SATELLITE SUPER-RESOLUTION: ON THE ROLE OF MULTI-DATE AND BAND-SHIFT

Anderson Nogueira Cotrim[1], Jérémy Anger[3,4], Gabriele Facciolo[3],
Helio Pedrini[1], Cid Adinam Nogueira Santos[2]

[1]Institute of Computing, University of Campinas, Campinas, SP 13083-852, Brazil
[2]Eldorado Research Institute, Campinas, SP 13083-898, Brazil
[3]Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190, Gif-sur-Yvette, France
[4]Kayrros SAS

***Abstract*—Super-resolution for satellite imagery has gained significant attention due to its potential to improve analytical products without requiring high-resolution imagery. In this work, we aim to study the multi- and single-image super-resolution problem, focusing on the interplay between multi-date information and band-shift characteristics of the imagery. For this study we propose a simulated dataset based on Landsat-8 imagery, aimed at replicating the band-shift properties of Sentinel-2 or their absence, while contemplating 20 multi-date frames for each scene. Our experiments show that the multi-date information can yield gains in both scenarios over single-date super-resolution. Furthermore, we also highlight the diminishing impact of adding frames in both scenarios and the importance of auxiliary frame selection strategies, where temporally closer frames tend to yield better reconstruction results. These findings provide valuable insights into optimizing super-resolution techniques for remote sensing, with implications for high-revisit multi-spectral satellites.**

***Index Terms*—Super-resolution, multi-temporal, multi-date, band-shift, simulated dataset, satellite images**

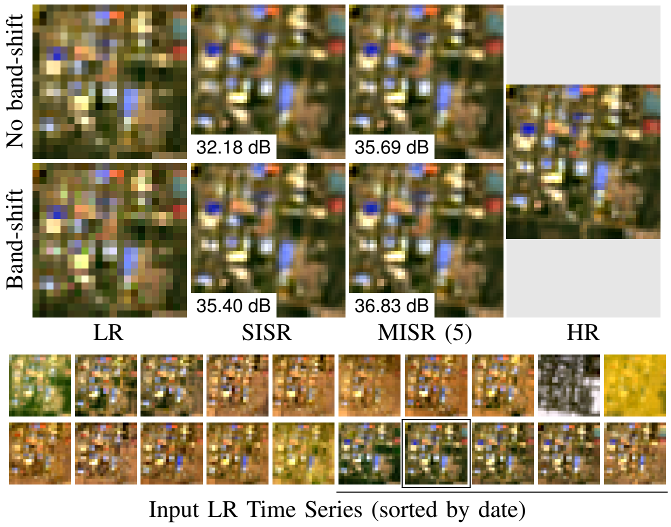

Fig. 1: Multi-Image (5 frames) without band-shift and single-image with band-shift achieve similar super-resolution results. The results of MSIR (20) are omitted as they are similar to the ones shown, having PSNRs 35.63 and 36.52 dB. When available, having both band-shift and multi-frames further improves the results, despite the changes due to seasonal variations, snow, clouds, vegetation, and human activities.

I. Introduction

Satellite imagery plays a key role in land and resource monitoring, but high-resolution (HR) data is often expensive and impractical for frequent analysis of dynamic phenomena.

To address the challenges of acquiring HR data, deep-learning based Super-Resolution (SR) techniques offer a cost-effective alternative, with the added benefit of being applicable to archived data [1–5]. Super-resolution can be categorized into Single-Image Super-Resolution (SISR) and Multi-Image Super-Resolution (MISR). By leveraging multiple aliased frames, MISR techniques have demonstrated superior reconstruction performance over SISR, particularly in general photography [6, 7]. In the field of remote sensing, MISR has been extensively studied, including applications with push-frame satellites [8–12] and in scenarios involving significant temporal gaps between image acquisitions [13–17]. Compared to SISR, MISR may offer a compelling advantage by leveraging the temporal abundance of revisited scenes to reconstruct finer spatial details, especially for satellites with high revisit rates.

In contrast to the MISR trend, Nguyen et al. [18] have demonstrated the importance of the combination of band-shifts and aliasing characteristics for SISR in satellite images. Their findings show that the presence of band-shift in Sentinel-2 imagery provides necessary information to perform super-resolution from a single frame. In this paper we study the interaction between this band-shift characteristic and the information brought by using multi-temporal frames.

