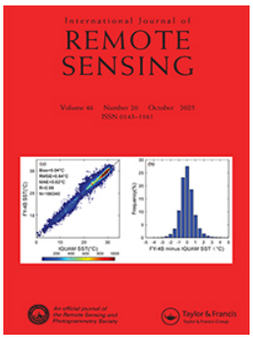

International Journal of Remote Sensing

ISSN: 0143-1161 (Print) 1366-5901 (Online) Journal homepage: www.tandfonline.com/journals/tres20

A holistic approach for multi-spectral Sentinel-2 super-resolution and spectral evaluation

David Major, Zsolt Horváth, Felix Kröber, Hannah Augustin, Martin Sudmanns, Petr Ševčík, Andrea Baraldi, Astrid Berg, Daniel Cornel & Dirk Tiede

To cite this article: David Major, Zsolt Horváth, Felix Kröber, Hannah Augustin, Martin Sudmanns, Petr Ševčík, Andrea Baraldi, Astrid Berg, Daniel Cornel & Dirk Tiede (2025) A holistic approach for multi-spectral Sentinel-2 super-resolution and spectral evaluation, International Journal of Remote Sensing, 46:20, 7437-7464, DOI: 10.1080/01431161.2025.2549132

To link to this article: https://doi.org/10.1080/01431161.2025.2549132

Published online: 07 Oct 2025.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2025, VOL. 46, NO. 20, 7437–7464
https://doi.org/10.1080/01431161.2025.2549132

Taylor & Francis
Taylor & Francis Group

A holistic approach for multi-spectral Sentinel-2 super-resolution and spectral evaluation

David Major [a*], Zsolt Horváth [a*], Felix Kröber [b,c], Hannah Augustin [c], Martin Sudmanns [c], Petr Ševčík [d], Andrea Baraldi [e], Astrid Berg [a], Daniel Cornel [a] and Dirk Tiede [c]

[a]VRVis GmbH, Vienna, Austria; [b]Forschungszentrum Jülich GmbH, Jülich, Germany; [c]Department of Geoinformatics, Paris Lodron Universität Salzburg, Salzburg, Austria; [d]EOX IT Services GmbH, Vienna, Austria; [e]Spatial Services GmbH, Salzburg, Austria

ABSTRACT

Images provided by the European Copernicus Sentinel-2 satellites are valuable and easily accessible sources of remote sensing data for tasks across various fields. These data have a high spectral and temporal resolution, but a rather low spatial resolution, limiting their applicability for many tasks. In agricultural tasks, such as crop monitoring of small land parcels, the use of these data for fine-scale analysis is contingent upon the enhancement of spatial resolution while maintaining spectral fidelity. In this work, we propose a comprehensive single-image super-resolution reconstruction workflow that ensures both properties and is divided into two parts. First, a deep learning-based super-resolution reconstruction approach is applied to improve the spatial resolution of multi-spectral Sentinel-2 images to 2.5 m. For this purpose, a novel method is applied to achieve super-resolution of multiple spectral bands where associated real-word reference data is only partially available. It learns to increase the spatial resolution while preserving spectral accuracy of 10 m bands using high-resolution data from an auxiliary satellite with spectral correspondence, and 20 m bands without reference data using synthetic Sentinel-2 pairs. Second, the suitability of the method to subsequent agricultural tasks is evaluated by measuring the discrepancy between the super-resolved and reference data through a novel spectral knowledge-based validation method. This method leverages mappings of reflectances to spectral categories that enable assessing the spectral fidelity of super-resolved outputs, which is complementary to existing image quality assessment metrics, but with greater depth. The promising spectral validation results suggest that our super-resolution reconstruction pipeline has a great potential for agricultural applications.

ARTICLE HISTORY
Received 31 March 2025
Accepted 11 July 2025

KEYWORDS
Sentinel-2; Multi-spectral; Super-resolution; Deep learning; Spectral validation

1. Introduction

Evaluating the spectral signature of agricultural land in a timely manner has become a crucial task in remote sensing to derive current quality measurements and enable automatic

CONTACT David Major dmajor@vrvis.at VRVis GmbH, Vienna, Austria
*Equal contribution

land monitoring. So far, this has been achieved with aerial and very high-resolution satellite imagery, which is constrained by high costs, limited availability, and limited spectral and temporal resolution. As a cost-effective alternative, remote sensing imagery with lower spatial resolution but high spectral quality and availability such as Sentinel-2 (S2) data is used. However, for regions with small-structured agricultural landscapes like Austria, the ground sampling distance (GSD) of S2 data, 10 m, 20 m or 60 m depending on the spectral band, is insufficient to assess fine-scale features critical for specific agricultural tasks such as land use classification.

Super-resolution reconstruction (SRR) can increase the data's spatial resolution without degrading their temporal and spectral quality to meaningfully employ S2 imagery for such fine-scale applications. The requirement for temporal fidelity excludes the use of multi-image SRR approaches that fuse spectral information of multiple datatakes and prevent analysis of spectral changes over time. An alternative to these approaches is single-image SRR that increases the spatial resolution of a single low-resolution (LR) datatake image using high-resolution (HR) reference data. However, the availability of such reference data is often restricted to Red, Green, Blue (RGB) and Near-Infrared (NIR) bands, whereas agricultural applications require super-resolution across Red Edge (RE) and Short Wavelength Infrared (SWIR) bands. Moreover, the high cost of acquiring HR reference data necessitates a careful selection and compilation of high-quality samples to train deep learning (DL) – based SRR approaches that learn the relationship between LR and HR representations from data.

Once a DL-based SRR model has been trained and is capable of generating HR outputs, a comprehensive and expressive evaluation framework is required to assess the model's applicability to subsequent agricultural tasks. Traditional accuracy metrics such as the peak signal-to-noise ratio (PSNR) and the mean squared error (MSE) primarily focus on intensity differences by measuring Euclidean distances between the spectral signatures of reference data and SRR outputs. However, relevant spectral inconsistencies are not necessarily discoverable by employing aggregative distance metrics. Starting from a given spectral signature, a multivariate displacement vector by a given metric distance x can have different implications depending on its orientation. For the same x, the resulting spectral signature can either (a) still characterize the same land use/cover type, (b) reflect a different land use/cover type or (c) transform the given signature towards a physically implausible signature. These different types of changes allow for different conclusions to be drawn about the SRR capabilities of a model, and they will have different impacts on downstream models, e.g. for land cover or land use classification. Therefore, they should be given different significance in the evaluation of SRR results despite the same distance-metric change. Consequently, we argue that there is a need to explicitly assess the spectral fidelity of SRR outputs and to provide means for interpreting possible systematic biases of SRR outputs semantically.

To address the above challenges, we propose a pipeline that is holistic in nature, as it integrates both a DL-based single-image SRR approach and an extensive spectral validation framework, thereby providing a comprehensive method to assess the SRR-applicability for follow-up agricultural tasks. Our SRR method integrates two complementary convolutional neural networks (CNNs) to increase the spatial resolution of the 10 m and 20 m bands of S2 data to 2.5 m while preserving its high spectral quality. Similar to related works (Galar et al. 2020; Romero, Luis, and Vilaplana 2020) that target SRR between

matching 10 m S2 and HR bands from a different sensor, our first model super-resolves the four 10 m bands and learns the SRR mapping using co-registered HR PlanetScope SuperDove (SD) data (Planet Labs PBC 2024). More than that, we extend the SRR capability of our approach by employing a second model, that is trained to upscale six 20 m bands without associated HR references utilizing synthetic S2 pairs, guided by additional HR input data. Exemplary output of our SRR approach is illustrated in Figure 1. The suitability of the SRR pipeline for subsequent tasks is evaluated by the knowledge-based spectral categorization system called Satellite Image Automatic Mapper (SIAM) (Baraldi 2011). It allows capturing spectral changes between reference and reconstructed data according to predefined, human-interpretable categories, and is capable of augmenting the semantic depth of existing image reconstruction quality metrics. We present promising results utilizing the SIAM system, which indicates that our SRR pipeline has a high potential for subsequent agricultural applications.

Our *contributions* can be summarized as follows:

- We introduce a novel combination of two CNN-based single-image SRR models for enhancing the spatial resolution of ten S2 bands with GSDs of 10 m representing RGB +NIR and 20 m displaying RE+SWIR spectra. This combination captures LR and HR relationships from both real-word and simulated pairs, providing a greater alignment with observed data than methods trained entirely on artificially generated pairs. More specifically, our approach learns to super-resolve RGB+NIR bands using a limited amount of HR SD reference data. Furthermore, the RE+SWIR bands' SRR mapping is learned from synthetic S2 data, and their spatial structure is improved by HR RGB+NIR images as additional input.
- We demonstrate how our knowledge-based, interpretable evaluation approach substantially expands the informative value of existing metrics for image reconstruction quality. Leveraging the SIAM framework, we gain systematic qualitative

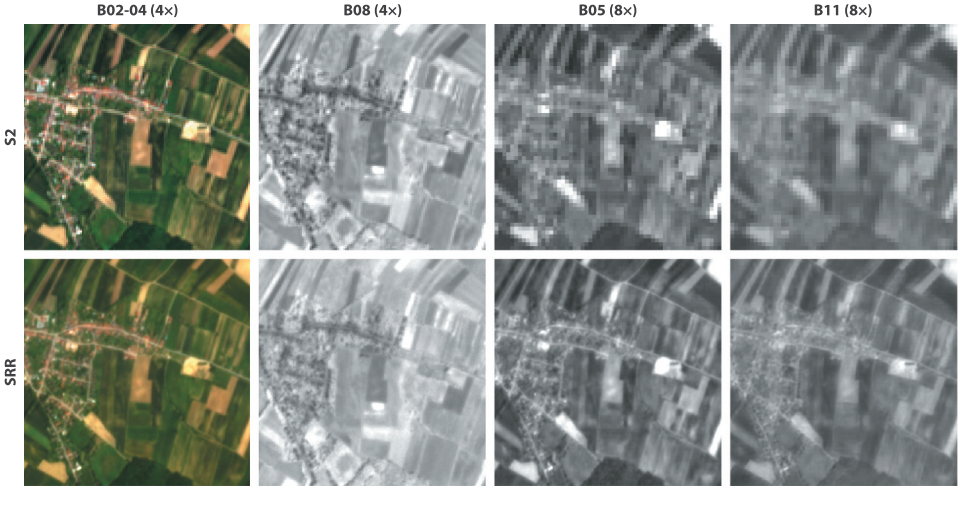

Figure 1. Exemplary output from our SRR approach (bottom row) compared to the S2 input (top row). The output spatial resolution is 2.5 m for all 10 m and 20 m bands, increasing the spatial resolutions of input bands B02, B03, B04 (RGB) and B08 (NIR) by a factor of 4 and of all RE and SWIR bands by a factor of 8.

as well as quantitative insights into the spectral fidelity of the SRR outputs. Analyses are performed at various levels of aggregation, showcasing the flexibility of our framework to derive high-level summaries as well as in-depth evaluations of the reconstruction quality.

2. Related work

2.1. Remote sensing super-resolution reconstruction

SRR aims to reconstruct a fine-grained HR image from an LR image with coarse details (Anwar, Khan, and Barnes 2020). Classical approaches (Yang et al. 2010; H. Zhang and Huang 2011) were recently superseded by DL methods (Dong et al. 2016; Kim, Kwon Lee, and Mu Lee 2016; Lim et al. 2017) due to various reasons. DL-based approaches automatically learn meaningful features from data as opposed to hand-crafted features designed by experts for classical methods. Moreover, DL-based approaches are capable of achieving higher performance by leveraging more advanced algorithms in combination with deeper network architectures and have become computationally efficient through the explosion of Graphical Processing Unit (GPU) availability (Tsagkatakis et al. 2019). In the context of remote sensing, single-image SRR maps an LR image of a single datatake to its HR version, whereas within multi-image SRR, LR images of multiple datatakes are mapped to a single or multiple HR-reconstructed versions (Razzak et al. 2023). In addition, satellite images appear in different resolutions covering multiple spectral bands, most commonly RGB and NIR followed by bands capturing higher wavelengths. In light of the aforementioned benefits, this work addresses a DL-based single-image multi-spectral SRR approach, and related work is discussed in detail below.

