



OPEN

DATA DESCRIPTOR

SEN2NAIP: A large-scale dataset for Sentinel-2 Image Super-Resolution

Cesar Aybar¹✉, David Montero², Julio Contreras¹, Simon Donike¹, Freddie Kalaitzis³ & Luis Gómez-Chova¹

The increasing demand for high spatial resolution in remote sensing has underscored the need for super-resolution (SR) algorithms that can upscale low-resolution (LR) images to high-resolution (HR) ones. To address this, we present SEN2NAIP, a novel and extensive dataset explicitly developed to support SR model training. SEN2NAIP comprises two main components. The first is a set of 2,851 LR-HR image pairs, each covering 1.46 square kilometers. These pairs are produced using LR images from Sentinel-2 (S2) and corresponding HR images from the National Agriculture Imagery Program (NAIP). Using this cross-sensor dataset, we developed a degradation model capable of converting NAIP images to match the characteristics of S2 imagery (S_{2like}). This led to the creation of a second subset, consisting of 35,314 NAIP images and their corresponding S_{2like} counterparts, generated using the degradation model. With the SEN2NAIP dataset, we aim to provide a valuable resource that facilitates the exploration of new techniques for enhancing the spatial resolution of Sentinel-2 imagery.

Background & Summary

Over the past few decades, remote sensing has experienced significant growth, with an escalating demand for updated high-resolution (HR) imagery. This demand arises from various real-world applications, such as crop delineation¹ and urban planning², where precise spatial information is crucial for decision-making. However, due to technical and economic constraints, open-access Earth observation missions often face challenges in delivering imagery at the desired high spatial resolution. Consequently, there is a pressing need for sophisticated super-resolution (SR) models to enhance the ground sampling distance of publicly available low-resolution (LR) imagery.

Creating training datasets for super-resolution in remote sensing is particularly challenging due to the limited availability of HR multispectral imagery. Although platforms such as PlanetScope, SPOT-6/7, and WorldView-3 provide resolutions more than twice that of freely available data like Sentinel-2 (S2), their use is limited by proprietary licenses and restricting accessibility for open research and public applications. Additionally, it is crucial to select an LR image within an appropriate time frame to ensure that it accurately corresponds to the HR image representation. Under this situation, researchers primarily adopt two methodologies: the synthetic and cross-sensor approaches. The synthetic method generates LR images by applying a degradation model to HR images³. Conversely, the cross-sensor approach utilizes real LR satellite images from a different sensor³.

Each approach has its advantages and limitations. The *synthetic* approach excels in providing “controlled environments”, where the LR image is an exact downsampled version of the HR image, making it ideal for testing new SR networks. However, relying solely on synthetic datasets can introduce significant distribution shifts during inference when applied to real-world LR imagery⁴. Prominent examples of synthetic datasets include WHU-RS19⁵, RSSCN7⁶, and AID⁷. In contrast, *cross-sensor* datasets face challenges in the scalability of the number of samples, as they require the collection of collocated LR-HR image pairs within a short time frame to minimize surface and atmospheric discrepancies. Even with simultaneous image acquisition, reflectance values may differ due to variations in sensor calibration and viewing angles. A notable example of a cross-sensor dataset is SEN2VEN μ S³, which harmonizes Sentinel-2 (LR) and Venus (HR) imagery before applying SR methods.

¹Image Processing Laboratory, University of Valencia, Valencia, Spain. ²Institute for Earth System Science and Remote Sensing, Leipzig University, Leipzig, Germany. ³Oxford Applied and Theoretical ML Group, University of Oxford, Oxford, UK. ✉e-mail: cesar.aybar@uv.es

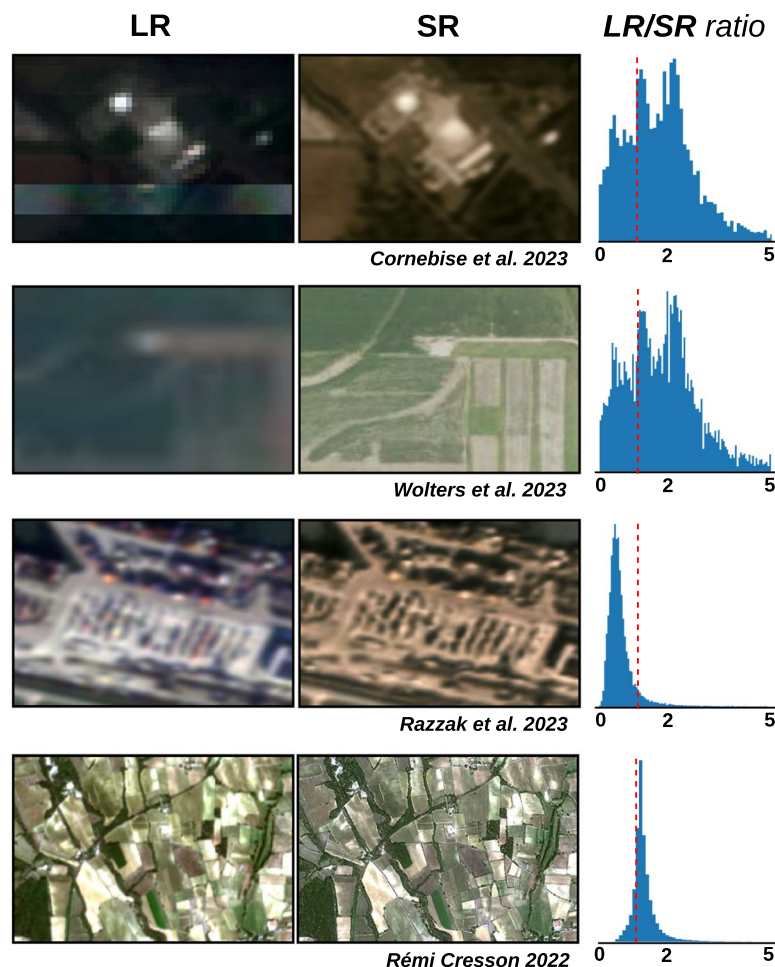


Fig. 1 Comparison of an S2 image with its corresponding super-resolved counterpart highlights the spectral inconsistency of state-of-the-art SR models. The third column displays the pixel-level ratio between the S2 image and the super-resolved results downsampled to the original S2 resolution of 10 m. The models evaluated include Worlstrat⁸ (first row), AllenAISR⁹ (second row), Razzak¹⁰ (third row), and SR4RS¹¹ (fourth row).