To explore these challenges and opportunities, we developed a simulated dataset based on Landsat-8 imagery that approximates the band-shift and aliasing characteristic of Sentinel-

This work was supported by the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES), grant number 88881.981726/2024-01. The authors also would like to thank the Eldorado Research Institute for their support and the National High-Performance Computing Center in São Paulo (CENAPAD) for its computational infrastructure. Centre Borelli is also with Université Paris Cité, SSA and INSERM. This work was performed using HPC resources from GENCI–IDRIS (grants 2024-AD011012472R3 and 2024-AD011012453R3).

2, and provides 20 dates per location. Using this dataset, we conducted experiments training a convolutional neural network (with a permutation invariant architecture based on SBFBurst [19]) in different scenarios, evaluating the potential contributions of multi-date fusion in comparison to band-shift characteristics (Figure 1). The contributions of this paper are:

- A simulated dataset[1] for 2×SR [9, 18] contemplating both multi-date acquisitions and band-shift characteristic.
- A comprehensive study that separately evaluates the contributions of band-shift and multi-date information in reconstructing super-resolved frames, as well as their combined effects.
- An evaluation of the impact of frame selection, including an analysis of the number of frames used during training and inference, and the effect of selecting frames based on acquisition dates.

The choice to work with simulated data is justified by the complexities of real-world datasets. Using cross-sensor imagery for high and low-resolution images involves significant radiometric analysis and careful consideration of the time gap between acquisitions [20]. Simulated datasets offer controlled conditions, enabling targeted investigation of specific variables (e.g., band-shift and multi-date), providing insights into the benefits and limitations of each for real-world applications.

II. Related Work

The PROBA-V super-resolution challenge [21] has become a pivotal benchmark for advancing multi-temporal super-resolution techniques in remote sensing. Numerous methods have been proposed, including HighResNet [13], DeepSUM [14], DeepSUM++ [15], PIUNet [16], and TR-MISR [17]. However, as highlighted by Nguyen et al. [22], the absence of a reference frame was a significant limitation of the challenge, which was addressed in some later studies [9, 10, 23, 24].

Another limitation of the PROBA-V challenge is that it does not extend to multi-spectral pushbroom satellites, where band shifts play a significant role in SISR, as noted by Nguyen et al. [18]. The DeepSent paper [1] addressed the multi-temporal aspect jointly with multi-spectral and also proposed a simulation pipeline [25] for training their model; however, since it used a degraded version of Sentinel-2 imagery, band shifts may not have been fully accounted for in their experiments. In contrast, Okabayashi et al. [3] utilized real Sentinel-2 data, but the absence of a reference frame remains a potential limitation of their study.

While there have been considerable efforts in multi-temporal super-resolution [1, 3, 26, 27], most studies lack a comprehensive evaluation of how multi-date information contributes relative to other key factors in single-image super-resolution, such as band shifts, as emphasized by Nguyen et al. [18].

[1]Simulated dataset is available at: https://doi.org/10.5281/zenodo.15387917

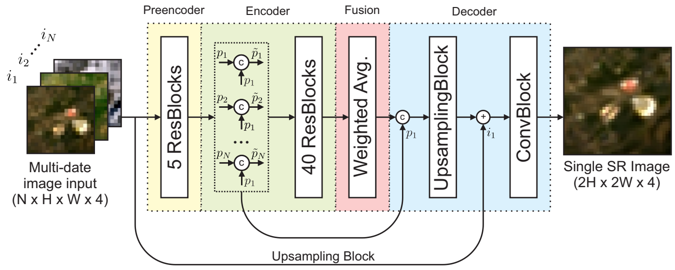

Fig. 2: An abstraction of the adapted SBFBurst for 2×SR, with all alignment-related modules removed. The fusion module was borrowed from [7], while the upsampling block utilizes pixel shuffling [28]. For more details, please refer to [19].

III. Super-resolution Method

For this work, we focus on the architecture of SBFBurst [19] (Figure 2), as it is permutation-invariant, reference-aware [22], and offers flexibility in controlling the number of frames during both training and inference. This design allows us to use different number of frames at inference for a given trained network.

To simplify the analysis, we did not employ optical flow alignment as originally implemented in SBFBurst [19]. Given the registration accuracy of Landsat-8 and the proposed simulation pipeline we ensured that frame misalignment is bounded to the subpixel level. Empirical tests further confirmed that embedding alignment provided no measurable improvement in our specific scenario. For the single-image setting, the architecture is adapted by removing the fusion block.

For the multi-date setting, at inference, the inputs of the network are the images closest in time to the reference. In Section V we perform an ablation to study the impact of this strategy against randomly sampling the time series.

