [3] is a high-performance method for multi-frame super-resolution which was developed in the context of the ESA Kelvin competition, and in-fact, achieved first place during the competition. The method is composed of three modules integrated into an end-to-end convolutional neural network. The network accepts a sequence of bicubic-up sampled and registered images and processes each one independently. The extracted features are subsequently introduced to the RegNet subnetwork to compute registration filters to register the feature maps. In the last step, the FusionNet subnetwork merges the features of the images to produce a residual image. The residual image is then added element-wise to the average of the registered input to obtain the SR image

3D WDSR [5] is a variant of residual network architecture that operates on pre-registered images and achieves the best performance post-mortem, i.e., performance was validated on testing data after the official end of the ESA Kelvin completion. The method employs a 3D convolution layer followed by a sequence of wide-residual modules as the building blocks and a residual path with an interpolated frame (the central one), either using bicubic interpolation or a 3D convolutional layer.

RAMS [6] is a variant of the 3D WDSR that introduces an attention mechanism. Specifically, a tensor of  $T$  single-channel LR images is introduced to the proposed model and two branches are followed. The main branch employed 3D convolutional for feature extraction, while the feature attention mechanism focuses on the most promising inner representations. At the same time, a global residual path introduces an attention path for enforcing temporal consistency.

Last, HighRes-Net [4] is a recently proposed deep learning architecture that learns to fuse an arbitrary number of low-resolution frames with implicit co-registration to a reference frame. The model consists of four modules, namely (i) co-registration, (ii) fusion, (iii) up-sampling, and (iv) registration-at-the-loss. The method operates on unregistered sequences by learning a global fusion operator that is applied recursively on an arbitrary number of low-resolution examples. During the training stage only, an additional network, the ShiftNet is trained by learning to align the SR output to a ground-truth example.

## PROPOSED FRAMEWORK

The proposed MFSR system will be composed of two subsystems, the SR module, which will be trained on the ground and eventually deployed on OPS-SAT, and the self-supervision module, which is part of the ground-based training procedure (will not be deployed). **Figure 1** **Error! Reference source not found.** presents the current version of the MFSR system architecture, which is composed of two modules, namely the OPS-SAT onboard component and the GROUND-based component. The proposed systems consist of two modules, one deployed on the spacecraft (inference) and one on the ground (training), where we assume the unrestricted availability of computational resources.

Specifically, The Registration (REG) module accepts as input a sequence of LR images and produces another LR image sequence where images are aligned with respect to a reference frame. The 3D Super Resolution (3D-SR) module accepts

as input a sequence of LR images and produces an HR image through a method based on a 3D Residual Convolutional Neural Network (3D-RCNN) architecture. The Degradation (DEG) module accepts an HR image and produces downgraded LR images, which are generated through a forward image formation model. Finally, we employ the Root Mean Squared error as the loss function.