While current state-of-the-art SR networks^{8–11} excel in enhancing high-frequency details, they often struggle to maintain coherence, particularly in the spectral domain¹² (Fig. 1). This misalignment arises from inadequate harmonization in cross-sensor approaches¹². In remote sensing, pixel values represent more than mere color intensity; they provide radiance observations crucial for retrieving biophysical variables and ensuring accurate environmental interpretation^{13–15}. Therefore, it is imperative to utilize transparent datasets that not only facilitate effective model training but also ensure accurate spectral and spatial consistency.

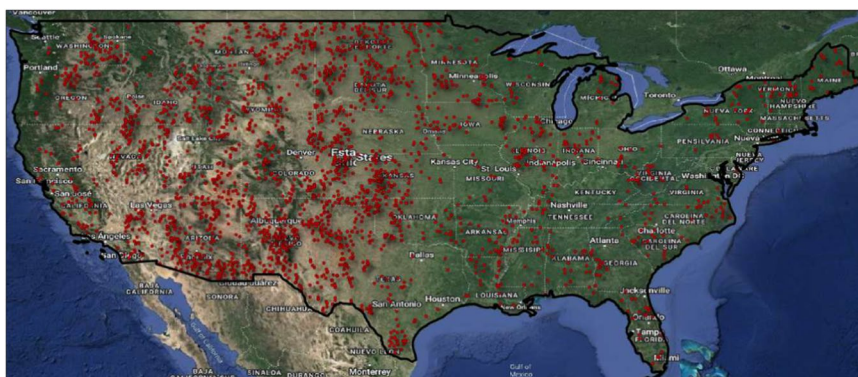
In this paper, we present an innovative hybrid dataset that combines the advantages of both synthetic and cross-sensor approaches, effectively addressing the issue highlighted in Fig. 1. First, a cross-sensor dataset comprising 2,851 image pairs of Sentinel-2 (S2) at 10 m and aerial images from the National Agriculture Imagery Program (NAIP) at 2.5 m is crafted to improve our understanding of the complex relationship between S2-NAIP image pairs (Fig. 2 cross-sensor dataset). With this dataset, we focus on learning an optimal point spread function (PSF), spectral alignment function, and the noise pattern between NAIP and S2. By accurately modeling these components, we can synthetically generate LR counterparts, removing the need to capture real HR and LR images simultaneously. This significantly enhances dataset scalability. Using the learned degradation model, a second dataset is created, consisting of 17,657 locations (Fig. 2 synthetic dataset), where two NAIP images (oldest and more recent) are selected per location, and their LR counterparts are synthetically generated ($S2_{like}$). NAIP was chosen over other HR products due to its extensive coverage, public availability, and public domain license, making it ideal for creating large-scale training datasets.

Methods

This section introduces the data products and details all the preprocessing steps for generating the cross-sensor and synthetic datasets. Figure 3A presents the workflow for crafting the cross-sensor dataset, while Fig. 3B depicts the workflow utilized for generating the synthetic dataset.

Data products. *Sentinel-2.* The S2 mission comprises two nearly identical satellites, Sentinel-2A and Sentinel-2B, launched in June 2015 and March 2017, respectively. Their Level 2 (L2A) products offer estimates of

Cross-Sensor Dataset (2 851 ROIs)



Synthetic Dataset (17 657 ROIs)

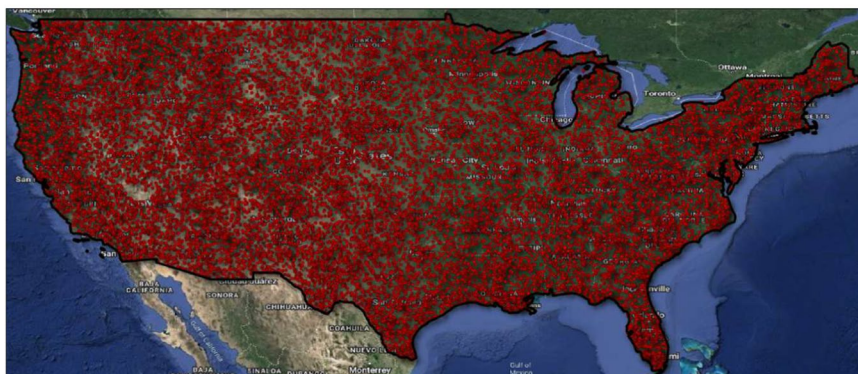


Fig. 2 The locations of cross-sensor (A) and synthetic (B) regions of interest (ROIs) within the SEN2NAIP dataset. Unlike the cross-sensor subset, each ROI within the synthetic subset comprises two NAIP images.

surface reflectance values across 13 spectral channels, covering the entire globe every five days. The S2 imagery is freely distributed under an open data policy. In the SEN2NAIP dataset, we focus on S2 10-meter bands, which include red, green, blue, and near-infrared bands (RGBNIR).

National Agriculture Imagery Program (NAIP). The National Agriculture Imagery Program (NAIP) captures aerial imagery of the contiguous United States and provides updates every three years (five years before 2009). The image acquisition mainly occurs from June to August, when the country experiences the peak growing seasons. While images are available from 2002, only those from 2011 onward are considered for SEN2NAIP, as they always have complete RGBNIR bands and a stable ground sampling distance less than or equal to 1.0 meters. In SEN2NAIP, we aim for $4\times$ super-resolution, so we download NAIP images from GEE, setting the scale to 2.5 meters. GEE stores images using a pyramiding policy, which aggregates 2×2 blocks of pixels. Therefore, instead of downloading images at their native resolution (~ 1 m), we retrieve data from a higher level of the pyramid. This approach reduces the volume of data downloaded while preserving the necessary details for our $4\times$ super-resolution tasks.

Cross-sensor dataset. The cross-sensor dataset begins by generating 18,482 random regions of interest (ROIs) across the contiguous United States using a hard-core point model¹⁶. This model ensures that each point is at least 5 km apart from any other point, effectively preventing spatial overlap between images. Subsequently, we remove S2-NAIP pairs captured with a difference higher than 1 day to maintain similar landscape conditions. In addition, we use the CloudSEN12 UNetMobV2¹⁷ cloud detection algorithm to automatically exclude pairs that contain clouds in the S2 image. The time and cloud filters reduced by 61% the original number of potential images. S2 and NAIP images that pass these initial screenings are retrieved using the R client for Google Earth Engine¹⁸.

To harmonize the reflectance values between NAIP and S2 images, we apply histogram matching to each NAIP band using S2 as a reference. This results in $NAIP_h$. It is well-known that histogram matching can be effective for adjusting the overall reflectance intensity distribution of images, but it struggles to account for local variations, failing to preserve finer details like local contrast and texture. However, in our dataset, we work with small patches, 484×484 pixels for HR and 121×121 pixels for LR images (see Record section). This patch-based approach helps mitigate the issue of local variance, as smaller regions are more homogeneous, allowing histogram matching to perform more accurately and retain local details.