IV. Simulated Dataset

Source imagery. To study the effect of band shift, the number of dates, and the temporal sampling strategy, we have built a synthetic dataset. This dataset is based on the Landsat-8 level 2 collection, with a global world coverage, sampled over 548 locations. For each sampled location, we extracted a time series of the B2, B3, B4 and B5 channels (at 30m/px), for 20 dates with less than 5% cloud coverage on the entire scenes. This leads to a total of 10960 images. Figure 3 shows the cumulative distribution of the number of days between two images, 60% of the data have a gap of less than 32 days (two repeat cycles).

To simulate our ideal images, we start by filtering the Landsat images with a small Gaussian low-pass filter of $\sigma = 0.5$. This shall prevent aliasing in the simulation, which we will re-introduce ourselves from the simulation. Furthermore, we introduce a random subpixel shift (± 0.5px) at each date to simulate small georeferencing errors [29]. The resulting images are considered as ideal, and the simulator is used to derive the ground-truth high-resolution and the low-resolution images used for the study.

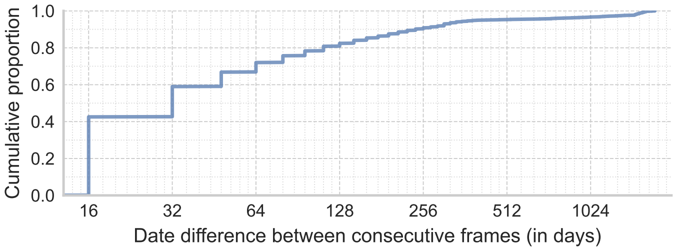

Fig. 3: Cumulative distribution of the difference between the dates (in days) of the simulated dataset.

Simulation. The main characteristic of this dataset resides in the downsampling simulation to provide *band-shift* and *no band-shift* low-resolution (LR) inputs, alongside the high-resolution targets. Let u be the ideal Landsat-8 imagery as described above. The simulation emulates the observation of an image from an inclined orbit, its sampling by a sensor, and the resampling that is typically performed in the ground segment to align the images. This captured by the formula

$$v_s = \text{RECTIFY}(\text{SAMPLE}_s(\text{OBSERVE}_s(u))) + n_s, \quad (1)$$

where s specifies the scenario (HR, *no band-shift* LR, *band-shift* LR), and n is white Gaussian noise of standard deviation $\sigma = 5$ (only for LR scenarios).

The OBSERVE step consists in viewing the scene from a given orbit. Here, we approximate this step by rotating the ideal scene u by an angle of 12 ± 0.5 degrees. The center of rotation is randomly selected from the central part of the image. For *band-shift* LR, each band is affected by slightly different rotation parameters (angle and center). This rotation corresponds to a change of viewpoint, as if the scene was acquired from an orbit that is slightly different at each date and band. While not accurate compared to a real Sentinel-2 acquisition simulator, we found that this simple step was sufficient to replicate some of the per-band aliasing patterns that we can seen on Sentinel-2 imagery, and less limited than the simulation based on translation [18].

The SAMPLE step realizes the sampling of the ideal scene to a fixed resolution grid. If the scenario is HR, then this step is the identity. Otherwise, a 2×2 binning operation is used to reduce the resolution to 60m/px, thus introducing alias in the LR frames. This operation corresponds to the integration of the photons on the sensor. No low-pass filter is applied since we require the optical system of the HR and LR to be identical, as the goal of super-resolution is to go from an aliased image to a well-sampled image.

The RECTIFY step is the inverse of OBSERVE. It undoes the exact rotation that was applied on each band, similar to the orthorectification process that is performed to go from a sensor geometry to a ground geometry. Note that the SAMPLE step is non-invertible when considering an LR scenario, thus the composition of all steps is not equivalent to just applying the SAMPLE operation on the original ideal image.

After simulation, the central part of each image is cropped to 100×100 for the LR sets, and 200×200 for the HR set.

	SISR	MISR		
Number of Frames	1	4	10	20
No band-shift	42.42	45.02	46.00	46.36
Band-shift	45.31	46.73	46.86	46.85
Band-shift gain	+2.89	+1.71	+0.86	+0.49

Table 1: PSNR (dB) results on test set for each trained model tested on according input number of frames. For MISR, we adopted the selection strategy of closest dates.