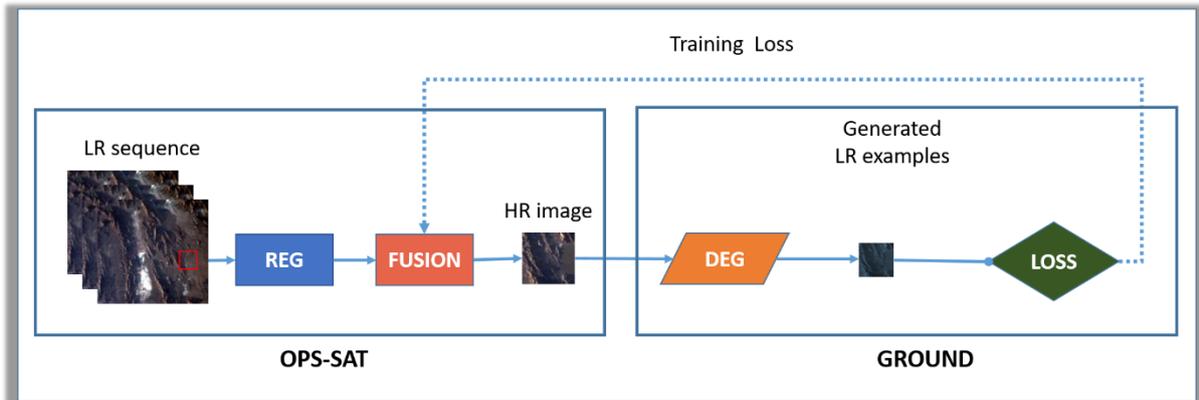


Figure 1: Block diagram of the proposed framework

The onboard module consists of two submodules: frame registration, and frame fusion. For the registration of frames, we assume one frame as the reference frame and register two leading and two following frames concerning the selected one.

We employ the *phase\_cross\_correlation* method which is part of the *skimage* library. The approach is based on an efficient subpixel image registration by cross-correlation on the frequency domain, optimized for reduced memory and processing requirements [7]. We consider integer pixel translation between frames, based on the assumed smooth motion of the satellite in orbit.

For the image fusion component, we consider a Convolutional Neural Network based on the architecture of the 3D Wide-Residual Convolutional Neural Network [5]. The network is composed of a sequence of residual blocks where the input is added to the output both as is and after going through a 3D convolution layer with the ReLU activation function. Furthermore, the weights are normalized through the Weight normalization layer.

Last, for the process of degradation, we consider spatial subsampling following a low pass filtering process to emulate the degradation process.

## OPS-SAT OBSERVATIONS

We consider actual observations acquired from OPS-SAT at different locations. Each frame is approximately 4 Mpixels (2000x2000 pixels) and in each case, a sequence of 15-20 frames in quick succession is acquired. The currently available images exhibit very specific characteristics, directly related to the imaging pipeline of OPS-SAT. In Figure 2, we present some exemplary frames and their associated (single-band) histograms of pixel values (calibrate digital counts).