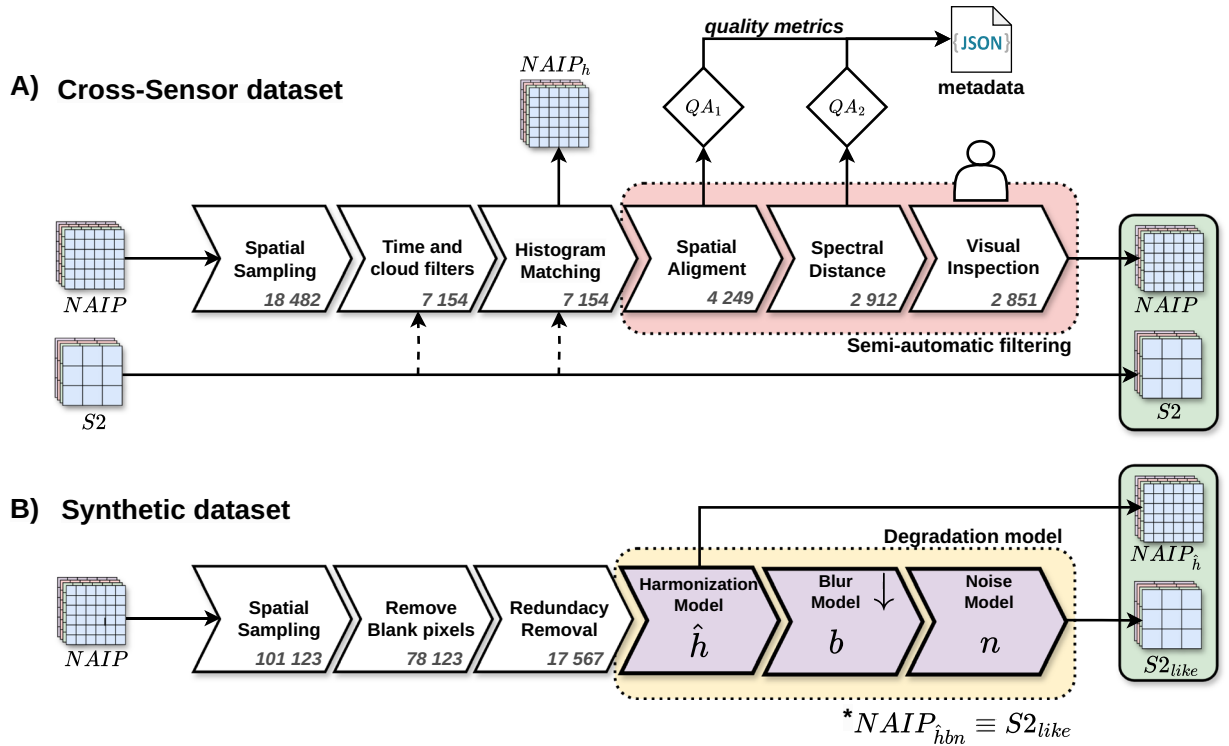


Fig. 3 A high-level summary of our workflow to generate SEN2NAIP datasets. The numbers represent the ROIs retained after each processing block. **(A)** The workflow for the cross-sensor SEN2NAIP: we assess one spatial quality metric (QA_1) and one spectral quality metric (QA_2) for each LR-HR pair. **(B)** The workflow for the synthetic SEN2NAIP: purple blocks denote the degradation model trained in the cross-sensor dataset.

After reflectance values correction, we employ the *opensr-test* framework¹² to assess the spatial (QA_1) and spectral (QA_2) alignment of the $S2$ - $NAIP_h$ image pairs. The QA_1 spatial quality metric is estimated by calculating the mean absolute error (MAE) between the ground control points identified by the LightGlue¹⁹ and DISK²⁰ algorithms. On the other hand, the QA_2 spectral quality metric is estimated by the average spectral angle distance of each band. We discard image pairs with $QA_1 > 1$ pixel or $QA_2 > 2$ deg to ensure minimal distortion due to external factors affecting super-resolution. Finally, to guarantee the dataset’s highest quality, we meticulously examine the remaining $S2$ - $NAIP$ pairs via a visual inspection review process. Any $S2$ - $NAIP$ pairs identified with saturated or defective pixels and any noticeable inconsistencies observed during the visual inspection were excluded.

A realistic degradation model. SR addresses an inverse problem, forcing the reconstruction of the original HR image from its degraded LR counterpart. This task is naturally ill-posed, as several potential HR images can correspond to a single LR image. The relationship between the LR (i.e., $S2$) and HR (i.e., $NAIP$) images conventionally is defined through the function:

$$S2_{like} = \Phi(NAIP; \theta) \downarrow_{bi} + n, \tag{1}$$

where Φ is the blurring function, \downarrow_{bi} is the bilinear downsampling operator, and n is the additive noise pattern. In a cross-sensor setting, the reflectance estimation can significantly vary between sensors. Therefore, the degradation model must consider the spectral alignment between the two sensors to obtain realistic LR counterparts. We propose a slightly variation of the Eq. (1) as:

$$NAIP_{\hat{h}} = \Upsilon(NAIP) \tag{2}$$

$$S2_{like} = \Phi(NAIP_{\hat{h}}; \theta) \downarrow_{bi} + n \tag{3}$$

where Υ is a harmonization model that transforms $NAIP$ into $NAIP_{\hat{h}}$ that closely match the reflectance values of $S2$. Utilizing the cross-sensor dataset, we obtain the Υ , Φ , and n models that emulate $S2$ imagery (i.e., $S2_{like}$) from $NAIP$. We randomly split the cross-sensor dataset into train and test, with 80% of the data allocated for training and the remaining 20% for testing. To learn the noise pattern n , we include cloud-free $S2$ images from the CloudSEN12¹⁷ dataset. As noise is an inherent attribute of $S2$ images, the $NAIP$ counterpart is unnecessary.