2.1.1. Single-image super-resolution

One potential approach for training a super-resolution DL model is to utilize a supervised setup with co-registered LR/HR images of analogous spectral bands from satellites with different image resolutions. Galar et al. (2020) utilize co-registered S2 and PlanetScope (PS) imagery to learn the SRR mapping from a GSD of 10 m to 5 m and 2.5 m of RGB and NIR bands and apply enhanced deep residual networks (EDSRs) (Lim et al. 2017) as a DL model. Romero, Luis, and Vilaplana (2020) propose a method using a generative adversarial network (GAN) to super-resolve RGB and NIR bands of S2 images by a factor of 5 utilizing reference data from the Worldview satellites. They employ the architecture of the enhanced super-resolution generative adversarial network (ESRGAN) (X. Wang et al. 2019) and outperform related SRR methods using standard image quality assessment (IQA)-metrics. Tao et al. (2021) follow a similar GAN-based approach to enhance RGB optical imagery from the ExoMars Trace Gas Orbiter by a resolution magnification factor of 4. They use images from the Mars Reconnaissance Orbiter for reference and present a modified SRR model that contains architectural building blocks of the ESRGAN (X. Wang et al. 2019) and the ESRGAN+ (Rakotonirina and Rasoanaivo 2020) in order to achieve high-quality SRR results. Wolters, Bastani, and Kembhavi (2023) compare the performance of DL-based methods for 4$\times$ SRR of S2 images using HR aerial NAIP images. They mainly investigate multi-temporal RGB S2 samples as input, but pass them aggregated to well-established single-image approaches such as SRCNN (Dong et al. 2016) and ESRGAN (X. Wang et al. 2019). The performance of

these models is compared to a novel diffusion-based technique called SR3 (Ho, Jain, and Abbeel 2020). Moreover, they evaluate the SRR outputs by downstream tasks and conclude that both ESRGAN and SR3 generate higher-quality images than the other methods, but the former achieves better accuracy for the downstream task and is better at avoiding incorrect or misleading results.

2.1.2. Multi-spectral super-resolution

Images captured by specific satellites offer a rich multi-spectral resource of information, such as the thirteen bands provided by the S2 satellites of the European Copernicus programme. However, such spectral abundance is not common in satellite imagery, making the problem of supervised SRR infeasible for certain spectral bands. Therefore, recent research focuses on methods where SRR is solved partially without access to LR/HR image pairs. The DSen2 approach of Lanaras et al. (2018) super-resolves the S2 bands having 20 m and 60 m GSD to the RGB band resolution of 10 m using a CNN with residual blocks. They create synthetic LR/HR pairs for training where original images are downscaled, train the SRR model with the generated pairs and apply the model to the original images assuming self-similarity of relations between bands of different resolutions within a certain scale-range. Salgueiro, Marcello, and Vilaplana (2021) propose an extension to the above method by modifying the CNN architecture. Vasilescu, Datcu, and Faur (2023) extend the same method by adding a consistency and a synthesis term to the loss to ensure that the degraded version of the model output matches the input and vice versa.

Similar to our approach, there are methods that address both, super-resolving by factors of up to 8 and focusing on SRR of the majority of spectral bands (not just RGB and NIR), but with different configurations and architectures compared to our work. Gupta, Mishra, and Zhang (2024) present a method for multi-spectral fusion and SRR of S2 images with 10 m and 20 m GSD. Their network architecture consists of two encoder blocks, one for each resolution, a feature fusion module, a generative latent bank captured by a StyleGAN (Karras et al. 2020), and a pixel attention-guided decoder block at the end. Unlike our pipeline, which super-resolves images 4× and 8×, their SRR method is designed for magnification factors of 2 and 4 and is trained by the SEN2VENμS dataset (Michel et al. 2022). Tarasiewicz et al. (2023) extract data along the temporal dimension in addition and apply multi-spectral fusion to S2 data. First, data are fused along the temporal axis, followed by fusion of the spectral bands, and ultimately, the data are super-resolved by convolution and pixel shuffling layers. Their DeepSent pipeline super-resolves images by factors of 3, 6 and 18, and in contrast to our work, it is trained completely by simulated LR/HR S2 image pairs. The method for joint SRR of S2 10 m and 20 m GSD spectral bands introduced by Latte and Lejeune (2020) is the closest to our work. Their SRR network achieves a 4× and 8× upscaling depending on the resolution. It is trained by synthetic LR/HR S2 pairs and receives four 10 m GSD bands and six 20 m GSD bands from S2 data as input. It is then transformed and upsampled by standard and sub-pixel convolutions and finally fused with PS images. The output is processed by four further residual dense blocks to obtain the final SRR results. Contrary to our work, their model expects images from both S2 and PS as input for inference.

2.2. Evaluation of SRR models

Different approaches can be distinguished for evaluating SRR models (P. Wang, Bayram, and Sertel 2022; Z. Wang, Chen, and Hoi 2020), all falling under the broader category of IQA. The first and, so far, dominant branch of methods focuses on evaluating SRR performance in a full-reference scenario, meaning with the use of a reference image. Commonly used metrics in this context are MSE and PSNR. Both measure pixel-wise differences between the SRR image and the reference image. Despite their simplicity and widespread use, these pixel-wise, distance-metric-based measures have been criticized for their poor correspondence with human perception and their sensitivity to small spatial translations (Z. Wang and Bovik 2009). Working towards more perceptually motivated distances, the similarity index measure (SSIM) (Z. Wang et al. 2004) leverages statistics computed within a moving window to compare images in terms of luminance, contrast and structural patterns. Furthermore, DL-based metrics, using latent space representations of pre-trained networks, have been proposed as effective perceptual metrics. The learned perceptual image patch similarity (LPIPS) (R. Zhang et al. 2018a) is one of them, highlighting the increasing importance of perceptual quality assessments (Blau et al. 2019). However, all of the previously mentioned metrics lack a focus on spectral reconstruction capabilities relevant and specific to the remote sensing domain.

Only recently have quality assessment frameworks been tailored to remote sensing, going beyond the current set of general-purpose metrics. Aybar et al. (2024) propose a protocol focused on assessing consistency, synthesis and correctness properties of SRR images. In the realm of consistency, they differentiate between metrics for reflectance, spectral and spatial consistency. Even so, employing MSE for reflectance consistency and spectral angle distance for spectral consistency does not acknowledge domain-specific expertise on the physical meaning of spectral signatures.

The complex relationship between phenomena observed on Earth's surface and their band-specific reflectance is not captured by measuring distances and angles uniformly across bands. Since the meaning and validity of a spectral signature mainly depends on its overall shape, inter-band relationships independent of absolute reflectances should be taken into account for assessing multi-spectral SRR capabilities. Additionally, means to characterize the observed reflectance deviations semantically are largely missing with the current set of metrics. A mere quantification of differences itself does not provide any further means of describing and categorizing the nature of these differences.

The second branch of methods evaluates SRR models indirectly via a downstream task. In the field of remote sensing, Razzak et al. (2023) and Chen et al. (2023) assess the effect of SRR models on the task of building footprint extraction, Li et al. (2021) and Xie et al. (2022) investigate the potential added value for land cover classification, and Kawulok et al. (2024) evaluate hyper-spectral SRR for the prediction of soil nitrogen content, water quality and air pollution. While these evaluation approaches are more orientated towards assessing the practical value of the SRR data, they are inherently very specific to single applications. This makes the downstream approach poorly suited as a general means for comparing the performance of different SRR models. One would need an extensive benchmark with a set of diverse downstream tasks to obtain robust evidence for the overall performance of an SRR model. In addition, standardized models would have to be kept available in such a benchmark

and these would also have to be calibrated using reference data, as systematic biases could not be uncovered if a downstream model was retrained only using the SRR data. Finally, it should be noted that downstream model-based evaluations again provide limited possibilities for characterizing the flaws of a given SRR model. High-level descriptions of SRR behaviour on downstream tasks may be insightful for the corresponding application domain, but offer few starting points for diagnosing the original SRR model and its spectral reconstruction capabilities. A knowledge-based, spectral categorization approach, as a preliminary stage of a downstream classification task, offers an adequate substitute to address the described weaknesses and to evaluate an SRR model with regard to its downstream capacities while remaining application-independent.

2.3. Knowledge-based spectral categorisation of remote sensing imagery

Classification of remote sensing imagery based on expert knowledge of spectral signatures is a long-standing field (Ton, Sticklen, and Jain 1991; Wharton 1987). A relevant driver of knowledge-based approaches was the emergence of Object-Based Image Analysis (OBIA) (Blaschke 2010) that extended previously prevailing approaches with a focus on pixel-based spectral characteristics by fostering the inclusion of knowledge on spatial and contextual object properties. In practice, it is common to encode any remote-sensing-specific knowledge about entities of interest into deterministic rule-sets representing knowledge bases. However, such knowledge bases are usually unsuitable for the purposes of general SRR evaluation, as they are often application-specific. They usually lack a mutually exclusive and collectively exhaustive classification design, or they have an unbalanced semantic depth of the different classes, orientated towards the specifics of the respective task. Moreover, many knowledge-based approaches are sensor-specific. This amounts to a lack of transferability of classification rules between different sensors, which is an inherent necessity when evaluating SRR.

With SIAM, Baraldi (2011) introduces an algorithm that generates application-agnostic categorizations of multi-spectral data in a fully automated manner using a knowledge-based decision tree. Operating as a point-wise operator, it maps reflectances into a discrete and finite vocabulary of semi-symbolic spectral categories. The categories' symbolization is indicative of the colour perceived by the human eye. The categories' descriptions capture definite properties of their specific underlying spectral signature in words. This makes the reflectances interpretable by humans. SIAM was validated on spatial data at a continental-scale (Baraldi et al. 2018a, 2018b) and has been leveraged in a variety of knowledge-based downstream systems. This includes prototypical implementations (Arvor et al. 2021) and operational semantic data cubes (Sudmanns et al. 2021). Comparable knowledge-based categorization systems with multi-sensor capabilities, like IMPACT (Simonetti, Marelli, and Eva 2015), are less established, and their outputs are more granular regarding the number of categories.

We integrate SIAM into our SRR evaluation framework in order to retain the flexibility and applicability of task-independent evaluations (as described in Section 2.2) while providing more extensive, expert knowledge-grounded insights into SRR model performance. This not only benefits model benchmarking and diagnostics, but can also be used to derive insights into expected success when using the SRR model in different

downstream domains. In terms of evaluation scope, the proposed approach can therefore be understood to stand between both of the branches of current evaluation methods – the full reference, task-independent approach and the downstream evaluation as detailed in the previous section. In terms of evaluation depth and foundation, the proposed approach complements current methods by introducing a novel, remote-sensing-specific, knowledge-based evaluation process.

3. Super-resolution reconstruction

In this section, we give an overview of our approach to single-image multi-spectral SRR of S2 data super-resolving bands with both 10 m and 20 m GSDs. We start with the preparation of input data required for training in Section 3.1, continue with the description of building blocks of our multi-spectral SRR pipeline (from Section 3.2 to Section 3.4) and present the implementation details in Section 3.5.

3.1. Input data

Our technique is designed to super-resolve LR data in the form of Sentinel-2 (S2) Level-2A imagery without the need for HR reference data as input during inference. This is accomplished in two separate passes for the 10 m and 20 m bands, as explained below. The reason for this is that the SRR of S2 10 m bands can be guided by very high-resolution data provided by PlanetScope SuperDove (SD) Level-3B data with a GSD of 3 m. In order to serve as a reference for SRR training, the LR and HR data require a very high degree of spectral, spatial and temporal correspondence.

3.1.1. Spectral correspondence

The spectral bands present in both S2 and SD data are listed in Table 1. Of the eight SD bands, six correspond to S2 bands, but we disregard the S2 aerosol band because of its GSD of 60 m and its limited use for land cover assessment. In terms of wavelength range, the Blue, Green and Red bands correspond sufficiently well between the two imagery products. The NIR band of SD is much narrower than the S2 B08 and corresponds more

Table 1. Overview of spectral bands of the input data and exploited correspondences between products (blue).