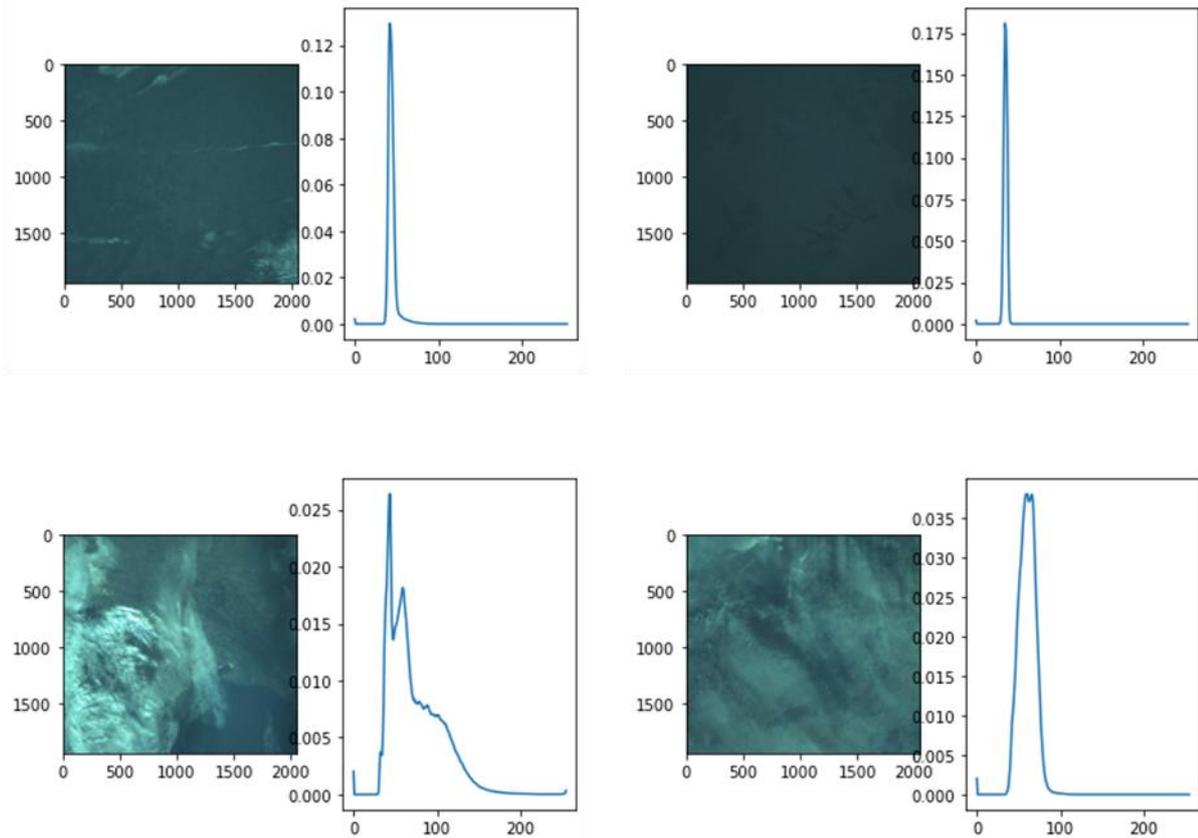


Figure 2: Example images and corresponding pixel value histograms.

These figures indicate that the currently available dataset is composed of images that exhibit a very low dynamic range, typically centered at some relatively low value. To increase the perceptual quality, we apply a contrast-enhancement process through the *exposure.equalize\_hist* function of *skimage* library. In Figure 3, we showcase different examples, of the raw (top row) and the contrast-enhanced (second row) versions of OPS-SAT image sequences.