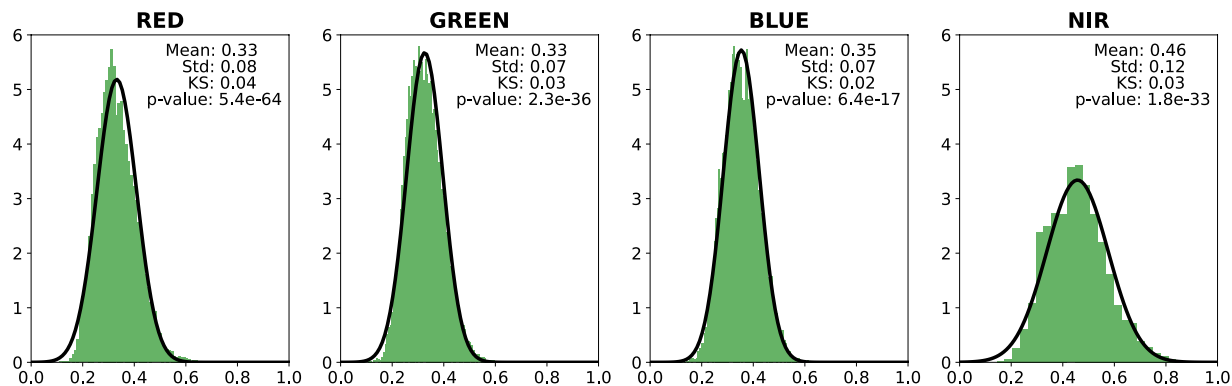


Fig. 4 Probability density function of the optimal gamma factor for each spectral band. The mean (Mean) and standard deviation (Std) are indicated for each band, along with the Kolmogorov-Smirnov (KS) statistic and p-value.

Harmonization model. One of the key challenges in scaling synthetic HR-LR datasets for super-resolution is the need for an LR reference at the same acquisition time. To overcome this, we propose three methods (i.e., statistical, deterministic, and variational) for correcting *NAIP* reflectance values without requiring an *S2* reference. Converting *NAIP* into *S2* can be seen as an “image color transfer” task. While a perfect match is not required, the harmonization must be accurate enough to “trick” the SR models into interpreting *NAIP*-degraded images as real *S2* data. This preprocessing step helps create HR-LR pairs without reflectance shifts, avoiding the problems shown in Fig. 1.

In the *statistical* approach, we determine the best gamma correction for each RGBNIR band within the pair of *NAIP* and *S2* images, setting a four-dimensional gamma correction vector. Next, we use a multivariate normal distribution to model the distribution of gamma values (see Fig. 4). During inference, the $NAIP_{\hat{i}}$ is obtained by applying the 50th-percentile gamma correction values to the original *NAIP* image bands to obtain $NAIP_{\hat{i}}$.

Figure 4 displays the density distributions of the optimal gamma factor for each spectral band. The Red, Green, and Blue panels show similar mean values, around 0.33 to 0.35, with standard deviations ranging from 0.07 to 0.08. The NIR band, however, has a notably higher mean value of 0.46 and a larger standard deviation of 0.12, indicating a higher degree of variability. The Kolmogorov-Smirnov (KS) statistic values are small, suggesting a good fit with the expected distribution.

For the *deterministic* model (Fig. 5A), we train a U-Net²¹ architecture with EfficientNet-B0²² as its backbone. The input data consists of *NAIP* imagery degraded to a 10-meter ($NAIP_{10m}$) resolution using simple bilinear interpolation with an anti-aliasing filter (i.e., a triangular blurring function). The target data comprises *S2* imagery. During inference, $NAIP_{\hat{i}}$ is obtained by a three-step process: (1) degrade *NAIP* to 10 m; (2) use the U-Net to estimate a harmonized version with reflectance closer to *S2*; and (3) apply histogram matching to correct the original *NAIP* reflectance values based on the U-Net output. Deep learning models for image-to-image tasks often introduce blurriness or lose fine details such as textures²³, so we use the U-Net predictions to just fine-tune the original *NAIP* reflectance values (Fig. 5A).

For the *variational* model (Fig. 5B), we disaggregate each band from *S2* and $NAIP_{10m}$ into a 1D Tensor containing the number of values inside a histogram bin. Each histogram was structured into 120 bins from 0 to 1, transforming an image pair into two tensors with dimensions of (4, 120). Then, we use this transformed version of the dataset to train a variational autoencoder²⁴ (VAE) that learns to transform the histogram of *NAIP* into the histogram of *S2*. For inference, $NAIP_{\hat{i}}$ is obtained in a four-step process. First, we obtain $NAIP_{10m}$. Second, we obtain the histograms for each band. Third, we utilize the trained VAE to obtain the $S2_{like}$ histogram. Fourth, we adjust the original *NAIP* reflectance values using the $S2_{like}$ histogram.

Blurring model. For the blurring model Φ , we fine-tune the width of the Gaussian degradation kernel by comparing *S2* imagery and $NAIP_{\hat{i}}$ using the MAE metric. We apply a specific Gaussian kernel to blur the $NAIP_{\hat{i}}$ image, then downsample it using bilinear interpolation without anti-aliasing filtering. In Fig. 6, we present the error curves (MAE) for the RGBNIR bands. The best sigma values are 3.0, 2.9, 2.9, and 3.4 for the RGBNIR bands, respectively.

Noise model. To accurately capture the noise complexity of *S2* images, we propose using the PD-Denoising model²⁵ to predict the noise distribution (Supplementary Figure S1A). We applied the PD-Denoising model to all the RGB cloud-free Sentinel-2 images in the CloudSEN12 dataset¹⁷. This allowed us to create a reflectance-to-noise matrix (Supplementary Figure S1B). This matrix correlates the reflectance values to empirical noise distributions, providing insight into varying noise characteristics. During the inference phase, the noise pattern, denoted as n , is generated by sampling from the reflectance-to-noise matrix for each reflectance value in the LR image. However, due to the computational cost of the previous method, we also proposed a simpler model. This alternative utilizes a Gaussian noise pattern with a zero mean and a standard deviation of 0.012, which represents the