V. Experiments

A. Experiments Settings

We divided the our dataset into training, validation, and test sets using a 60-20-20 split, ensuring that samples in each set are geographically distinct. We trained our model in different settings depending on the number of frames used at training time and the presence or absence of band-shift in the simulation:

- **Input characteristic**: *band-shift* and *no band-shift.*
- **Training frames**: 1 (SISR), 4, 10 and 20.

In all trainings we used L1 loss as the cost function and optimized until convergence on the validation set. In multi-frame scenarios, auxiliary frames were randomly selected during the training phase. As mentioned earlier, at inference time the inputs of the network are the images closest in time to the reference. For evaluation, we employed the PSNR (dB) metric with a peak reference of 16384.

B. Results and Discussion

Table 1 summarizes the PSNR results over the test set for all the SISR and MISR configurations (with and without band-shift), while using the same number of frames at training and at inference. Additionally, Figure 4 extends the evaluation considering more frames at the inference, and randomly selecting the time series.

The gain of MISR over SISR is significant. From Table 1, we see that the gain of using multiple date images over SISR is significant, when using 20 dates we have a gain of 3.94 dB and 1.54 dB in the *no band-shift* and *band-shift* scenarios respectively. This range of gain is coherent with the literature [1, 26, 27]. From Figure 4, we can see that compared to what band shift information can brings to SISR, MISR *no band-shift* can achieve similar effectiveness, using 8 frames (random strategy) and with approximately 5 frames when using the strategy of selecting the frames with closest dates to the reference frame.

The gain of band-shift. From Table 1, we can see that the impact of band-shift is particularly significant in the SISR configuration [18], yielding a notable gain of 2.89 dB. For MISR, the band-shift gain diminishes as the number of input frames increases, dropping from 1.71 dB (4 frame) to 0.49 dB (20 frames). This can be attributed to the fact that in the *band-shift* setting, a single frame already contains a substantial amount of information for super-resolution. As a result, adding

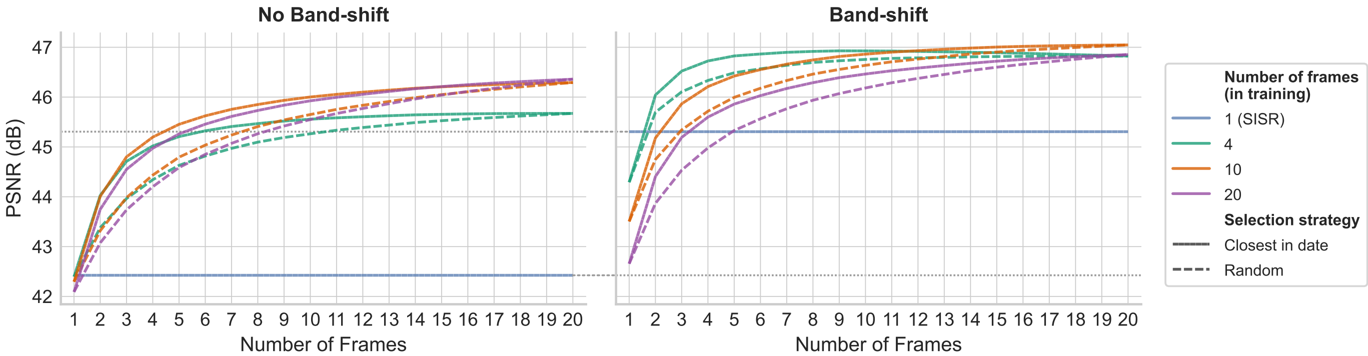

Fig. 4: PSNR vs. input frames for models trained with different numbers of input frames for *band-shift* and *no band-shift* scenarios. Solid lines depict the gain of using the auxiliary frame selection strategy of taking the closest dates over random selection (dashed lines).

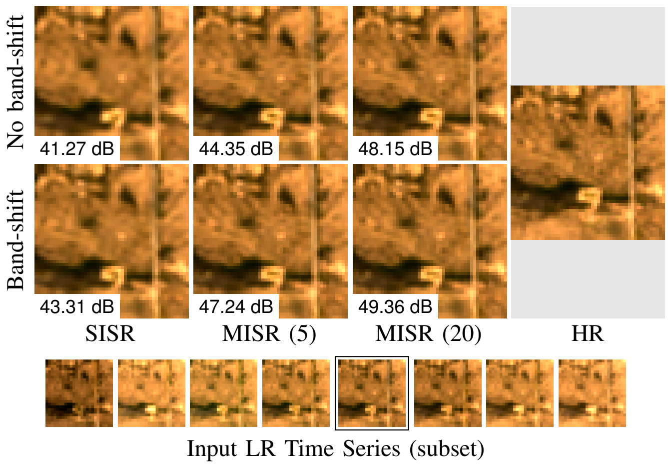

Fig. 5: Examples of SR outputs obtained for each studied scenario where the input time series exhibits minimal structural changes. MISR (5) and MISR (20) refer to model outputs using 5 and 20 input frames during inference, respectively.

more frames does not improve the results as significantly as in the *no band-shift* case.