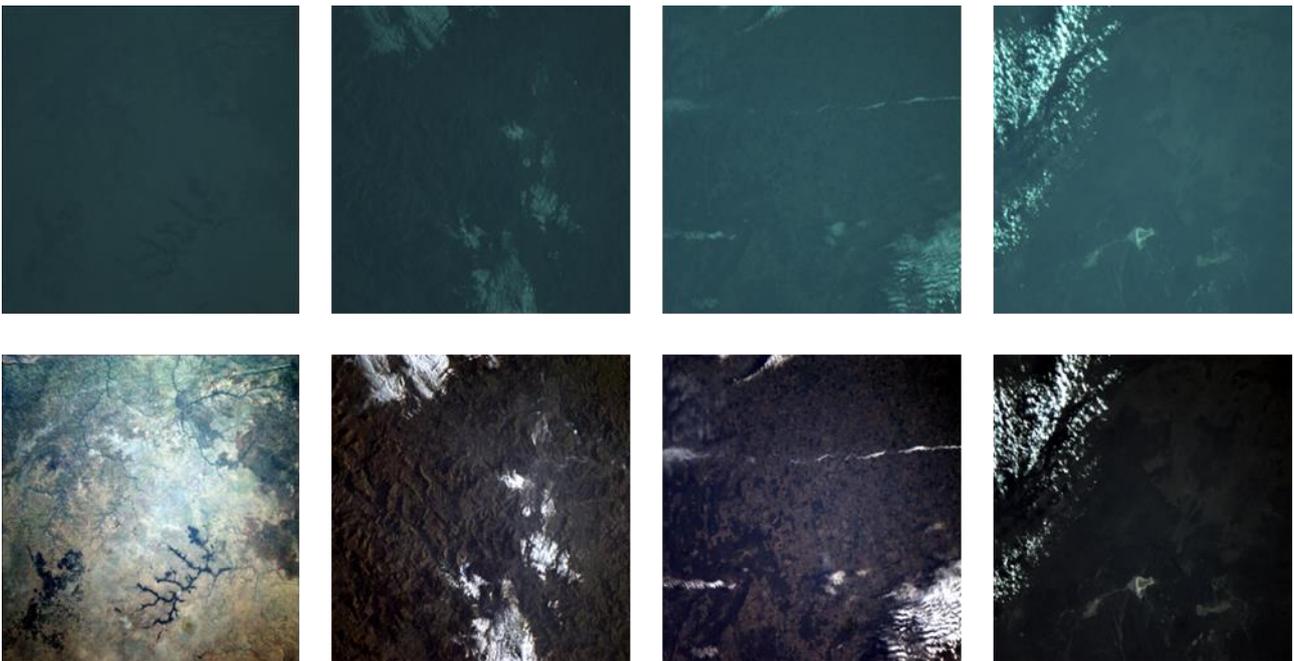


Figure 3: Raw (top) and contrast-enhanced (bottom) images from OPS-SAT

## ON-BOARD VS OFF-BOARD PROCESSING

The proposed MFSR approach is predicated on the application of advanced learning-based methods to enable novel capabilities. More specifically, the underlying assumption of this project is that storing/transmitting a single high-resolution image is “better” than storing/transmitting a sequence of low-resolution frames.

To quantify the performance gains, we consider the amount of data, in terms of Mpixels, that need to be stored and transmitted to the ground. In this scenario, the merit of on-board processing versus off-board/ground processing can be enumerated. The following table considers the specific scenario for OPS-SAT and quantifies the performance gain, where values above 1 indicate that performing the onboard processing is leading to reduced amounts of data.

Table 1: Description on the 5-frame operating point.

|                             |    |             |               |
|-----------------------------|----|-------------|---------------|
| x-axis resolution (Mpixels) | 2  |             |               |
| upscale factor              |    | x2          | x4            |
| Mpixels/frame               | 4  | 16          | 64            |
| # frames                    | 5  | 1           | 1             |
| Total Mpixels               | 20 | 16          | 64            |
| <b>Gain</b>                 |    | <b>1.25</b> | <b>0.3125</b> |

The table above considers two cases of upscale factors, namely x2 and x4 when applied on sequences of 5 frames. The results indicate that a significant benefit is offered in the case of x2, however, for higher upscaling factors, the proposed approach actually suffers a loss in terms of performance. To actually observe performance gains for upscaling factor x4, a sequence of at least 18 frames must be encoded as shown in the following table.

Table 2: Table 1: Description on the 18-frame operating point.

|                             |    |            |              |
|-----------------------------|----|------------|--------------|
| x-axis resolution (Mpixels) | 2  |            |              |
| upscale factor              |    | x2         | x4           |
| Mpixels/frame               | 4  | 16         | 64           |
| #frames                     | 18 | 1          | 1            |
| Total Mpixels               | 72 | 16         | 64           |
| <b>Gain</b>                 |    | <b>4.5</b> | <b>1.125</b> |

## EXPERIMENTAL SETUP AND RESULTS

Unlike the case of PROBA-V where the satellite itself can acquire images with different spatial resolutions based on user input, the OPS-SAT platform acquires images of (approximately) 2000x2000 pixels at a fixed Ground Sampling Distance (GSD). In this case, training a machine learning-based super-resolution module by minimizing the error between predicted and actual HR image is not directly applicable.

Before deploying the onboard module, the 3D CNN must be trained on the ground. While much of the research in this domain was conducted within the scope of the ESA Kelvin competition where LR/HR image pairs from the PROBA-V satellite were utilized, in the case of OPS-SAT such examples are not available.

Specifically, we will consider frames for the available sequences and artificially reduce the spatial resolution to generate the LR samples. The LR samples will be introduced to the training model and the objective will be to estimate high-resolution images which in this case will be the original HR frames, as shown in Figure 4.