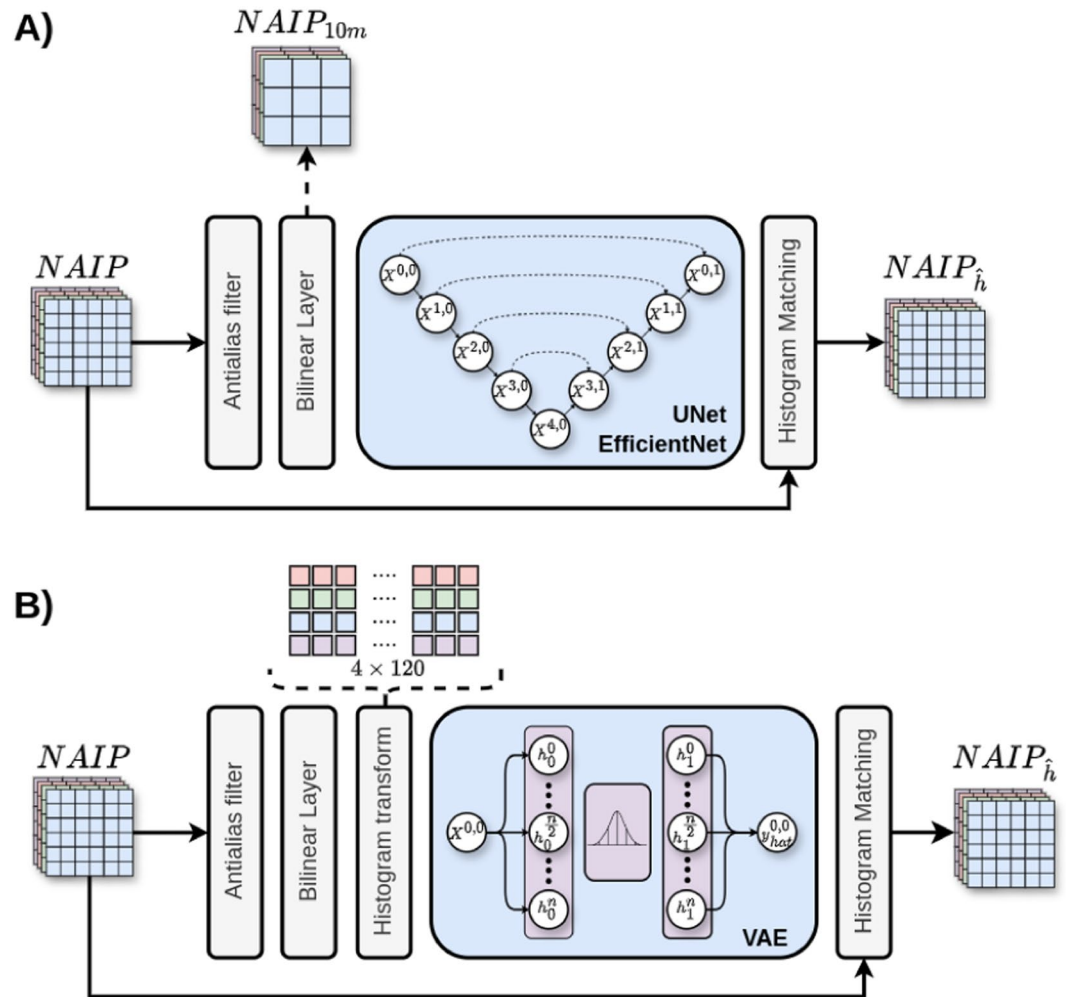


Fig. 5 Flowchart diagrams for the deterministic (A) and variational (B) harmonization approaches.

average standard deviation across various reflectance ranges (bins of 0.05) derived from the signal-to-noise matrix shown in Supplementary Figure S1B. The final noise is then scaled proportionally to the square root of the mean squared reflectance value for each pixel, ensuring that the noise level adapts to the intensity of the signal.

Synthetic dataset. The synthetic dataset does not require a pre-existing LR pair; instead, it is generated using the model trained in the previous section. As in the cross-sensor dataset (Fig. 3B), we use hard-core point modeling to produce 101,123 randomly generated ROIs across the contiguous United States. By removing the need for NAIP images to be paired with simultaneous images from S2, we are able to generate a much larger dataset. Each ROI contains an *early* (oldest) and *late* (most recent) NAIP image to track land cover changes.

While performing visual inspections on the cross-sensor dataset, we observed that many NAIP images had blank values. This issue occurred because the sampling was too close to the NAIP scene borders. To tackle this problem, we designed a straightforward yet effective blank identification system based on a 19×19 kernel to calculate image variance. Any *early* or *late* NAIP image with a zero variance was identified as containing blank values, and subsequently, the ROI was removed. This filtering resulted in a 22.7% reduction in our dataset.

The NAIP data program focuses on capturing images of agricultural areas during the growing seasons, potentially leading to a biased sampling distribution. To enhance the diversity of the dataset, we utilize a semantic clustering filter that integrates Contrastive Language-Image Pre-training²⁶ (CLIP) and MiniBatchKMeans²⁷. Recent studies have demonstrated the efficacy of CLIP in semantic understanding of super-resolved remote sensing imagery⁹. Our approach uses CLIP with MiniBatchKMeans, a KMeans algorithm variant optimized for large datasets, to cluster images based on semantic similarities. We establish an 18,000-cluster framework and randomly select one image per cluster, resulting in a 77.5% reduction in the dataset (Fig. 7). While the reduction percentage may seem excessive, the dataset is still well-spatially distributed (Fig. 2) and extensive, with 17,567 ROIs and 35,134 images (approximately 170 billion pixels).

Finally, for each NAIP *early* and *late* images, we generated three $NAIP_{\hat{h}}-S2_{like}$ image pairs using the degradation model presented in Eq. (2).

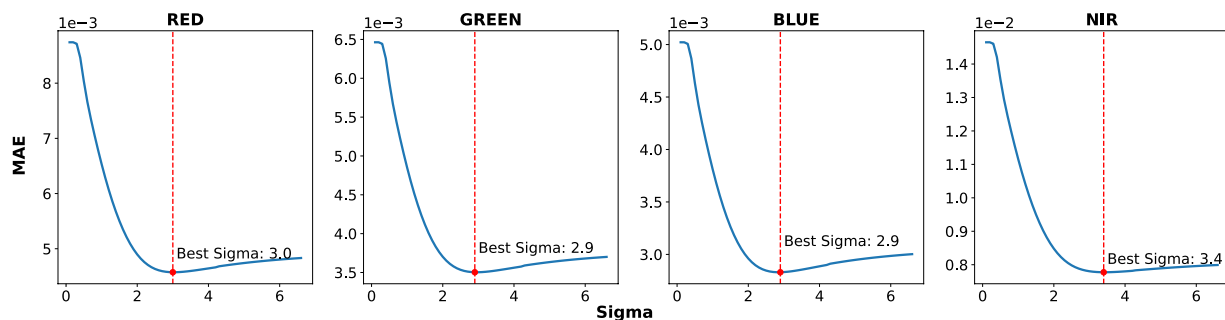


Fig. 6 Error curves (MAE) for the RGBNIR bands: associated error (y-axis) for each Gaussian kernel's sigma value (x-axis). The optimal sigma value is highlighted by a dashed red vertical line.