Name	S2 (Range, GSD)	SD (Range)
Aerosol	B01 (433 – 453 nm, 60 m)	B1 (431 – 452 nm)
Blue	B02 (458 – 523 nm, 10 m)	B2 (465 – 515 nm)
Green I	–	B3 (513 – 549 nm)
Green	B03 (543 – 578 nm, 10 m)	B4 (547 – 583 nm)
Yellow	–	B5 (600 – 620 nm)
Red	B04 (650 – 680 nm, 10 m)	B6 (650 – 680 nm)
Red Edge	B05 (698 – 713 nm, 20 m)	B7 (697 – 713 nm)
Red Edge	B06 (733 – 748 nm, 20 m)	–
Red Edge	B07 (773 – 793 nm, 20 m)	–
NIR	B08 (785 – 900 nm, 10 m)	–
Narrow NIR	B8A (855 – 875 nm, 20 m)	B8 (845 – 885 nm)
Water Vapor	B09 (935 – 955 nm, 60 m)	–
SWIR/Cirrus	B10 (1360 – 1390 nm, 60 m)	–
SWIR1	B11 (1565 – 1655 nm, 20 m)	–
SWIR2	B12 (2100 – 2280 nm, 20 m)	–

closely to S2's Narrow NIR band B8A, which has a GSD of 20 m. This means that the important S2 10 m band B08 does not have a perfectly corresponding band in SD, but since the wavelength range of SD B8 is completely contained within S2 B08, we still use SD B8 as reference data for the SRR of S2 B08, as indicated with blue colouring in Table 1. Besides B8, B7 is the only other SD band corresponding to a 20 m band of S2. While it would be possible to incorporate this band as reference data for S2 B05, we would need to treat this band differently from all other S2 20 m bands that do not have corresponding HR data. Instead, we disregard the SD band B7 for supervised training and opt for a unified approach across all S2 20 m bands by using synthetic data.

3.1.2. Spatial correspondence

For training of the RGB+NIR SRR, we focus on two areas of interest of size 24 × 24 km in Austria (see Figure 2), coinciding with the SD grid cells *33N_17E-215N* and *33N_23E-222N*, which are contained completely within S2 granules *33TVM* and *33UWP*, respectively. The first area of interest encompasses an alpine region characterized by forests and pastures, while the second area of interest is predominantly composed of flat agricultural land, as illustrated by the example in Figure 1. Visual inspection revealed that within these regions, geographic features between the two imagery products are not always perfectly aligned, preventing a correct input pairing for supervised training. This is why we apply a non-linear co-registration of the SD data onto the S2 true-colour image using the AROSICS (Scheffler et al. 2017) approach integrated in the *Coregistration* plugin for QGIS (Llano 2021). The co-registration already introduces a resampling step with nearest-neighbour interpolation, so, similar to Galar et al. (2020), we decided to also change the spatial resolution of the SD data from a GSD of 3 m to 2.5 m, which we define as the target resolution for our SRR approach. This enables a more convenient SRR of the S2 10 m and 20 m bands with an integer magnification factor of 4 and 8, respectively.

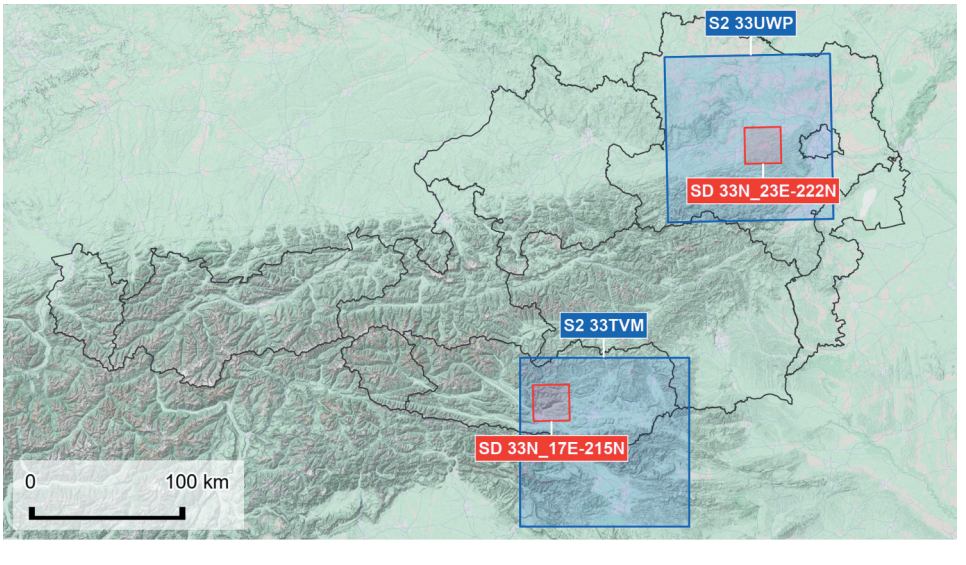

Figure 2. Map of the two areas of interest (red) within Austria, coinciding with SD granules and contained within the S2 granules (blue).

3.1.3. Temporal correspondence

We maximize correspondence between LR/HR training data pairs by considering only granules of datatakes at the same date during the year 2021. This minimizes the change in soil and vegetation conditions between the two imagery products. Furthermore, we exclude all pairs of granules that exhibit a cloud or snow cover of more than 10% within the areas of interest using the semantic EO data cube approach (Augustin et al. 2019). This leaves 21 and 34 pairs of granules of the highest quality for the two areas of interest.

3.1.4. Patch subdivision

For patch-wise training, we separate the selected image data into individual bands and subdivide them into patches of 1280 m $\times$ 1280 m size, i.e. 512 px $\times$ 512 px for all SD bands, 128 px $\times$ 128 px for S2 10 m bands and 64 px $\times$ 64 px for S2 20 m bands. In order to make the training invariant to the placement of patch borders, we introduce an overlap of 160 m, i.e. 64 px for SD, 16 px for S2 10 m and 8 px for S2 20 m, between neighbouring patches on all sides. Finally, we mark patches as invalid and exclude them if any band of the corresponding pair of granules exhibits more than 1% null values, which happens mostly at the borders of datastrips. In total, we obtain 11,463 valid patches of corresponding S2 and SD data for SRR of S2 10 m bands. This number is relatively low and is due to the costly procurement of SD data, but the training data exhibits high quality resulting from their careful selection. As the SRR of S2 20 m bands does not require HR data, we are not limited to the area of the two defined SD grid cells, but can use the entire data of the S2 granules. Performing the same patch subdivision results in a second dataset containing 288,647 patches.

The patches of both sets are split into training, validation and test sets with a ratio of 8:1:1. This means that there are 8988 training, 1240 validation and 1235 test pairs of S2 and SD patches for SRR of the 10 m S2 bands. For the SRR of the 20 m S2 bands, 230,722 training, 28,892 validation and 29,033 test patches are considered. We follow a random sampling strategy for assigning patches to each of the groups, aiming to maximize heterogeneity of surface representation in each set and to avoid spatial and temporal overlaps between them. The assignment algorithm starts by selecting a random patch and all direct neighbour patches. It then collects corresponding valid patches that do not contain artefacts such as clouds or snow across all datatakes of the dataset. The samples are assigned to the appropriate group and the process is repeated until the target split ratio is reached for all groups – training, validation and test. The sampling process is depicted in Figure 3.

3.2. SRR of RGB+NIR bands

Utilizing the RGB and NIR images of SD $A = \{\mathbf{B2}_{SD}, \mathbf{B4}_{SD}, \mathbf{B6}_{SD}, \mathbf{B8}_{SD}\}$ with a GSD of 2.5 m as references, our first step super-resolves the corresponding 10 m GSD spectral images of S2 $B = \{\mathbf{B02}_{S2}, \mathbf{B03}_{S2}, \mathbf{B04}_{S2}, \mathbf{B08}_{S2}\}$ by a factor of 4. Further, let $\mathbf{y}_A \in \mathbb{R}^{W\times H\times 4}$ and $\mathbf{y}_B \in \mathbb{R}^{\frac{W}{4}\times\frac{H}{4}\times 4}$ be the observed intensities of all bands contained in sets A, B. Hence, we are seeking the function

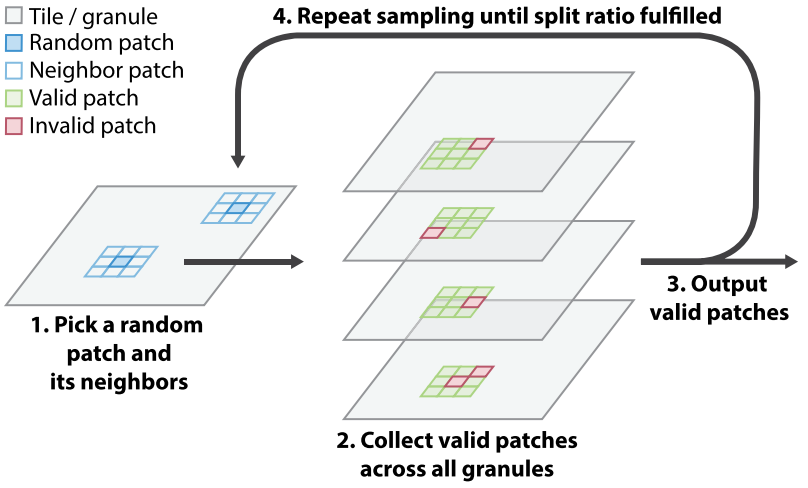

Figure 3. Overview of sampling valid patches that are distributed into training, validation and test sets.

$$\mathcal{F}_{4\times}: \begin{array}{l} \mathbb{R}^{\frac{W}{4}\times\frac{H}{4}\times 4} \rightarrow \mathbb{R}^{W\times H\times 4}, \\ \mathbf{y}_B \mapsto \mathbf{x}_B, \end{array} \tag{1}$$

where $\mathbf{x}_B \in \mathbb{R}^{W\times H\times 4}$ denotes the super-resolved bands of B with 2.5 m spatial resolution. Modelling the SRR function $\mathcal{F}_{4\times}$ of Equation 1 is straightforward by CNN architectures. As mentioned by Y. Zhang et al. (2018b), residual and dense connections help within CNNs to make use of hierarchical feature handling from the LR image to the HR output. Residual connections support feature reuse, while dense blocks are inserted for feature exploration. For these reasons, we employ a SRR network architecture containing residual-in-residual dense blocks (RRDBs) followed by upsampling and convolution layers proposed by Rakotonirina and Rasoanaivo (2020). Due to the excellent satellite image SRR performance of GANs (Romero, Luis, and Vilaplana 2020; Wolters, Bastani, and Kembhavi 2023), we follow the work of X. Wang et al. (2019) and Rakotonirina and Rasoanaivo (2020) and obtain our RGB+NIR SRR model using the ESRGAN+ architecture, consisting of a generator and a discriminator network that are trained jointly to achieve high-quality image reconstruction. Consequently, the SRR loss function $\mathcal{L}_{SRR}$ consists of a perceptual dissimilarity term $\mathcal{L}_{Perc}$, a GAN term $\mathcal{L}_{SRR}^{Ra}$ (see Equation 3), and a pixel-wise mean absolute error (MAE) $\mathcal{L}_{Pix}$. The loss function is formulated by

$$\mathcal{L}_{SRR} = \mathcal{L}_{Perc} + \lambda\mathcal{L}_{SRR}^{Ra} + \gamma\mathcal{L}_{Pix}. \tag{2}$$

The contribution of the loss terms in Equation 2 is controlled by the scaling factors λ and γ. To obtain $\mathcal{L}_{Perc}$, the MAE is computed between feature outputs of the SRR result and the HR image using a pretrained CNN like the VGG model (Simonyan, Zisserman, and Zisserman 2014).

Following X. Wang et al. (2019), a relativistic average discriminator $\mathcal{D}_{Ra}$ is utilized to compute $\mathcal{L}_{SRR}^{Ra}$. It estimates the probability of a real (reference) image being considered relatively more realistic than fake (SRR) images by $\mathcal{D}_{Ra}(\mathbf{y}_A, \mathbf{x}_B) = \sigma(\mathcal{C}(\mathbf{y}_A) - \mathbb{E}_{\mathbf{x}_B\sim\mathcal{X}_B}[\mathcal{C}(\mathbf{x}_B)])$ and vice versa. $\mathbf{y}_A$ and $\mathbf{x}_B$ denote our SD reference (real) and SRR (fake) images with their

respective sets $\mathcal{Y}_A$ and $\mathcal{X}_B$ the sets of our reference and SRR images, $\sigma(.)$ the sigmoid activation function, $\mathcal{C}(.)$ the linear output, and $\mathbb{E}_{\mathbf{y}_A \sim \mathcal{Y}_A}$ and $\mathbb{E}_{\mathbf{x}_B \sim \mathcal{X}_B}$ the expectations over reference and SRR images. Finally, $\mathcal{L}_{SRR}^{Ra}$ is obtained by the formula

$$\mathcal{L}_{SRR}^{Ra} = -\mathbb{E}_{\mathbf{y}_A \sim \mathcal{Y}_A}[\log(1 - \mathcal{D}_{Ra}(\mathbf{y}_A, \mathbf{x}_B))] - \mathbb{E}_{\mathbf{x}_B \sim \mathcal{X}_B}[\log(\mathcal{D}_{Ra}(\mathbf{x}_B, \mathbf{y}_A))]. \tag{3}$$

Besides $\mathcal{L}_{SRR}$, the adversarial version of $\mathcal{L}_{SRR}^{Ra}$ denoted by $\mathcal{L}_{D}^{Ra}$ is optimized in an extra step in order to train the discriminator $\mathcal{D}_{Ra}$. For more details, refer to X. Wang et al. (2019).