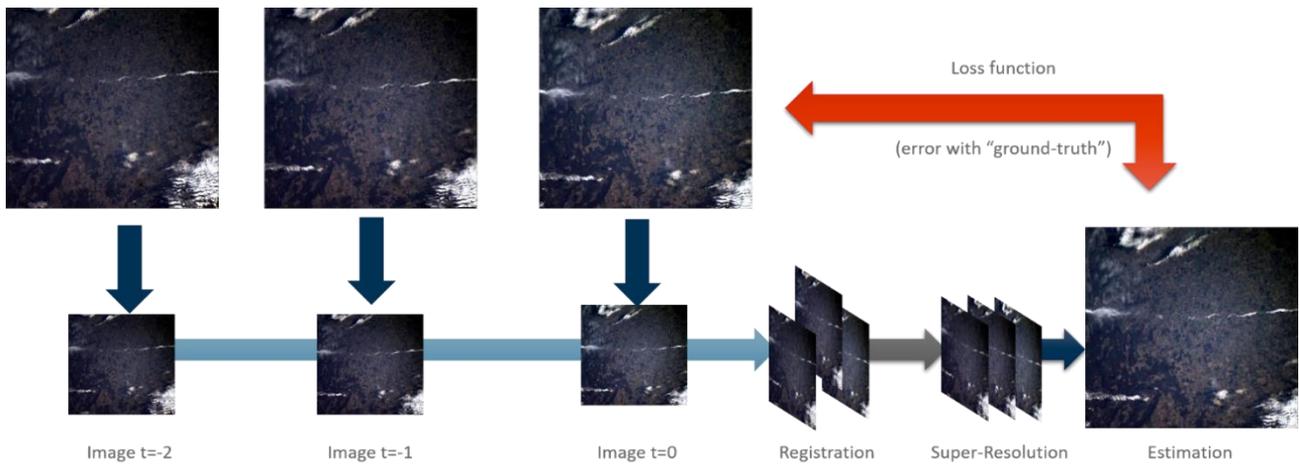


Figure 4: Training process of the proposed approach.

Given the computational and communications resources of OPS-SAT, we will focus on the case of x2 super-resolution for the onboard MFSR. We believe that a super-resolution factor of an x2 factor will be able to demonstrate the potential of the method, minimize the additional requirements, and offers quantified benefits in terms of “effective” compression.

We will consider processing sequences of 5 consecutive low-resolution frames and extracting a single high-resolution one. The initial acquisition parameters are the same as the ones considered during past video acquisition campaigns.

Table 3: Performance on validation examples

| Validation Example | Error Metric | Interpolation | Proposed      |
|--------------------|--------------|---------------|---------------|
| 20220228_102105    | PSNR         | 41.09         | <b>42.93</b>  |
|                    | MSE          | 0.0088        | <b>0.0071</b> |
| 20220124_035501    | PSNR         | 41.45         | <b>43.26</b>  |
|                    | MSE          | 0.0084        | <b>0.0068</b> |

We present indicative experimental results in 5 and Figure 6. In these figures, the first column shows ground-truth images, the second column corresponds to interpolation-based single image super-resolution and the third column correspond to results from the proposed approach. Furthermore, the first row presents unprocessed observation, the second-row images are contrast-enhanced versions of the same images and the third row correspond to the focused images of selected regions.

## DISCUSSION

Our analysis indicates that for specific operating conditions, i.e. relationship between the number of frames and increase in resolution, the proposed approach can reduce the amount of data that needs to be stored and communicated. Furthermore, by performing the proposed approach onboard the spacecraft, the gains in terms of quality (PSNR) are available in-situ, facilitating any subsequent processes like object detection.

## ACKNOWLEDGEMENTS

This work was founded by the OPS-SAT Experiments campaign of the ESA OSIP program and European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 842560.