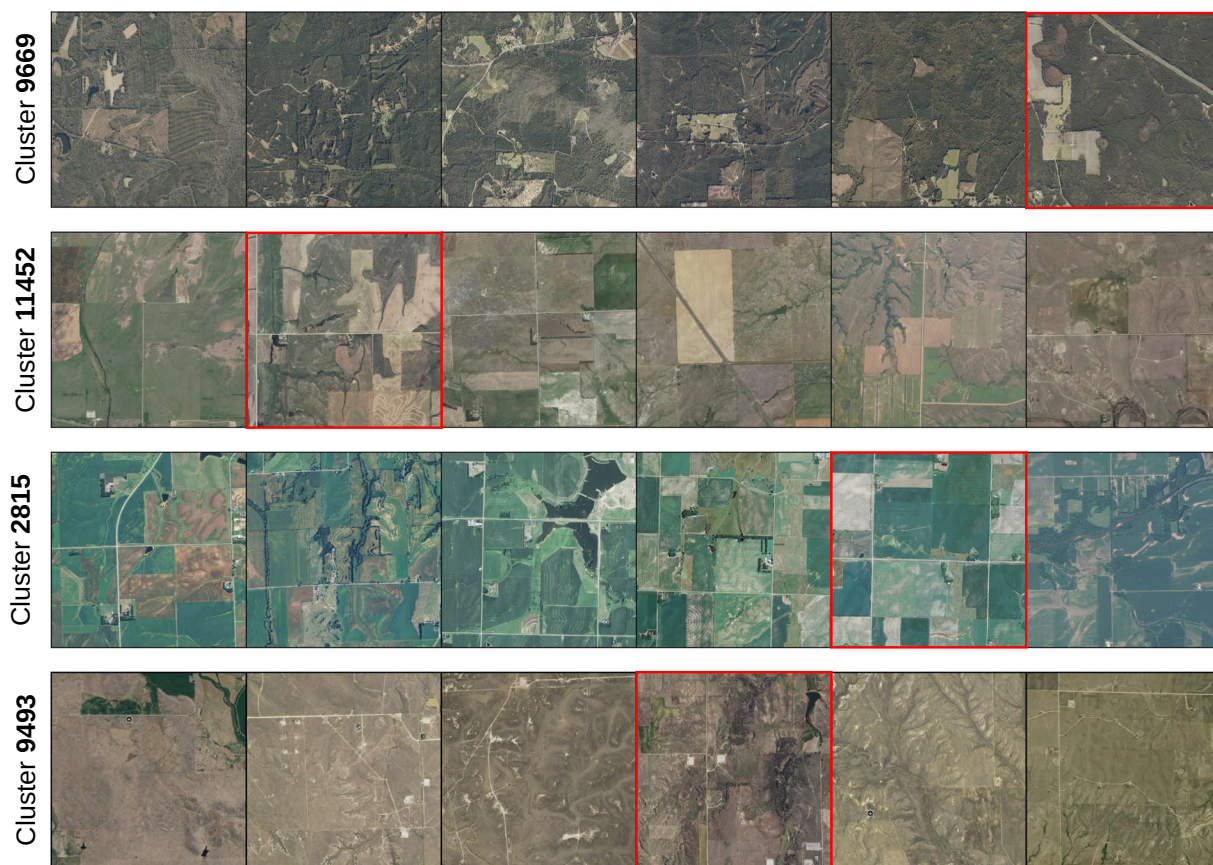


Fig. 7 Examples of the clusters formed. The highlighted image in red indicates the one chosen for the SEN2NAIP dataset.

Data Records

The dataset is available online via Science Data Bank²⁸. It is divided into two main sets: *cross-sensor* and *synthetic* (Fig. 8). These sets contain a total of 20,508 regions of interest (ROIs), with the *cross-sensor* set having 2,851 ROIs and the *synthetic* set having 17,657 ROIs.

Within the *cross-sensor* directory, every ROI subfolder contains two images: LR.tif and HR.tif. These images are stored in GeoTIFF format and contain the RGBNIR spectral bands. The LR image refers to the S2 imagery at 10 meters and an image shape of $4 \times 121 \times 121$ pixels. The HR.tif is a downsampled version of NAIP imagery, with a resolution of 2.5 meters to set a $4 \times$ scaling factor w.r.t. Sentinel-2. The *synthetic* directory's structure is slightly different. Each ROI folder contains two subfolders, *early* and *late*, each corresponding to different time acquisitions. The NAIP images maintain the RGBNIR spectral bands; however, unlike the *cross-sensor* dataset, their dimensions are set at $4 \times 1100 \times 1100$. Only histograms with 120 bins per spectral band were stored for each degradation model (statistical, deterministic, and variational) to prevent the dataset from becoming too large. Finally, Table 1 provides a detailed overview of the metadata for the acquired images, which are stored in the JSON format as metadata.json.

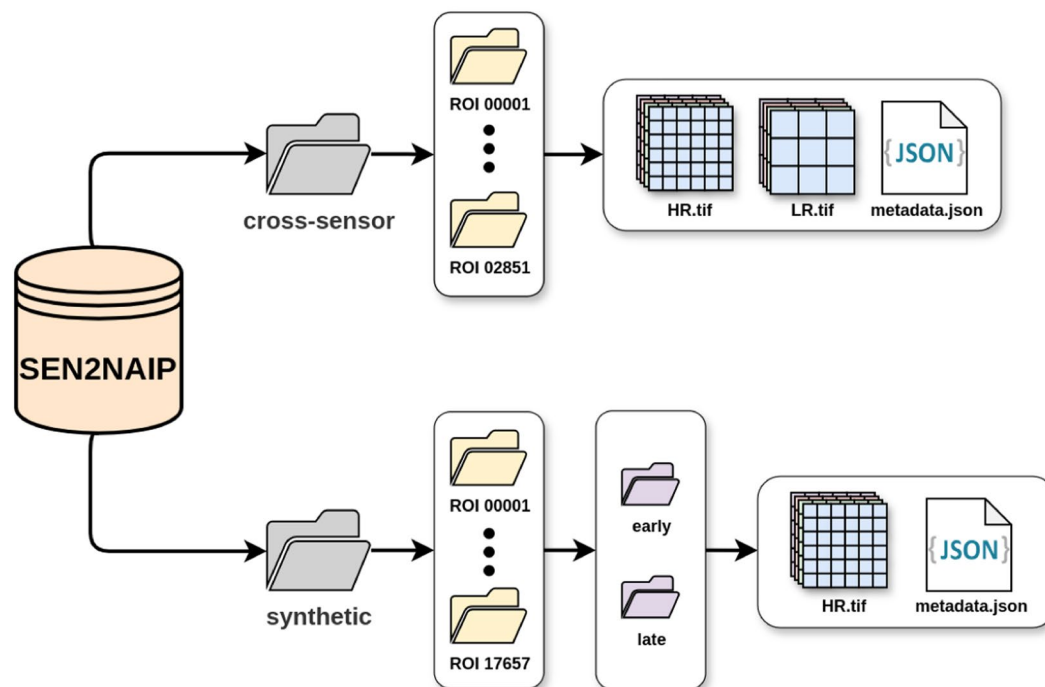


Fig. 8 The SEN2NAIP dataset is organized into a hierarchical folder structure. The dataset is divided into cross-sensor and synthetic images at the top level (gray folders). The subsequent level (denoted by yellow folders) arranges the images based on their geographic location (ROI). Within the synthetic division, an additional split encompasses early and late images, each characterized by the time acquisition. Finally, each folder contains an LR-HR pair with a metadata.json file that details the specifics of the data.