3.3. SRR of RE+SWIR bands

In addition to the 10 m GSD RGB+NIR S2 bands, we aim to reconstruct the 20 m GSD S2 images $C = \{\mathbf{B05}_{S2}, \mathbf{B06}_{S2}, \mathbf{B07}_{S2}, \mathbf{B8A}_{S2}, \mathbf{B11}_{S2}, \mathbf{B12}_{S2}\}$ by a factor of 8. Since there are no references available for these bands, we rely on the already reconstructed RGB+NIR data $\mathbf{x}_B$ (see Equation 1. Let $\mathbf{y}_C \in \mathbb{R}^{\frac{W}{8} \times \frac{H}{8} \times 6}$ be the observed intensities of all bands contained in set C. Formally, we are seeking the function

$$\begin{aligned} \mathcal{F}_{8\times}: \quad & \mathbb{R}^{W \times H \times 4} \times \mathbb{R}^{\frac{W}{8} \times \frac{H}{8} \times 6} \rightarrow \mathbb{R}^{W \times H \times 6}, \\ & (\mathbf{x}_B, \mathbf{y}_C) \mapsto \mathbf{x}_C, \end{aligned} \tag{4}$$

where $\mathbf{x}_C \in \mathbb{R}^{W \times H \times 6}$ denotes the super-resolved bands of C with a 2.5 m spatial resolution. As stated above, HR reference images are missing to build a model for $\mathcal{F}_{8\times}$ in Equation 4 Following the idea of transferability of resolution relationships due to self-similarity introduced by Lanaras et al. (2018), we downsample $\mathbf{y}_B$ during training by a factor of 2 and $\mathbf{y}_C$ by a factor of 8 to create synthetic versions $\hat{\mathbf{y}}_B$ and $\hat{\mathbf{y}}_C$ (see Downsampling in Figure 4a). Consequently, the network learns the SRR relationship $(\hat{\mathbf{y}}_B, \hat{\mathbf{y}}_C) \mapsto \hat{\mathbf{x}}_C$ during training, where $\hat{\mathbf{x}}_C \approx \mathbf{y}_C$, which is then applied to $(\mathbf{x}_B, \mathbf{y}_C) \mapsto \mathbf{x}_C$ during inference. As the basis for $\hat{\mathbf{y}}_B$, we rely on the extensively available S2 reference data (see Section 3.1) during training and use the original 10 m bands $\mathbf{y}_B$ instead of reconstructed 2.5 m bands $\mathbf{x}_B$. To build a model for $\mathcal{F}_{8\times}$ in Equation 4, we apply the EDSR-like (Lim et al. 2017) network architecture using residual connections that was utilized by the DSen2 approach of Lanaras et al. (2018). After feeding the inputs, as a pre-step, the network upsamples the input $\hat{\mathbf{y}}_C$ using bilinear interpolation to the size of $\hat{\mathbf{y}}_B$. As the loss function, we use the pixel-wise MAE between the reference image and the SRR result.

3.4. Multi-spectral SRR pipeline

As outlined in Section 3.2, the RGB+NIR SRR network facilitates a complex GAN-based design in order to learn a high-quality mapping between structures and to preserve the spectral properties of the reference data. Conversely, the RE+SWIR SRR network described in Section 3.3 utilizes a more compact design, as spatial structure can be directly inferred from an additional input containing HR data. A schematic overview of the training process for our multi-spectral SRR pipeline is presented in Figure 4a. Training of the two networks can be done in parallel. The RGB+NIR SRR network uses S2 data $\mathbf{y}_B$ as input and SD data $\mathbf{y}_A$ as the reference. The RE+SWIR SRR network uses synthetic downscaled versions $\hat{\mathbf{y}}_B$ and $\hat{\mathbf{y}}_C$ of the RGB+NIR bands of S2, denoted by $\mathbf{y}_B$ and the RE+SWIR bands of S2, denoted by $\mathbf{y}_C$, as inputs. Moreover, it uses the original data $\mathbf{y}_C$ as the reference to obtain the trained

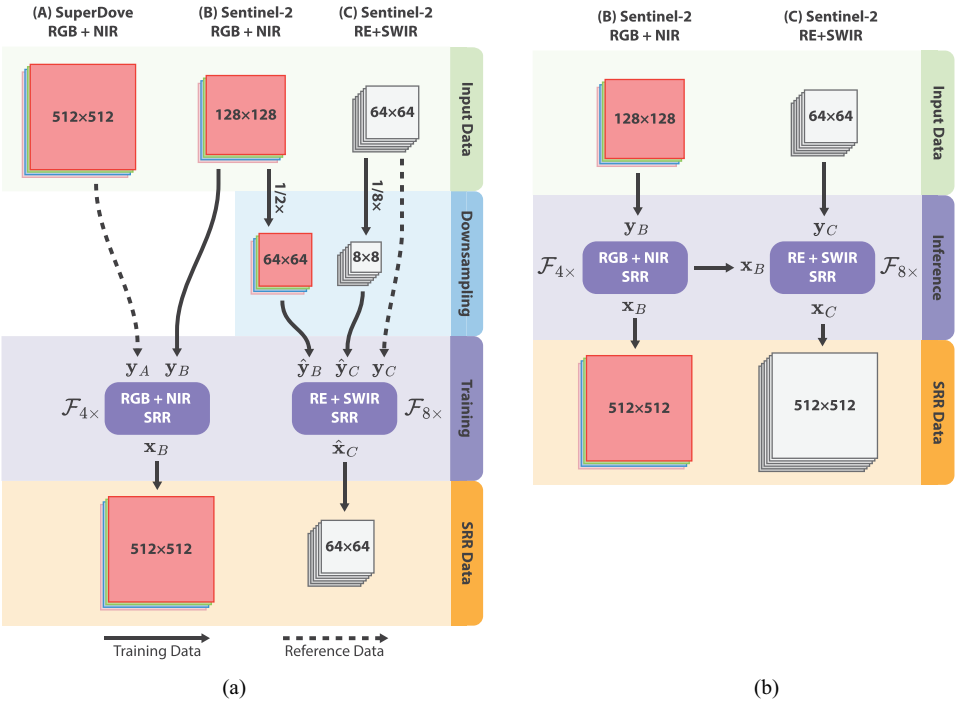

Figure 4. Schematic overview of the training (a) and inference (b) pipelines for the RGB+NIR and RE +SWIR bands. (a) To train the RGB+NIR SRR network, corresponding bands from (A) SuperDove and (B) Sentinel-2 patches are used. Training the RE+SWIR network depends only on Sentinel-2 data ((B) and (C)). SuperDove patches (A) serve only as reference and Sentinel-2 ones (C) are used as both training and reference data. (b) To super-resolve the RGB+NIR bands, only the original (B) Sentinel-2 RGB+NIR bands are required. The output from this step is then used as input for the RE+SWIR SRR model along with the original (C) Sentinel-2 RE+SWIR bands.

weights of the network. During inference (depicted in Figure 4b), first, a ten-band S2 patch with bands from sets *B* and *C* is extracted and is divided into two image sets $\mathbf{y}_B$ and $\mathbf{y}_C$. Next, the RGB+NIR SRR is applied to $\mathbf{y}_B$ which generates $\mathbf{x}_B$, and subsequently the RE +SWIR SRR model yields $\mathbf{x}_C$ from the inputs $\mathbf{x}_B$ and $\mathbf{y}_C$. The final SRR result for a queried ten-band S2 patch is formed by $(\mathbf{x}_B, \mathbf{x}_C)$.

3.5. Implementation

Following Rakotonirina and Rasoanaivo (2020), the RGB+NIR SRR network consists of 23 RRDBs from the ESRGAN+ architecture. In addition, Gaussian noise is added to the output of each residual dense block using the nESRGAN+ version as our final architecture. It is trained in a first phase with a batch size of 4 using the pixel loss $\mathcal{L}_{Pix}$ for 1M iterations with a learning rate of 2×10^{-4} decayed by 0.5 after 200K, 400K, 600K and 800K iterations. In a second phase, the generator is initialized by the trained weights of the first phase and GAN training (c.f. Equation 2) is utilized. Here, $\mathcal{L}_{SRR}^{Ra}$ is scaled by $\lambda = 0.005$, $\mathcal{L}_{Pix}$ is scaled by $\gamma = 0.01$, and for $\mathcal{L}_{Perc}$ an ImageNet-pretrained VGG19 model (Simonyan, Zisserman, and

Zisserman 2014) is applied. $\mathcal{L}_{Perc}$ is calculated as the sum of the RGB-based and NIR-based perceptual loss terms. We have found empirically that computing the features at the twelfth convolution layer level of the VGG19 leads to better SRR performance than at the sixteenth level as proposed by X. Wang et al. (2019). The second training phase is executed for 500K iterations using a learning rate of 1×10^{-4} decayed by 0.5 after 50K, 100K, 200K and 300K iterations. At inference, final SRR images are generated using averaged trained weights from the first and the second training phases as proposed by X. Wang et al. (2019).

Following Lanaras et al. (2018), the RE+SWIR SRR network consists of a convolutional layer and a ReLU activation at the start, followed by 6 residual blocks and another convolutional layer at the end with a skip connection. Each residual block contains 2 convolution layers with a ReLU in between, followed by a residual scaling of 0.1 and an addition of the skip connection in the end. The network is trained with a batch size of 128 for 3.5M iterations using a learning rate of 7×10^{-4} and the weights with the highest validation PSNR are selected for inference.

The (GDAL/OGR contributors 2025) Python library was used to prepare our image data for further processing, and PyTorch (Paszke et al. 2019) was utilized to implement the training and inference routines of our multi-spectral SRR pipeline.

4. Spectral evaluation

In this section, we present the extended spectral evaluation of SRR outputs, which is built on top of the aforementioned SIAM system (Baraldi 2011) to produce knowledge-based, spectral categorizations of both SRR outputs and corresponding reference data. The physical model-based expert system is capable of handling any multi-spectral image data that are radiometrically calibrated to at least top of atmosphere (TOA) reflectance. SIAM is a point-wise operator and maps reflectances into a discrete and finite vocabulary of semi-symbolic spectral categories. The categories do not represent immediate land use or land cover classes, as these high-level semantic concepts cannot be derived in an unambiguous way solely on a per-pixel basis. Confusions of concepts that may have the same spectral signatures, such as between shadows and water, can only be resolved by adding spatial context information (texture, object size, shape and neighbourhood). SIAM categories represent an intermediate level of semantic enrichment that can be derived from the spectral information, and their visualized symbolization is indicative of the colour perceived by the human eye. Using SIAM, the extensive multi-dimensional continuous data space is reduced down to essential information components that can be represented as an 8-bit discrete output raster. The derived categories are application-independent, mutually exclusive and collectively exhaustive and are shown in Table 2.

SIAM has several sensor modes, allowing it to be applied to a range of multi-spectral inputs despite different available bands. For SRR data, the *Landsat-like* sensor mode is used, which incorporates six bands (Red, Green, Blue, NIR, SWIR1 and SWIR2) to output the spectral categorization. For the SD data, the *very high-resolution (VHR)-like* sensor mode is used, utilizing four bands (Red, Green, Blue and NIR). Both of

Table 2. SIAM categories with their colour symbolization, abbreviations and descriptions.

	SVHNIR	Strong vegetation with high NIR
	SVLNIR	Strong vegetation with low NIR
	AVHNIR	Average vegetation with high NIR
	AVLNIR	Average vegetation with low NIR
	WV	Weak vegetation
	VSH_VWA_TWASH	Vegetation in shadow or vegetation in water or turbid water or shadow
	SHRBRHNIR	Shrub rangeland with high NIR
	SHRBRLNIR	Shrub rangeland with low NIR
	HRBCR	Herbaceous rangeland
	WR	Weak rangeland
	PB	Pit bog
	GH	Greenhouse
	VBBB	Very bright Bare soil (barren land) or built-up
	BBB	Bright bare soil (barren land) or built-up
	SBB	Strong bare soil (barren land) or built-up
	ABB	Average bare soil (barren land) or built-up
	DBB	Dark bare soil (barren land) or built-up
	WBB_BBSH	Weak bare soil (barren land) or built-up or bare soil (barren land) or built-up in shadow
	NIRPBB	Near infrared-peaked bare soil (barren land) or built-up
	BA	Burned area
	DPWASH	Deep water or shadow
	SLWASH	Shallow water or shadow
	TWASH	Turbid water or shadow
	SASLWA	Salty shallow water
	CL	Cloud
	SMKPLM	Smoke plume
	TNCLV	Thin cloud over vegetation
	TNCLWA_BB	Thin clouds over water or bare soil (barren land) or built-up
	SN_WAICE	Snow or water ice
	SNSH	Snow in shadow
	SH	Shadow areas
	FLAME	Flame, active fire
	UN	Unknown
	MASKED_NO_DATA	No data

SIAM's sensor modes are applied to S2 data to better assess undesirable differences in spectral categorization compared to the SD and SRR data, and avoid false positives in the detection of spectral changes.