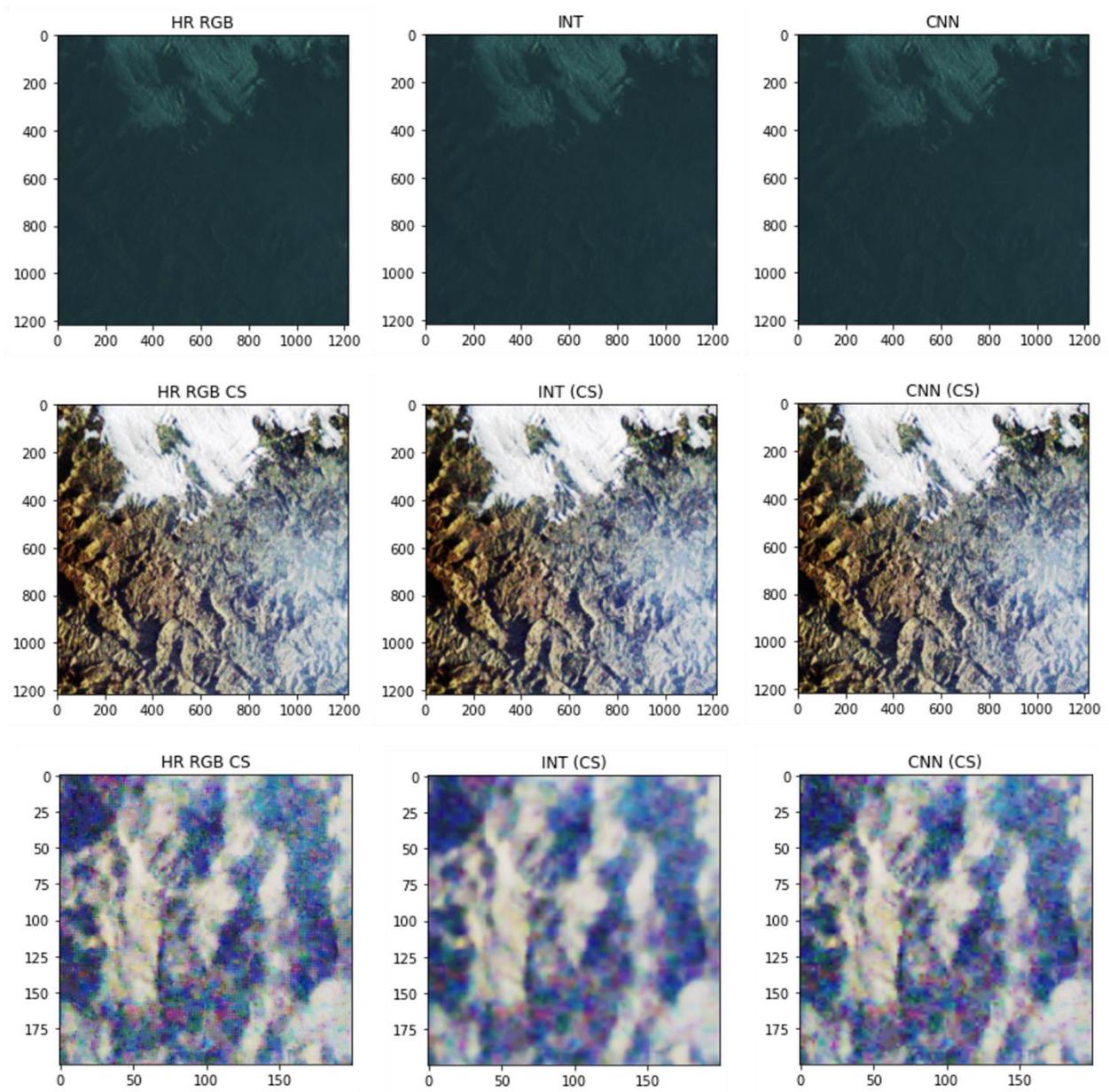


Figure 5: Example of original, interpolated, and recovered images.