| Field name | Description |
|----------------------------|--|
| roi_id | The region of interest ID. |
| test | Whether the ROI is part of training (train) or testing (test) dataset. |
| s2_id | Sentinel-2 product ID. |
| s2_date | Sentinel-2 acquisition date in ISO format. |
| s2_sen2cor_version | Sen2Cor configuration baseline used at the time of the product generation. |
| s2_view_off_nadir | The angle from the Sentinel-2 sensor between nadir and the scene center. |
| s2_view_sun_azimuth | Sentinel-2 sun azimuth angle. |
| s2_view_sun_elevation | Sentinel-2 sun elevation angle. |
| naip_id | NAIP product ID. |
| naip_date | NAIP acquisition date in ISO format. |
| proj:epsg | EPSG code. |
| proj:geometry | Footprint of the ROI. |
| proj:centroid | Centroid coordinates of the ROI in latitude and longitude. |
| proj:transform | The affine transformation coefficients. |
| QA_1 | Only available in the cross-sensor subset. The spatial quality metric. |
| QA_2 | Only available in the cross-sensor subset. The spectral quality metric. |
| $d_{sim \rightarrow real}$ | Only available in the synthetic subset. The histograms to obtain $NAIP_{hat}$ from raw NAIP imagery. |

Table 1. Metadata associated with each LR-HR image pair.

Technical Validation

Validation of synthesized Sentinel-2-like imagery. We evaluate the precision of degradation models (statistical, deterministic, and variational) in converting NAIP into S2 images. We consider different percentile values (10, 25, 50, 75, 90) for the statistical approach to assess the model sensitivity. The efficacy of the transformation is quantitatively summarized in Table 2, which presents a comparative analysis between $S2_{like}$ and S2. This analysis is conducted on the test subset derived from the cross-sensor dataset. To evaluate the performance of the degradation models comprehensively, we employ five metrics: Pearson correlation, which measures the linear correlation; Spectral angle distance, which assesses the similarity in the spectral signatures by computing the

| Method | Pearson Correlation | Spectral Angle Distance | PBIAS | MAE | RMSE |
|-------------------------------|---------------------|-------------------------|--------------|--------------|--------------|
| Raw | 0.717 | 0.389 | 3.381 | 0.344 | 0.336 |
| Statistical - 10th percentile | 0.716 | 0.366 | 0.562 | 0.095 | 0.075 |
| Statistical - 25th percentile | 0.722 | 0.333 | 0.795 | 0.078 | 0.060 |
| Statistical - 50th percentile | 0.726 | 0.316 | 1.073 | 0.076 | 0.059 |
| Statistical - 75th percentile | 0.727 | 0.312 | 1.384 | 0.092 | 0.077 |
| Statistical - 90th percentile | 0.726 | 0.318 | 1.743 | 0.129 | 0.114 |
| Deterministic | 0.972 | 0.139 | 1.063 | 0.040 | 0.033 |
| Variational | 0.926 | 0.288 | 0.936 | 0.063 | 0.052 |

Table 2. Comparative Analysis between S_2 and $S_{2_{like}}$ generated by the three proposed methods: statistical, deterministic, and variational. From the statistical method, various percentile values were examined.

angle between their vectors; Percentage Bias (PBIAS), which quantifies the average tendency of the degraded images to overestimate or underestimate; and mean absolute and root mean squared errors (MAE and RMSE), which measure the average absolute and squared differences, respectively. We estimated these metrics per image and then averaged the values to report a global value.

On average, the raw *NAIP* images exhibit reflectance values approximately three times greater than those of S_2 images, as evidenced by the PBIAS metric in Table 2. This difference is due to the different preprocessing steps each sensor uses. *NAIP* images undergo a Bidirectional Reflectance Distribution Function (BRDF) correction, which adjusts reflectance based on the statistical properties of the image before mosaicking. In contrast, S_2 images skip BRDF correction because automatic methods can be unreliable in certain locations and atmospheric conditions²⁹.

Based on the evaluation of various metrics, the deterministic approach that combines U-Net adaptation with histogram matching proved to be the most accurate method. It accurately simulates reflectance intensity (PBIAS) and successfully preserves the spectral signatures of S_2 images (SAD). In contrast, the statistical method could only produce stable results at the 50th percentile. This method was able to represent reflectance intensity, but it failed to preserve the inter-band relationships, as shown by the SAD, RMSE, and MAE metrics. Meanwhile, the variational approach yields results almost as accurate as the deterministic method but slightly inferior. However, it has the potential to include aleatory variations, which could be useful in enhancing the generalization capacity of SR networks. Supplementary Figure S1B complements the analysis by illustrating the reflectance distribution for each band across the different proposed degradation methods.

Figures 9 and 10 depict two common scenarios in the SEN2NAIP dataset. The deterministic approach demonstrates good performance in both images. However, other methods may also serve as appropriate degradation models under specific scenarios. For instance, in environments with minimal spatial autocorrelation, the statistical method at the 50th percentile often produces results comparable to the deterministic approach, as shown in Fig. 9. On the other hand, the variational method typically excels in high-contrast scenes, as illustrated in Fig. 10. These findings suggest that combining all degradation methods simultaneously may be beneficial when training SR networks to increase the dataset diversity.

Real-world SR network training. In this section, we explore the impact of various degradation models on the super-resolution of S_2 imagery. To ensure a fair comparison, we maintained the same SR lightweight network and identical training setup, altering only the degradation model used. The SR network is a simplified version of the VGG model, where the max-pooling layers were removed. We utilize the Adam optimizer, setting β_1 to 0.9 and β_2 to optimize a composite MAE plus adversarial loss function. To differentiate between real and super-resolved images, we utilized the PatchGAN discriminator. This discriminator assesses 16×16 patches, aggregating their scores to determine whether the image is *real* or *fake*. We use the S_2NAIP synthetic dataset for training, where the LR inputs and HR targets are represented by $S_{2_{like}}$ and $NAIP_h$, respectively. We have established five distinct degradation models (dataloader configurations):

- *Raw*: This configuration does not use a harmonization model to create $NAIP_h$. The $S_{2_{like}}$ pair is generated by applying an anti-aliasing filter, which is applied by default in PyTorch, and then using bilinear downsampling. This approach provides a baseline of not applying a harmonization model.
- *Statistical*: The $NAIP_h$ is derived by applying the statistical model set at the 50th percentile. This configuration offers insight into the impact of adjusting solely the reflectance intensity.
- *Variational*: This configuration utilizes the vanilla VAE model to generate $NAIP_h$ (Fig. 5B).
- *Deterministic*: Here, $NAIP_h$ is produced by applying the trained U-Net model (Fig. 5A).
- *All*: In this approach, all degradation models are employed together, with their application randomized within the dataloader.