The MISR gain depends on the scene stability. A challenge in using MISR is the temporal changes in the scene, which are naturally caused by seasonal variations, snow, clouds, vegetation, and human activities. Figure 1 shows examples of reconstructions using 5 frames. Although not displayed, the PSNR results for 20 frames indicate no improvement over those shown. This limited gain may be due to scene changes. Conversely, in Figure 5, we present a more extreme case where the scene is less prone to changes. In this case, a higher MISR gain of up to 7 dB can be observed.

Effect of the number of frames used in training. The number of frames seen during training can influence how the fusion module suppress or prioritizes frames for reconstructing the desired outcome. This can be observed in Figure 4.

In the *no band-shift* scenario, all models exhibit similar performance up to six frames. However, the model trained with four frames shows reduced effectiveness beyond this point, compared to those trained with 10 or 20 frames. Models trained with 10 and 20 frames show comparable performances, with the model trained on 10 frames consistently yielding slightly stronger results.

In *band-shift* scenario, we observe a more pronounced trend where the model trained with 4 frames performs better than the other models in the region below 10 frames. This contrasts with the *no band-shift* experiments, where the model trained with 4 frames under-performs for 6 frames or more. A possible explanation for this could be that since band-shift provides useful information for frame reconstruction, as demonstrated by Nguyen et al. [18], training with more frames might prevent the network from extracting information from this crucial characteristic, which is inherently present in each frame. Indeed, in all cases, the band-shift MISR model taking a single frame performs considerably worse than band-shift SISR, a trend that is not observed or is relatively weak in the *no band-shift* scenario when comparing MISR with the equivalent SISR model. This observation highlights the need for better learning mechanisms to capture the interplay between band-shift and temporal information effectively.

Closest date strategy improves the results. In both cases (*band-shift* and *no band-shift*), we observe a significant improvement when selecting frames with the closest dates to the reference image for inference (Figure 4). This suggests that the smaller the temporal gap between frames, the more information can be leveraged to reconstruct the selected frame, indicating that MISR can benefit even more when satellites have a higher revisit rate. Furthermore, since in this study the LR frames have a resolution of 60m/pixel, this effect could be even more pronounced at higher resolutions.

VI. Conclusions

In this study, we have shown that multi-date SR can lead to significant gains over SISR in both *no band-shift* and *band-shift* scenarios, through a new proposed dataset. Our results emphasize the importance of scene stability and frame selection strategies, with closer temporal proximity enhancing reconstruction quality. Future work will focus on transposing these results to real-world data and on improving the integration of band-shift and temporal data.

REFERENCES

[1] T. Tarasiewicz, J. Nalepa, R. A. Farrugia, G. Valentino, M. Chen, J. A. Briffa, and M. Kawulok, "Multitemporal and Multispectral Data Fusion for Super-Resolution of Sentinel-2 Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–19, 2023.

[2] M. Yang, D. Mao, Y. Zhang, Y. Huang, and J. Yang, "Super-Resolution Method for Synthetic Aperture Radar Image Based on Multi-Scale Feature Extraction," in *International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2024, pp. 3388–3391.

[3] A. Okabayashi, N. Audebert, S. Donike, and C. Pelletier, "Cross-Sensor Super-Resolution of Irregularly Sampled Sentinel-2 Time Series," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2024, pp. 502–511.

[4] N. L. Nguyen, J. Anger, A. Davy, P. Arias, and G. Facciolo, "L1BSR: Exploiting Detector Overlap for Self-Supervised Single-Image Super-Resolution of Sentinel-2 L1B Imagery," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 2012–2022.

[5] A. Ramirez-Jaime, N. Porras-Diaz, G. Arce, D. Harding, M. Stephen, and J. MacKinnon, "Super-Resolution of Satellite LIDARS for Forest Studies Via Generative Adversarial Networks," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2024, pp. 2271–2274.