To perform the SRR-inherent cross-sensor comparisons, we leverage a fundamental property of SIAM, which is the production of outputs according to a sensor-mode-agnostic, harmonizing taxonomy with 33 spectral categories. Given these 33 categorical outputs, a categorization of differences that occur between two outputs can be

performed using the two-dimensional change matrix of dimension 33 × 33. Following Baraldi et al. (2023), the 1089 potential combinations of SIAM categories are aggregated into a set of 30 change categories, grouping semantically similar changes to reduce the complexity of the change matrix. All pairs of SIAM categories that indicate a vegetation increase, for example, are subsumed under a single, new combination category. Twenty-one of these 30 aggregated change categories represent actual change categories, while the remaining ones, located on the diagonal of the change matrix, represent constant spectral signatures or negligible changes. In a final step, a weighting scheme assigns a severity level to each of the 21 change categories. Calculating unweighted confusion statistics on categorical maps proves inadequate due to the varying granularity in space discretization inherent to SIAM. For example, there are numerous vegetation-like categories, but only a few water-like categories. Changes from one vegetation-like category to another are likely, even with minor changes in the spectral signature, but they are less pronounced than a change from a vegetation-like category to a water-like category, for example. Changes within a super-set of categories (e.g. within vegetation or soil-like categories) are considered minor changes. All remaining change categories indicate major, more severe changes. The resulting SIAM change matrix is presented in Table 3 along with its legend in Table 4. Aggregated change categories are coloured and represented by a combination of letters and numbers, where the letter indicates the severity grouping and the number distinguishes between different types of changes from a semantic perspective.

Table 3. Change matrix for SIAM categories. The first column and row denote the SIAM categorizations for two data pairs T1 and T2. For a breakdown of the SIAM abbreviations, see Table 2. The changes are categorized within the matrix; a corresponding breakdown of these change categories can be found in Table 4.

T1 \ T2	SVHNIR	SVLNIR	AVHNIR	AVLNIR	WV	VSH_VWA_TWASH	SHRBRHNIR	SHRBRLNIR	HRBCR	WR	PB	GH	VBBB	BBB	SBB	ABB	DBB	WBB_BBSH	NIRPBB	BA	DPWASH	SLWASH	TMASH	SASLWA	CL	SMKPLM	TNCLV	TNCLWA_BB	SN	SNSH	SH	FLAME	UN	NO_DATA
SVHNIR	C_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
SVLNIR	B_2	C_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
AVHNIR	B_2	B_2	C_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
AVLNIR	B_2	B_2	B_2	C_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
WV	B_2	B_2	B_2	B_2	C_1	B_1	B_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
VSH_VWA_TWASH	B_2	B_2	B_2	B_2	B_2	C_1	B_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
SHRBRHNIR	B_2	B_2	B_2	B_2	B_2	B_2	C_1	B_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
SHRBRLNIR	B_2	B_2	B_2	B_2	B_2	B_2	B_2	C_1	B_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
HRBCR	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	C_1	B_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
WR	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	C_1	B_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
PB	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	C_1	B_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
GH	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	B_2	C_1	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_2	A_4	A_4	A_4	A_4	A_5	A_5	A_5	A_5	A_{11}	A_{11}	A_5	A_{17}	A_5	A_5
VBBB	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	C_2	B_3	B_3	B_3	B_3	B_3	B_3	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
BBB	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	C_2	B_3	B_3	B_3	B_3	B_3	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
SBB	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	B_3	C_2	B_3	B_3	B_3	B_3	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
ABB	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	B_3	B_3	C_2	B_3	B_3	B_3	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
DBB	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	B_3	B_3	B_3	C_2	B_3	B_3	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
WBB_BBSH	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	B_3	B_3	B_3	B_3	C_2	B_3	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
NIRPBB	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	B_3	B_3	B_3	B_3	B_3	C_2	B_3	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
BA	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	B_3	B_3	B_3	B_3	B_3	B_3	B_3	C_2	A_7	A_7	A_7	A_7	A_9	A_9	A_9	A_9	A_{11}	A_{11}	A_9	A_{17}	A_9	A_9
DPWASH	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_6	A_6	A_6	A_6	A_6	A_6	A_6	A_6	C_3	C_3	C_3	C_3	A_8	A_8	A_8	A_8	A_{11}	A_{11}	A_8	A_{17}	A_8	A_8
SLWASH	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_6	A_6	A_6	A_6	A_6	A_6	A_6	A_6	C_3	C_3	C_3	C_3	A_8	A_8	A_8	A_8	A_{11}	A_{11}	A_8	A_{17}	A_8	A_8
TWASH	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_6	A_6	A_6	A_6	A_6	A_6	A_6	A_6	C_3	C_3	C_3	C_3	A_8	A_8	A_8	A_8	A_{11}	A_{11}	A_8	A_{17}	A_8	A_8
SASLWA	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_3	A_6	A_6	A_6	A_6	A_6	A_6	A_6	A_6	C_3	C_3	C_3	C_3	A_8	A_8	A_8	A_8	A_{11}	A_{11}	A_8	A_{17}	A_8	A_8
CL	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	C_4	C_4	C_4	C_4	A_{10}	A_{15}	A_{18}	A_{17}	A_{18}	A_{18}
SMKPLM	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	C_4	C_4	C_4	C_4	A_{10}	A_{15}	A_{18}	A_{17}	A_{18}	A_{18}
TNCLV	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	C_4	C_4	C_4	C_4	A_{10}	A_{15}	A_{18}	A_{17}	A_{18}	A_{18}
TNCLWA_BB	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	C_4	C_4	C_4	C_4	A_{10}	A_{15}	A_{18}	A_{17}	A_{18}	A_{18}
SN	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{14}	A_{14}	A_{14}	A_{14}	A_{10}	A_{10}	A_{10}	A_{10}	C_5	C_5	A_{10}	A_{17}	A_{10}	A_{10}
SNSH	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{12}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{13}	A_{14}	A_{14}	A_{14}	A_{14}	A_{15}	A_{15}	A_{15}	A_{15}	C_5	C_6	A_{15}	A_{17}	A_{15}	A_{15}
SH	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	A_{18}	A_{18}	A_{18}	A_{18}	A_{10}	A_{15}	C_7	A_{17}	C_9	C_9
FLAME	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_1	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	A_{16}	A_{16}	A_{16}	A_{16}	A_{10}	A_{15}	A_{16}	C_8	A_{16}	A_{16}
UN	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	A_{18}	A_{18}	A_{18}	A_{18}	A_{10}	A_{15}	C_9	A_{17}	C_9	C_9
NO_DATA	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_5	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_9	A_8	A_8	A_8	A_8	A_{18}	A_{18}	A_{18}	A_{18}	A_{10}	A_{15}	C_9	A_{17}	C_9	C_9

Note that the definition of aggregation rules towards a set of aggregated change categories, as well as the assignment of severity, were carried out in an expert-based manner. Other experts may come to a different conclusion regarding the aggregation and categorization of the changes. Nevertheless, this approach enables a knowledge-based categorization of spectral changes in a transparent and reproducible manner.

Table 4. SIAM change categories grouped by severity.

major, severe changes		A_1	Vegetation total gain from bare soil or built-up or fire
		A_2	Vegetation total loss into bare soil or built-up
		A_3	Vegetation total gain from water (or shadow)
		A_4	Vegetation total loss into water (or shadow)
		A_5	Single-date vegetation
		A_6	Bare soil or built-up total gain from water (or shadow)
		A_7	Bare soil or built-up total loss into water (or shadow)
		A_8	Single-date water (or shadow)
		A_9	Single-date bare soil
		A_{10}	Single-date snow (or shadowed snow)
		A_{11}	Snow (or bright bare soil/built-up or cloud in VHR imagery) total gain
		A_{12}	Vegetation from snow (or bright bare soil/built-up or cloud in VHR imagery)
		A_{13}	Bare soil or built-up from snow (or bright bare soil/built-up or cloud in VHR imagery)
		A_{14}	Water (or shadow) from snow (or bright bare soil/built-up or cloud in VHR imagery)
		A_{15}	Single-date shadowed snow
		A_{16}	Single-date flame
		A_{17}	Active flame
		A_{18}	Single-date cloud
minor changes		B_1	Vegetation decrease
		B_2	Vegetation increase
		B_3	Within-bare soil or built-up change
no change		C_1	Constant vegetation
		C_2	Constant bare soil or built-up
		C_3	Constant water (or shadow)
		C_4	Constant cloud or single-date cloud
		C_5	Constant snow (or bright bare soil/built-up or cloud in VHR imagery)
		C_6	Constant shadowed snow
		C_7	Constant shadow
		C_8	Constant flame
		C_9	Constant unknown or noisy

5. Results and discussion

In this section, we present the results of our SRR approach with a main focus on the evaluation of the spectral stability and fidelity of the SRR output in relation to the S2 input, as well as the SD reference. In particular, we start with a qualitative inspection of the results (Section 5.1), followed by a standard quantitative evaluation using common IQA metrics (Section 5.2). Next, we present the complementary qualitative and quantitative results for our own extended evaluation framework (Section 5.3) and conclude with a report on the time performance of our proposed pipeline (Section 5.4).

5.1. Qualitative inspection

In Figure 1, we provide the SRR output (bottom) of a single S2 input patch (top) across several bands from the portion of our S2 test dataset corresponding to granule *33UWP* acquired in 2021 (see Section 3.1), to allow for a quick visual assessment of the achieved image quality. This visual comparison shows a substantially higher spatial resolution in the SRR output with a sharp delineation of the various different structures. This is most visible in the super-resolved 20 m bands B05 and B11. Also, the SRR appears to perform equally well on agricultural and build-up land without introducing any obvious artefacts.

In Figure 5, we provide a more detailed comparison of the S2 input and SRR output focusing on the spectral profiles derived from agricultural land parcels as required by many downstream tasks. We have selected patches from a different year (2023) and different locations in Austria not included in our training, validation and test data to demonstrate the transferability to unseen locations within the same biome. Again, the visual comparison of RGB bands of S2 (left) and SRR (middle) reveals sharper delineations of agricultural land parcels without introducing artefacts. However, the SRR output appears darker than the input, which suggests slightly lower reflectance values in the bands B02, B03 and B04. This is confirmed by a direct comparison of reflectance profiles (right). To obtain these reflectance profiles, the reflectance values of all pixels fully contained within the polygon of a parcel (red) have been averaged per band and then plotted. While the reflectance profiles are generally in good agreement, the reflectance in the RGB bands is generally slightly lower than in the input. On the other hand, the NIR band B08 exhibits a higher degree of variation and is often overestimated, in particular across bare soil (bottom row). One possible cause of this variation is that, as discussed in Section 3.1, we use SD band B8 as reference data for S2 band B08 during training, even though the latter has a considerably wider wavelength range, which leads to a lower spectral correspondence and possibly lower SRR performance than in the RGB bands. Furthermore, as reported by Yu-Hsuan et al. (2022), the SD data exhibit a generally higher variance in the NIR band as well as a slight underestimation of reflectance values on reflective surfaces. In summary, we can deduce from these observations that the SRR provides visually reasonable results. The extent to which the spectral deviations occur systematically is analysed in Section 5.3.