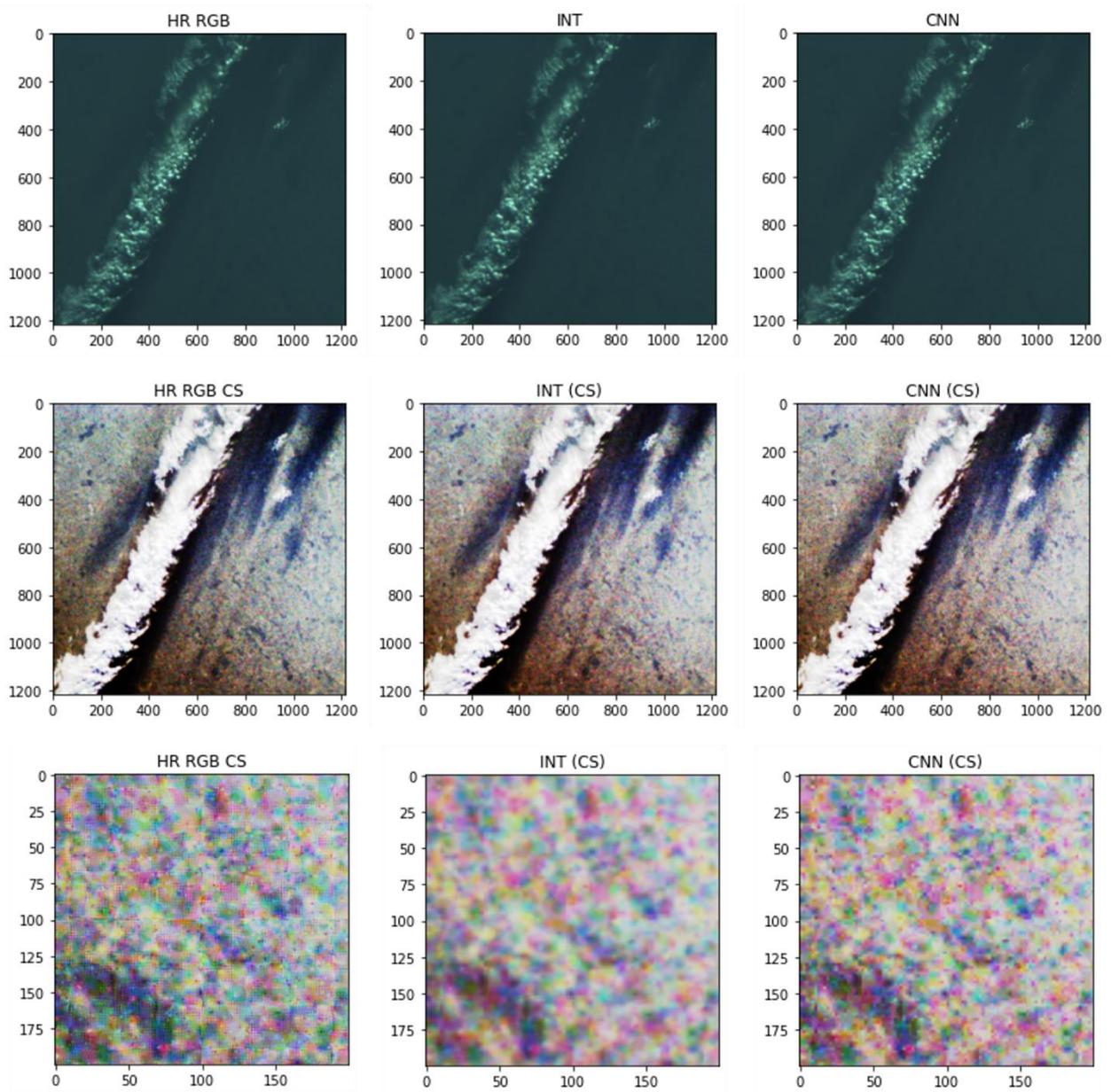


Figure 6: Example of original, interpolated, and recovered images.

## References

- [1] ESA's Kelvin Competition: <https://kelvins.esa.int/proba-v-super-resolution/>
- [2] PROBA-V data used for the competition: <https://kelvins.esa.int/proba-v-super-resolution/data/>
- [3] A. Bordone Molini et al., "DeepSUM: Deep neural network for Super-resolution of Unregistered Multitemporal images"
- [4] M. Deudon et al., "HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery", in arXiv:2002.06460, 2020
- [5] Francisco Dorr, "Satellite Image Multi-Frame Super Resolution Using 3D Wide-Activation Neural Networks", *Remote Sensing*, 12(22), 3812, 2020
- [6] Salvetti, Francesco, et al. "Multi-image super resolution of remotely sensed images using residual attention deep neural networks." *Remote Sensing* 12.14 (2020): 2207.
- [7] Almonacid-Caballer, J., Pardo-Pascual, J. E., & Ruiz, L. A. (2017). Evaluating Fourier cross-correlation sub-pixel registration in Landsat images. *Remote Sensing*, 9(10), 1051.
- [8] Lim, Bee, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. "Enhanced deep residual networks for single image super-resolution." In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 136-144. 2017.