[6] B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C.-K. Liang, M. Levoy, and P. Milanfar, "Handheld Multi-Frame Super-Resolution," *ACM Transactions on Graphics*, vol. 38, no. 4, Jul. 2019.

[7] G. Bhat, M. Danelljan, L. Van Gool, and R. Timofte, "Deep Burst Super-Resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2021, pp. 9209–9218.

[8] J. Anger, T. Ehret, C. de Franchis, and G. Facciolo, "Fast and Accurate Multi-Frame Super-Resolution of Satellite Images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. V-1–2020, p. 57–64, Aug. 2020.

[9] N. L. Nguyen, J. Anger, A. Davy, P. Arias, and G. Facciolo, "Self-Supervised Multi-Image Super-Resolution for Push-Frame Satellite Images," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Jun. 2021, p. 1121–1131.

[10] ——, "Self-Supervised Super-Resolution for Multi-Exposure Push-Frame Satellites," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2022, p. 1848–1858.

[11] ——, "Self-Supervised Push-Frame Super-Resolution With Detail-Preserving Control And Outlier Detection," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 131–134.

[12] J. Lafenetre, N. L. Nguyen, G. Facciolo, and T. Eboli, "Handheld Burst Super-Resolution Meets Multi-Exposure Satellite Imagery," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023, pp. 2056–2064.

[13] M. Deudon, A. Kalaitzis, I. Goytom, M. R. Arefin, Z. Lin, K. Sankaran, V. Michalski, S. E. Kahou, J. Cornebise, and Y. Bengio, "HighRes-Net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery," pp. 1–15, 2020.

[14] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3644–3656, 2020.

[15] ——, "DeepSUM++: Non-Local Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2020, pp. 609–612.

[16] D. Valsesia and E. Magli, "Permutation Invariance and Uncertainty in Multitemporal Image Super-Resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 1–12, 2022.

[17] T. An, X. Zhang, C. Huo, B. Xue, L. Wang, and C. Pan, "TR-MISR: Multiimage Super-Resolution Based on Feature Fusion With Transformers," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 1373–1388, 2022.

[18] N. L. Nguyen, J. Anger, L. Raad, B. Galerne, and G. Facciolo, "On The Role of Alias and Band-Shift for Sentinel-2 Super-Resolution," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, Jul. 2023, p. 4294–4297.

[19] A. Cotrim, G. Barbosa, C. Santos, and H. Pedrini, "Simple Base Frame Guided Residual Network for RAW Burst Image Super-Resolution," in *19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 3, INSTICC. SciTePress, 2024, pp. 77–87.

[20] J. Michel, E. Kalinicheva, and J. Inglada, "Revisiting Remote Sensing Cross-Sensor Single Image Super-Resolution: The Overlooked Impact of Geometric and Radiometric Distortion," Oct. 2024. [Online]. Available: https://hal.science/hal-04723225

[21] M. Märtens, D. Izzo, A. Krzic, and D. Cox, "Super-Resolution of PROBA-V Images using Convolutional Neural Networks," *Astrodynamics*, vol. 3, no. 4, p. 387–402, Aug. 2019.

[22] N. L. Nguyen, J. Anger, A. Davy, P. Arias, and G. Facciolo, "PROBA-V-REF: Repurposing the PROBA-V Challenge for Reference-Aware Super Resolution," in *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 2021, pp. 3881–3884.

[23] B. Pogodzinski, T. Tarasiewicz, and M. Kawulok, "Multi-Image Fusion for Super-Resolving Individual Sentinel-2 Images," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2024, pp. 1108–1112.

[24] M. Impieri, D. Valsesia, T. Bianchi, and E. Magli, "Degradation-Aware Self-Supervised Multi-Temporal Super-Resolution," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2024, pp. 1099–1102.

[25] T. Tarasiewicz, J. Nalepa, and M. Kawulok, "Semi-Simulated Training Data for Multi-Image Super-Resolution," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2022, pp. 481–484.

[26] P. Wolters, F. Bastani, and A. Kembhavi, "Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing," 2023. [Online]. Available: https://arxiv.org/abs/2311.18082

[27] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Deep Learning For Super-Resolution Of Unregistered Multi-Temporal Satellite Images," in *10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2019, pp. 1–5.

[28] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2016, pp. 1874–1883.

[29] R. Rengarajan, M. Choate, M. N. Hasan, and A. Denevan, "Co-registration Accuracy between Landsat-8 and Sentinel-2 Orthorectified Products," *Remote Sensing of Environment*, vol. 301, p. 113947, 2024.