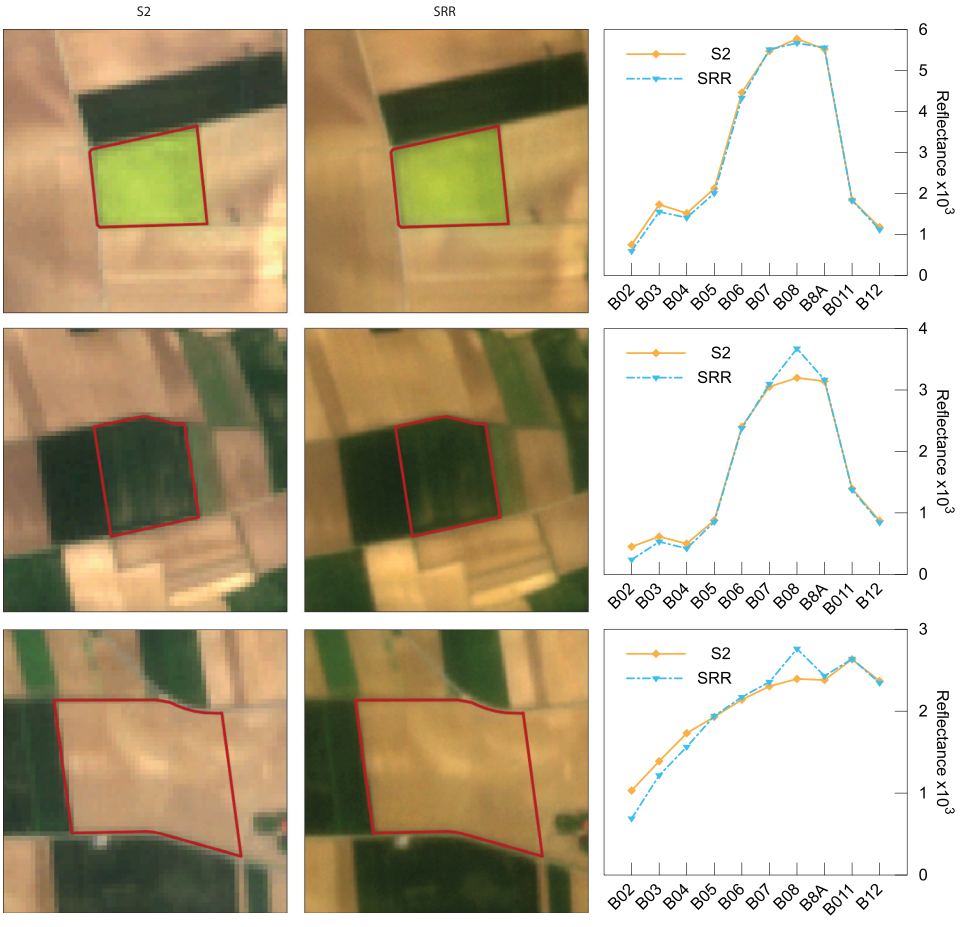

Figure 5. Results of our approach compared to the low-resolution input. RGB bands of S2 input from year 2023 (left) and SRR output (middle) as well as reflectance profiles (right) averaged over an agricultural land parcel (red).

5.2. Image quality assessment

To assess the quality of reconstructed images via common IQA metrics, we use the PSNR and the SSIM between reference test data and SRR outcomes. PSNR quantifies the difference between query samples using the logarithmic ratio of the maximum possible pixel value to the MSE. SSIM, on the other hand, evaluates quality by comparing structural information, luminance and contrast of query images (Z. Wang et al. 2004).

We perform the RGB+NIR SRR evaluation on the 1235 S2 and SD test pairs (see Section 3.1). The model achieves an average PSNR of 31.00 for the RGB bands and 26.23 for the NIR band. Moreover, it obtains an average SSIM of 0.911 for the RGB bands and 0.818 for the NIR band. The results indicate that both the PSNR and SSIM values calculated for the NIR band are inferior to those obtained for the RGB bands. As mentioned in Section 5.1, we attribute this again to the higher variance of the NIR signals in the SD imagery product, as observed, e.g., by Yu-Hsuan et al. (2022).

Moreover, we evaluate the performance of the RE+SWIR SRR model on the S2 RE, Narrow Near-Infrared (NNIR), and SWIR bands using the 29,033 S2 test patches (see Section 3.1). We

obtain an average PSNR of 38.34 and an average SSIM of 0.979 for the RE and NNIR bands (B05, B06, B07, B8A) and an average PSNR of 33.25 and an average SSIM of 0.939 for the SWIR bands (B11, B12). As discussed by Lanaras et al. (2018), the performance drop of the model for the SWIR bands can be attributed to the fact that these bands lie far outside the spectral range covered by high-resolution RGB+NIR images utilized as auxiliary input to the network.

The above IQA results exhibit a comparable trend to those reported in related works (described in Section 2.1). However, a direct quantitative comparison is not possible due to large differences in the datasets employed.

5.3. Spectral, knowledge-based evaluation

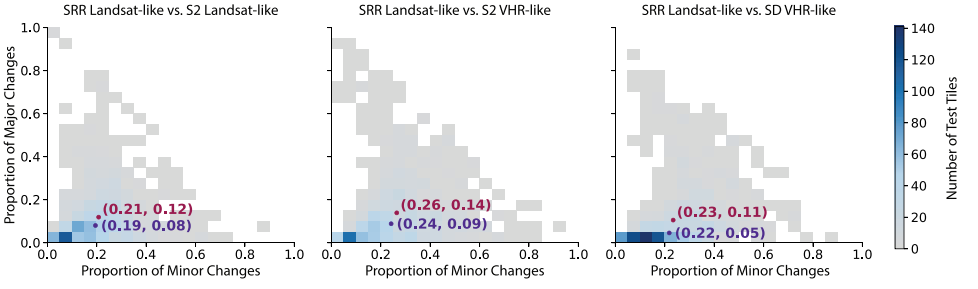

Figure 6. Bivariate frequencies of the patch-wise proportion of minor and major changes in the SIAM categorizations between SRR outputs and input data. Mean (red) and median (purple) values across all S2, SD and SRR patches of the 1235 test pairs are marked by dots. Definitions of minor and major changes are given in Table 4.

The SIAM change approach, outlined in Section 4, enables in-depth spectral validation. We start with a macroscopic comparison of global statistics of SIAM categorization changes between the SRR output and the input products provided in Figure 6. The histograms of Figure 6 show a binning of test patches according to their proportion of minor (horizontal axis) and of major (vertical axis) categorization changes relative to the total number of pixels in a patch. The applied transfer function reveals that most patches are binned in the bottom left part of the histograms, indicating a low number of changes in the majority of tiles. This is supported by the indicated mean and median proportion of minor and major categorization changes over all S2, SD and SRR patches of the 1235 test pairs (see Section 3.1). Summed up, the mean proportion of changed categories is equal to or below 40% in all three cases. Most of the changes are categorized as minor changes. For the S2 input, we can perform the underlying categorization either with six bands in *Landsat-like* mode (left) or with four bands in *VHR-like* mode (middle) and compare to the SRR categorizations in the same mode. In the SD input data, only the four bands required for *VHR-like* mode (right) are available. As noted in Section 4, using different sensor modes on the same data does not lead to the same categorization. Since SD can only be categorized in *VHR-like* mode, a fair assessment of the similarity between S2 and SD inputs and SRR outputs therefore involves a comparison of both input datasets in this mode. The corresponding histograms in Figure 6 show that the SRR outputs have a substantially greater spectral similarity to the SD data (right) than to the S2 data (middle), in particular with regard to major changes. The mean proportion of major changes amounts to 11% for SD (right) and 14% for S2 (middle).

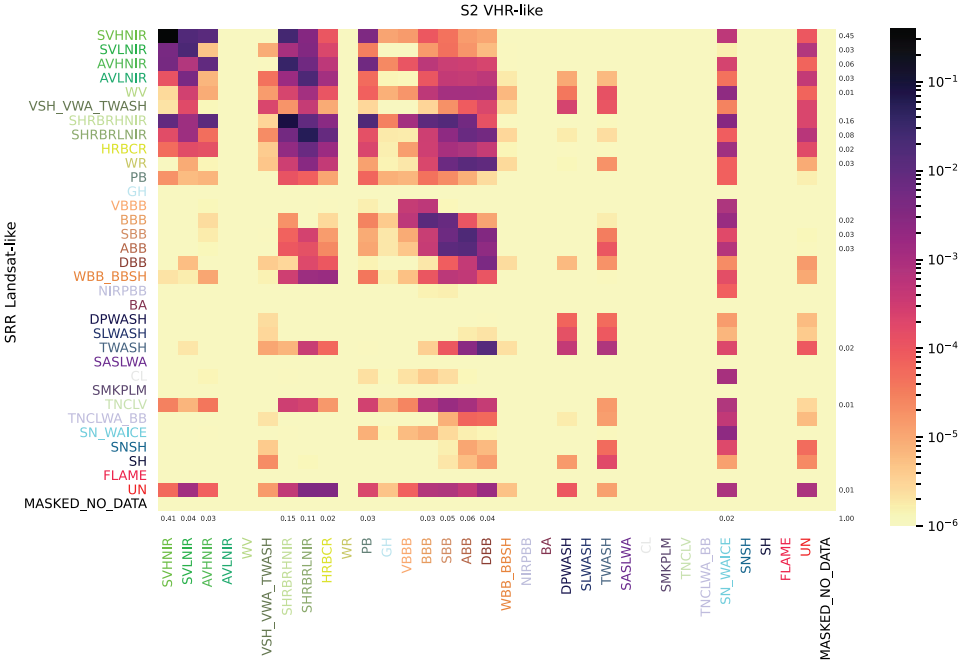

Figure 7. Bivariate histogram of SIAM changes showing a comparison of SRR with S2 categorizations. The individual cells represent the relative frequency of occurrence of a specific SIAM change across all S2 and SRR patches of the 1235 test pairs.

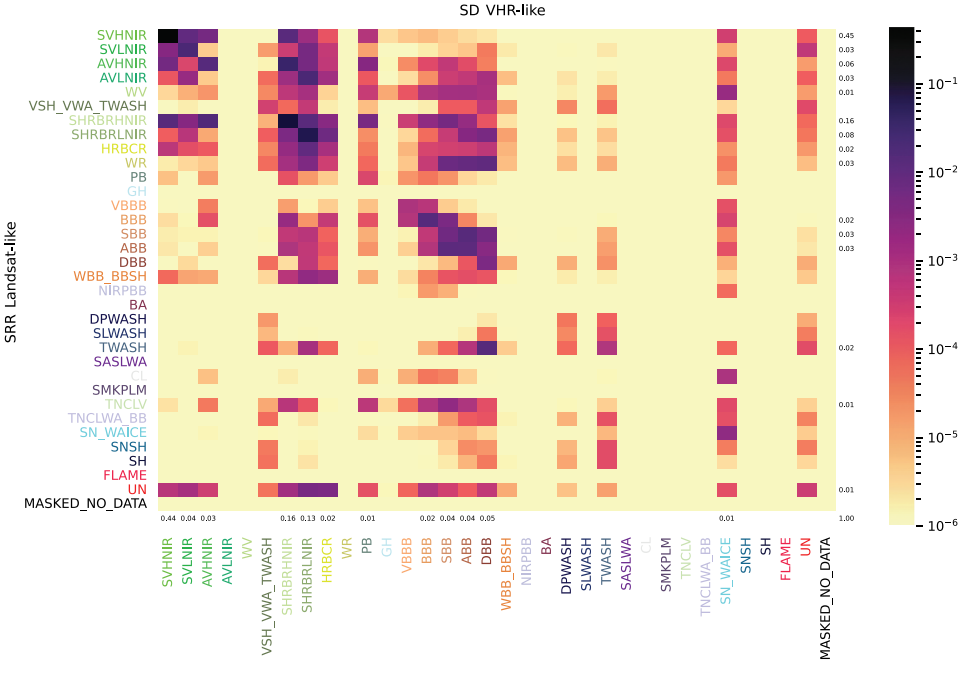

Figure 8. Bivariate histogram of SIAM changes showing a comparison of SRR with SD categorizations. The individual cells represent the relative frequency of occurrence of a specific SIAM change across all SD and SRR patches of the 1235 test pairs.

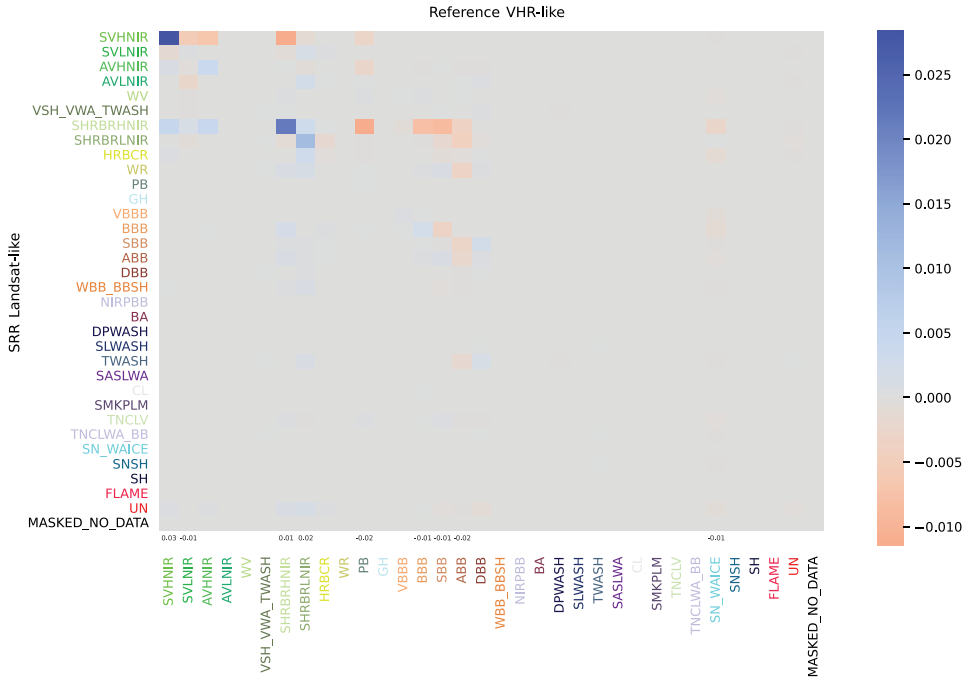

Figure 9. Differential bivariate histogram of individual SIAM changes. Subtracting the histogram in Figure 7 from the one in Figure 8, the differences in the proportion of category transitions between SD and S2, each as a reference to SRR outputs, are shown.

For a more detailed consideration of the comparisons of the SRR categorization with the reference data, a breakdown of the frequencies by individual spectral categories is presented in Figure 7 for S2 data and in Figure 8 for SD data. According to the selection of patches with a focus on agricultural areas, approximately 80% of all areas are categorized as vegetation-like, both in the original data and in the SRR outputs. This can be seen from the row and column totals in Figures 7 and 8. Confirming the previous observation on the dominance of non-severe changes, a large portion of the category transitions occur within super-sets of categories (e.g. within vegetation-like or within bare soil-like categories). Changes classified as severe are primarily transitions from bare soil categories in the original data to weak vegetation in the SRR outputs. The complementary change (i.e. pixels categorized as weak vegetation in the original data and categorized as bare soil in the SRR outputs) also occurs, albeit with lower frequency, as can be seen in the area below the main diagonal in Figures 7 and 8. This unidirectional shift in categorizations from bare soil to vegetation-like signatures is likely to be caused by elevated NIR reflectances in SRR outputs as previously noted in Section 5.1. The change matrix approach provides strong evidence that the previous qualitative observations are not merely isolated cases but point to a systematic bias. Other substantial changes identifiable in Figures 7 and 8 involve re-categorization of original dark soil pixels as water or shadow-like pixels in the SRR outputs. Again, the complementary process is much less pronounced here. Finally, the re-categorization of original pixels categorized as clouds, ice or snow also contributes to the substantial changes. The proportion of unknown spectral signatures averages 0.17% for SD, 0.37% for S2 and 1.43% for the SRR outputs. While the SRR data thus overall reconstructs the input data well, there are some systematic SRR model biases along with an increased proportion of signatures that cannot be interpreted physically using a knowledge-based classifier.

Comparing the spectral categorization of SRR outputs to S2 and SD categorizations in a differential manner (see Figure 9), it is evident that especially for strong vegetation categories encoded in the top left, the SRR outputs exhibit fewer changes in comparison with SD than with S2. The higher stability of the SIAM category changes for the SD-SRR comparison is apparent from the positive values on the main diagonal. The greater severity of changes in the comparison between S2 and SRR can be mainly inferred from two observations: First, a higher proportion of pixels are mapped as strong vegetation (SVHNIR), which were originally categorized as other types of vegetation (see first row in Figure 9). Second, a higher proportion of pixels are marked as shrubland (SHRBRHNIR), which were originally categorized as bog or bare soil (see seventh row in Figure 9). These observations, given by the aggregated statistics, are reflected in an example shown in Figure 10, where SIAM categorizations with the corresponding classification of change severities are visualized for a single patch from our 1235 test pairs (see Section 3.1). In this case, the severe changes originate mainly from soil-like and vegetation-like confusions between S2 and SRR, which are less prominent between SD and SRR. Also, the poorer spatial resolution of S2 leads to an increased level of changes, in particular at object boundaries. However, the proportion of these resolution-related, line-like changes is marginal compared to the areal changes of entire fields.

5.4. Time performance

We measure the inference throughput of our multi-spectral SRR pipeline (see Section 3.4) for super-resolving the 1280 m $\times$ 1280 m patches introduced in Section 3.1, as well as the throughput for SIAM categorizations. For every patch, the RGB+NIR SRR module reconstructs the four 10 m S2 bands, and the RE+SWIR SRR model upscales the six 20 m S2 bands. When applied to the 1235 test pairs (see Section 3.1), our pipeline, including both SRR modules, processes on average 11.11 patches per second on an NVIDIA L40S GPU based on all S2 patches, while the SIAM system evaluates on average 0.54 patches per second on an AMD Ryzen 7 5800H CPU with 32 GB RAM using all S2, SD and SRR patches.

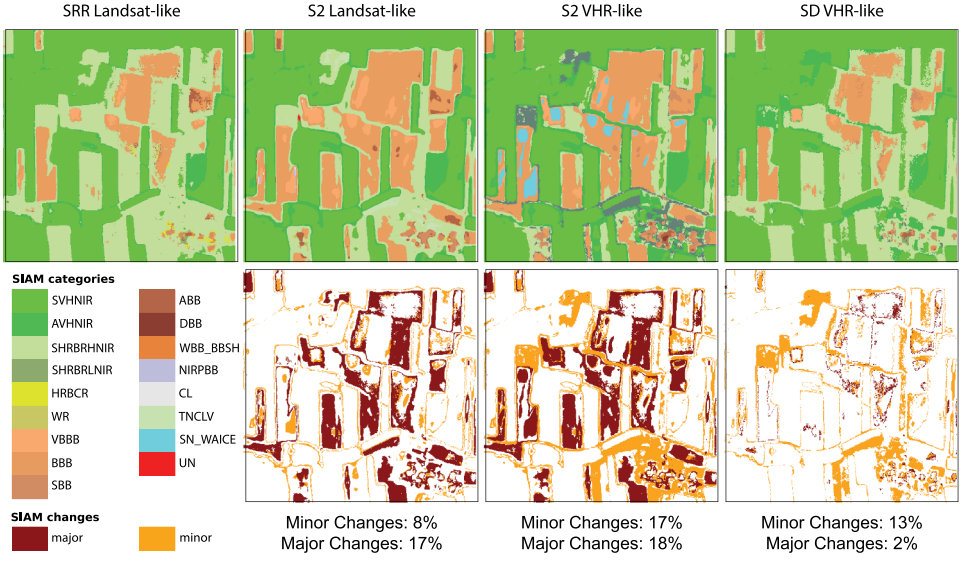

Figure 10. SIAM categorisations (top row) and change evaluations (bottom row) for an individual patch. Changes are measured relative to the SRR result. Severe changes are visualised in dark-red, minor changes in orange. The legend shows the colour scheme of the SIAM categorisations, for more details the reader is referred to Table 2.

6. Conclusion

We have introduced a holistic pipeline that upscales S2 imagery by a deep learning-based single-image super-resolution reconstruction approach and provides a comprehensive evaluation framework to assess the spectral fidelity of the reconstructed outcomes. Our approach focuses on maintaining validity across multiple spectral bands while improving the spatial resolution and comprises a novel combination of two CNNs dedicated to the SRR of 10 m bands and 20 m bands, respectively. Both models are designed to upsample images to a spatial resolution of 2.5 m. The first model is trained with corresponding pairs of S2 and SD images, while the second model is trained without reference images from an auxiliary satellite.

Beyond qualitative inspections and IQA-based quantifications, we adapted a novel knowledge-based spectral categorization system to derive in-depth insights on the spectral properties of the SRR approach on various aggregation levels. The promising results in all of these evaluation environments suggest that the presented SRR approach has a great potential for subsequent agricultural applications. Moreover, the spectral categorization framework allowed to derive quantitative statistics that helped uncover systematic biases of the SRR pipeline, which could be easily described due to the semantics inherent to our approach. Looking ahead, we firmly believe that this has the potential to facilitate iterative SRR model improvements as well as to enhance trust in SRR products, resulting in an improved applicability of SRR models for downstream tasks such as land use classification. Even though we discussed the results in the context of applications in the agricultural domain, we want to emphasize that neither our SRR approach nor our approach to spectral evaluation is restricted to these applications.

Acknowledgement(s)

This work contains modified Copernicus Sentinel data (2021, 2023) (European Commission 2024) as well as modified data by Planet Labs PBC (2021) (Planet Labs PBC 2024). We thank AgrarMarkt Austria for providing a visual evaluation of the created SRR data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Austrian Research Promotion Agency (FFG) under Grant FO999892195. The VRVis GmbH is funded by BMIMI, BMWET, Tyrol, Vorarlberg and Vienna Business Agency in the scope of COMET – Competence Centers for Excellent Technologies (879730, 911654) which is managed by FFG.

ORCID

David Major http://orcid.org/0000-0002-9091-3684
Zsolt Horváth http://orcid.org/0000-0001-7188-6564
Felix Kröber http://orcid.org/0009-0008-1711-1125
Hannah Augustin http://orcid.org/0000-0002-3334-5350

Martin Sudmanns http://orcid.org/0000-0002-0473-1260
Petr Ševčík http://orcid.org/0009-0005-1564-8525
Andrea Baraldi http://orcid.org/0000-0001-5196-9944
Astrid Berg http://orcid.org/0000-0002-2300-2661
Daniel Cornel http://orcid.org/0000-0002-2481-6720
Dirk Tiede http://orcid.org/0000-0002-5473-3344

Data availability statement

Due to privacy reasons, the data used in this study cannot be made publicly available.

References

Anwar, S., S. Khan, and N. Barnes. 2020. "A Deep Journey into Super-Resolution: A Survey." *ACM Computing Surveys* 53 (3): 1–34. https://doi.org/10.1145/3390462.

Arvor, D., J. Betbeder, F. R. Daher, T. Blossier, R. Le Roux, S. Corgne, C. A. da Silva Junior, V. de Freitas Silgueiro, and C. A. D. Silva Junior. 2021. "Towards User-Adaptive Remote Sensing: Knowledge-Driven Automatic Classification of Sentinel-2 Time Series." *Remote Sensing of Environment* 264:112615. https://doi.org/10.1016/j.rse.2021.112615.

Augustin, H., M. Sudmanns, D. Tiede, S. Lang, and A. Baraldi. 2019. "Semantic Earth Observation Data Cubes." *Data* 4 (3): 102. https://doi.org/10.3390/data4030102.

Aybar, C., D. Montero, S. Donike, F. Kalaitzis, and L. Gómez-Chova. 2024. "A Comprehensive Benchmark for Optical Remote Sensing Image Super-Resolution." *IEEE Geoscience & Remote Sensing Letters* 21:1–5. https://doi.org/10.1109/LGRS.2024.3401394.

Baraldi, A. 2011. "Satellite Image Automatic Mapper - A Turnkey Software Executable for Automatic Near Real-Time Multi-Sensor Multi-Resolution Spectral Rule-Based Preliminary Classification of Spaceborne Multi-Spectral Images." *Recent Patents on Space Technology* 1 (2): 81–106. https://doi.org/10.2174/1877611611101020081.

Baraldi, A., M. Laurence Humber, D. Tiede, and S. Lang. 2018a. "Geo-CEOS Stage 4 Validation of the Satellite Image Automatic Mapper Lightweight Computer Program for ESA Earth Observation Level 2 Product Generation - Part 1: Theory." *Cogent Geoscience* 4 (1): 1–46. https://doi.org/10.1080/23312041.2018.1467357.

Baraldi, A., M. Laurence Humber, D. Tiede, and S. Lang. 2018b. "Geo-CEOS Stage 4 Validation of the Satellite Image Automatic Mapper Lightweight Computer Program for ESA Earth Observation Level 2 Product Generation – Part 2: Validation." *Cogent Geoscience* 4 (1): 1–52. https://doi.org/10.1080/23312041.2018.1467254.

Baraldi, A., L. D. Sapia, D. Tiede, M. Sudmanns, H. L. Augustin, and S. Lang. 2023. "Innovative Analysis Ready Data (ARD) Product and Process Requirements, Software System Design, Algorithms and Implementation at the Midstream as *Necessary-But-Not-Sufficient* Precondition of the Downstream in a New Notion of Space Economy 4.0 - Part 1: Problem Background in Artificial General Intelligence (AGI)." *Big Earth Data* 7 (3): 455–693. https://doi.org/10.1080/20964471.2021.2017549.

Blaschke, T. 2010. "Object Based Image Analysis for Remote Sensing." *ISPRS Journal of Photogrammetry & Remote Sensing* 65 (1): 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004.

Blau, Y., R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor 2019. "The 2018 PIRM Challenge on Perceptual Image Super-Resolution." *ArXiv:1809.07517 [cs.CV]*, https://doi.org/10.48550/arXiv.1809.07517.

Chen, S., Y. Ogawa, C. Zhao, and Y. Sekimoto. 2023. "Large-Scale Individual Building Extraction from Open-Source Satellite Imagery via Super-Resolution-Based Instance Segmentation Approach." *ISPRS Journal of Photogrammetry & Remote Sensing* 195:129–152. https://doi.org/10.1016/j.isprsjprs.2022.11.006.

Dong, C., C. Change Loy, K. He, and X. Tang. 2016. "Image Super-Resolution Using Deep Convolutional Networks." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 38 (2): 295–307. https://doi.org/10.1109/TPAMI.2015.2439281.

European Commission. 2024. "Legal Notice on the Use of Copernicus Sentinel Data and Service Information." Accessed July 25, 2025. https://sentinels.copernicus.eu/documents/247904/690755/Sentinel_Data_Legal_Notice.

Galar, M., R. Sesma, C. Ayala, L. Albizua, and C. Aranda. 2020. "Super-Resolution of Sentinel-2 Images Using Convolutional Neural Networks and Real Ground Truth Data." *Remote Sensing* 12 (18). https://doi.org/10.3390/rs12182941.

GDAL/OGR contributors. 2025. *GDAL/OGR Geospatial Data Abstraction Software Library*, Open Source Geospatial Foundation. Accessed July 25, 2025. https://gdal.org .

Gupta, A., R. Mishra, and Y. Zhang. 2024. "Senglean: An End-to-End Deep Learning Approach for Super-Resolution of Sentinel-2 Multiresolution Multispectral Images." *IEEE Transactions on Geoscience & Remote Sensing* 62:1–19. https://doi.org/10.1109/TGRS.2024.3374575.

Ho, J., A. Jain, and P. Abbeel. 2020. "Denoising Diffusion Probabilistic Models." *Advances in Neural Information Processing Systems* 33:6840–6851. https://doi.org/10.48550/arXiv.2006.11239.

Karras, T., S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. 2020. "Analyzing and Improving the Image Quality of StyleGAN." In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Seattle, WA, USA (IEEE), 8110–8119. https://doi.org/10.1109/CVPR42600.2020.00813.

Kawulok, M., P. Kowaleczko, M. Ziaja, J. Nalepa, D. Kostrzewa, D. Latini, D. De Santis, et al. 2024. "Hyperspectral Image Super-Resolution: Task-Based Evaluation." *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing* 17:18949–18966. https://doi.org/10.1109/JSTARS.2024.3475644.

Kim, J., J. Kwon Lee, and K. Mu Lee. 2016. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks." In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Las Vegas, NV, USA (IEEE), 1646–1654. https://doi.org/10.1109/CVPR.2016.182.

Lanaras, C., J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler. 2018. "Super-Resolution of Sentinel-2 Images: Learning a Globally Applicable Deep Neural Network." *ISPRS Journal of Photogrammetry & Remote Sensing* 146:305–319. https://doi.org/10.1016/j.isprsjprs.2018.09.018.

Latte, N., and P. Lejeune. 2020. "Planetscope Radiometric Normalization and Sentinel-2 Super-Resolution (2.5 M): A Straightforward Spectral-Spatial Fusion of Multi-Satellite Multi-Sensor Images Using Residual Convolutional Neural Networks." *Remote Sensing* 12 (15). https://doi.org/10.3390/rs12152366.

Li, Y., Z. Du, S. Wu, Y. Wang, Z. Wang, X. Zhao, and F. Zhang. 2021. "Progressive Split-Merge Super Resolution for Hyperspectral Imagery with Group Attention and Gradient Guidance." *ISPRS Journal of Photogrammetry & Remote Sensing* 182:14–36. https://doi.org/10.1016/j.isprsjprs.2021.09.023.

Lim, B., S. Son, H. Kim, S. Nah, and K. Mu Lee. 2017. "Enhanced Deep Residual Networks for Single Image Super-Resolution." In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* Honolulu, HI, USA (IEEE), 1132–1140. https://doi.org/10.1109/CVPRW.2017.151.

Llano, X. C. 2021. "Coregistration." Accessed December 4, 2024. https://github.com/SMByC/Coregistration-Qgis-processing.

Michel, J., J. Vinasco-Salinas, J. Inglada, and O. Hagolle. 2022. "SEN2VENμS, a Dataset for the Training of Sentinel-2 Super-Resolution Algorithms." *Data* 7 (7). https://doi.org/10.3390/data7070096.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, et al. 2019. "PyTorch: An Imperative Style, High-Performance Deep Learning Library." *Advances in Neural Information Processing Systems* 32. https://doi.org/10.48550/arXiv.1912.01703.

Planet Labs PBC. 2024. "Planet Application Program Interface: In Space for Life on Earth." Accessed July 25, 2025. https://api.planet.com .

Rakotonirina, N. C., and A. Rasoanaivo. 2020. "ESRGAN+: Further Improving Enhanced Super-Resolution Generative Adversarial Network." In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* Barcelona, Spain (IEEE), 3637–3641. https://doi.org/10.1109/ICASSP40776.2020.9054071.

Razzak, M. T., G. Mateo-Garcia, G. Lecuyer, L. Gómez-Chova, Y. Gal, and F. Kalaitzis. 2023. "Multi-Spectral Multi-Image Super-Resolution of Sentinel-2 with Radiometric Consistency Losses and Its Effect on Building Delineation." *ISPRS Journal of Photogrammetry & Remote Sensing* 195:1–13. https://doi.org/10.1016/j.isprsjprs.2022.10.019.

Romero, S., J. M. Luis, and V. Vilaplana. 2020. "Super-Resolution of Sentinel-2 Imagery Using Generative Adversarial Networks." *Remote Sensing* 12 (15): 2424. https://doi.org/10.3390/rs12152424.

Salgueiro, L., J. Marcello, and V. Vilaplana. 2021. "Single-Image Super-Resolution of Sentinel-2 Low Resolution Bands with Residual Dense Convolutional Neural Networks." *Remote Sensing* 13 (24): 5007. 24. https://doi.org/10.3390/rs13245007.

Scheffler, D., A. Hollstein, H. Diedrich, K. Segl, and P. Hostert. 2017. "AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data." *Remote Sensing* 9 (7): 1–21. https://doi.org/10.3390/rs9070676.

Simonetti, D., A. Marelli, and H. Eva. 2015. *Impact: Portable GIS Toolbox for Image Processing and Land Cover Mapping*. Luxembourg (Luxembourg): Publications Office of the European Union. https://doi.org/10.2788/143497.

Simonyan, K., A. Zisserman, and A. Zisserman. 2014. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *ArXiv:1409.1556 [cs.CV]*. https://doi.org/10.48550/arXiv.1409.1556.

Sudmanns, M., H. Augustin, L. Van Der Meer, A. Baraldi, and D. Tiede. 2021. "The Austrian Semantic EO Data Cube Infrastructure." *Remote Sensing* 13 (23): 4807. https://doi.org/10.3390/rs13234807 .

Tao, Y., S. J. Conway, J.-P. Muller, A. R. D. Putri, N. Thomas, and G. Cremonese. 2021. "Single Image Super-Resolution Restoration of TGO CaSSIS Colour Images: Demonstration with Perseverance Rover Landing Site and Mars Science Targets." *Remote Sensing* 13 (9). https://doi.org/10.3390/rs13091777.

Tarasiewicz, T., J. Nalepa, R. A. Farrugia, G. Valentino, M. Chen, J. A. Briffa, and M. Kawulok. 2023. "Multitemporal and Multispectral Data Fusion for Super-Resolution of Sentinel-2 Images." *IEEE Transactions on Geoscience & Remote Sensing* 61:1–19. https://doi.org/10.1109/TGRS.2023.3311622.

Ton, J., J. Sticklen, and A. K. Jain. 1991. "Knowledge-Based Segmentation of Landsat Images." *IEEE Transactions on Geoscience & Remote Sensing* 29 (2): 222–232. https://doi.org/10.1109/36.73663.

Tsagkatakis, G., A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides. 2019. "Survey of Deep-Learning Approaches for Remote Sensing Observation Enhancement." *Sensors* 19 (18). https://doi.org/10.3390/s19183929.

Vasilescu, V., M. Datcu, and D. Faur. 2023. "A CNN-Based Sentinel-2 Image Super-Resolution Method Using Multiobjective Training." *IEEE Transactions on Geoscience & Remote Sensing* 61:1–14. https://doi.org/10.1109/TGRS.2023.3240296.

Wang, P., B. Bayram, and E. Sertel. 2022. "A Comprehensive Review on Deep Learning Based Remote Sensing Image Super-Resolution Methods." *Earth-Science Reviews* 232:104110. https://doi.org/10.1016/j.earscirev.2022.104110.

Wang, X., K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. 2019. "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks." In *European Conference on Computer Vision (ECCV) Workshops* Munich, Germany L. Leal-Taixé and S. Roth , 63–79. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-11021-5_5.

Wang, Z., and A. C. Bovik. 2009. "Mean Squared Error: Love it or Leave It? A New Look at Signal Fidelity Measures." *IEEE Signal Processing Magazine* 26 (1): 98–117. https://doi.org/10.1109/MSP.2008.930649.

Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13 (4): 600–612. https://doi.org/10.1109/TIP.2003.819861.

Wang, Z., J. Chen, and S. C. H. Hoi. 2020. "Deep Learning for Image Super-Resolution: A Survey." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 43 (10): 3365–3387. https://doi.org/10.1109/TPAMI.2020.2982166.

Wharton, S. 1987. "A Spectral-Knowledge-Based Approach for Urban Land-Cover Discrimination." *IEEE Transactions on Geoscience & Remote Sensing* GE-25 (3): 272–282. https://doi.org/10.1109/TGRS.1987.289799.

Wolters, P., F. Bastani, and A. Kembhavi. 2023. "Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing." *ArXiv:2311.18082 [cs.CV]*, https://doi.org/10.48550/arXiv.2311.18082.

Xie, J., L. Fang, B. Zhang, J. Chanussot, and S. Li. 2022. "Super Resolution Guided Deep Network for Land Cover Classification from Remote Sensing Images." *IEEE Transactions on Geoscience & Remote Sensing* 60:1–12. https://doi.org/10.1109/TGRS.2021.3120891.

Yang, J., J. Wright, T. S. Huang, and Y. Ma. 2010. "Image Super-Resolution via Sparse Representation." *IEEE Transactions on Image Processing* 19 (11): 2861–2873. https://doi.org/10.1109/TIP.2010.2050625.

Yu-Hsuan, T., K. Johansen, B. Aragon, M. M. El Hajj, and M. F. McCabe. 2022. "The Radiometric Accuracy of the 8-Band Multi-Spectral Surface Reflectance from the Planet SuperDove Constellation." *International Journal of Applied Earth Observation and Geoinformation* 114 (103035): 1–19. https://doi.org/10.1016/j.jag.2022.103035.

Zhang, H., and B. Huang. 2011. "Scale Conversion of Multi Sensor Remote Sensing Image Using Single Frame Super Resolution Technology." In *19th International Conference on Geoinformatics* Shanghai, China (IEEE), 1–5. https://doi.org/10.1109/GeoInformatics.2011.5980856.

Zhang, R., P. Isola, A. A. Efros, E. Shechtman, and O. Wang. 2018a. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric." *ArXiv:1801.03924 [cs.CV]*, https://doi.org/10.48550/arXiv.1801.03924.

Zhang, Y., Y. Tian, Y. Kong, B. Zhong, and Y. Fu. 2018b. "Residual Dense Network for Image Super-Resolution." In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Salt Lake City, UT, USA (IEEE), 2472–2481. https://doi.org/10.48550/arXiv.1802.08797.