In all dataloaders except for Raw, $S_{2_{like}}$ is produced using the pre-trained Gaussian blurring model and bilinear downsampling (with no anti-aliasing filter) on $NAIP_h$. Additionally, simple Gaussian random noise, as described in the noise modeling section, is incorporated into the process. We repeat the training for each degradation model five times and report the best result. To conduct a thorough evaluation, we utilized two distinct

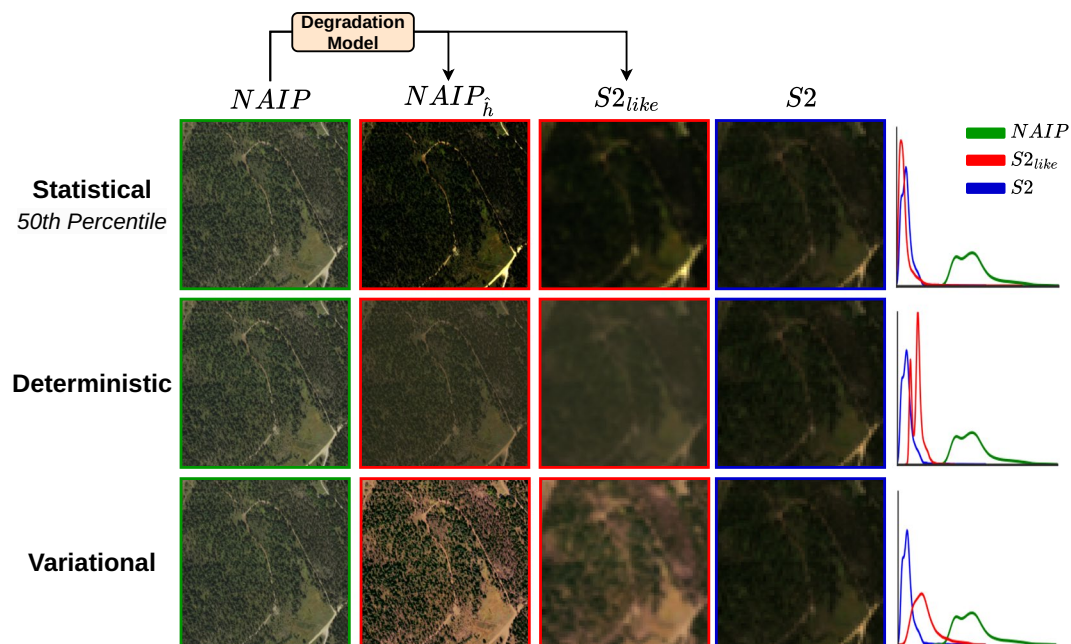


Fig. 9 Comparative visualization of three image degradation methods applied to NAIP imagery against Sentinel-2 data, with the final column showing the histograms of pixel values.

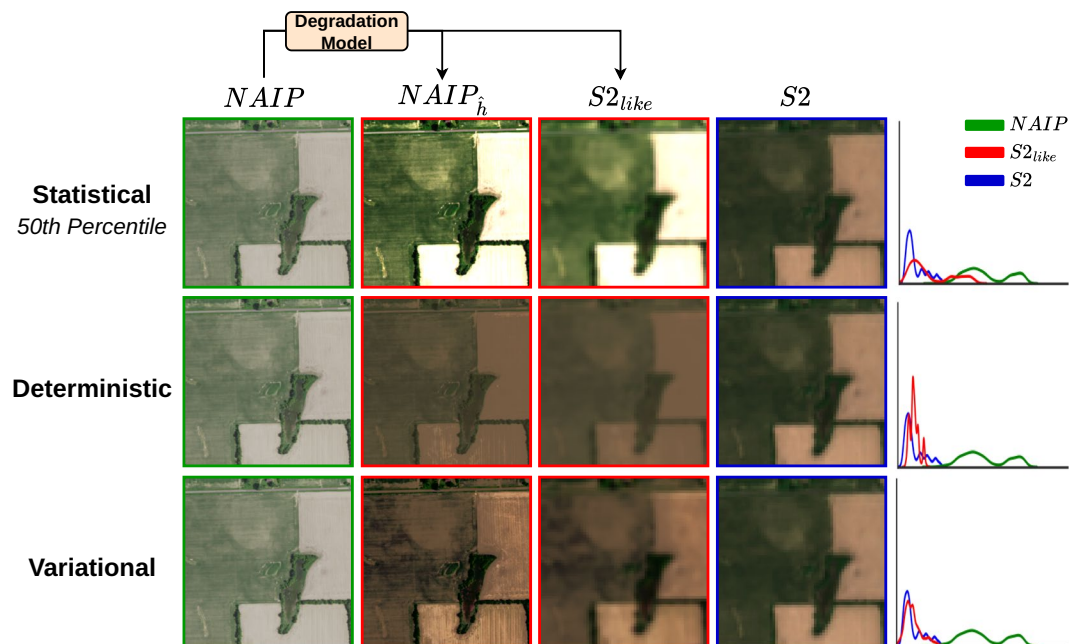


Fig. 10 Comparative visualization of three image degradation models applied to NAIP imagery against Sentinel-2 data (high-contrast example), with the final column showing the histograms of pixel values.

cross-sensor datasets that were not utilized during the degradation model's training. These independent datasets consist of S2-SPOT and S2-NAIP image pairs, as described in the *opensr-test* paper¹².

To assess the impact of the degradation model in super-resolution, we focus on the omission metric within the *opensr-test* framework. The omission denotes the missing high-frequency pattern that the SR model failed to incorporate¹². The higher the omission score, the more conservative the SR network is.

The findings are summarized in Fig. 11, revealing that although all the SR networks perform similarly well on synthetic test data (see the first panel in Fig. 11), their effectiveness varies considerably when tested on cross-sensor datasets. The SR network trained with the raw dataloader struggles to introduce high-frequency details in real S2 imagery, resulting in outputs that closely correspond to a simple bilinear interpolation (Fig. 12).

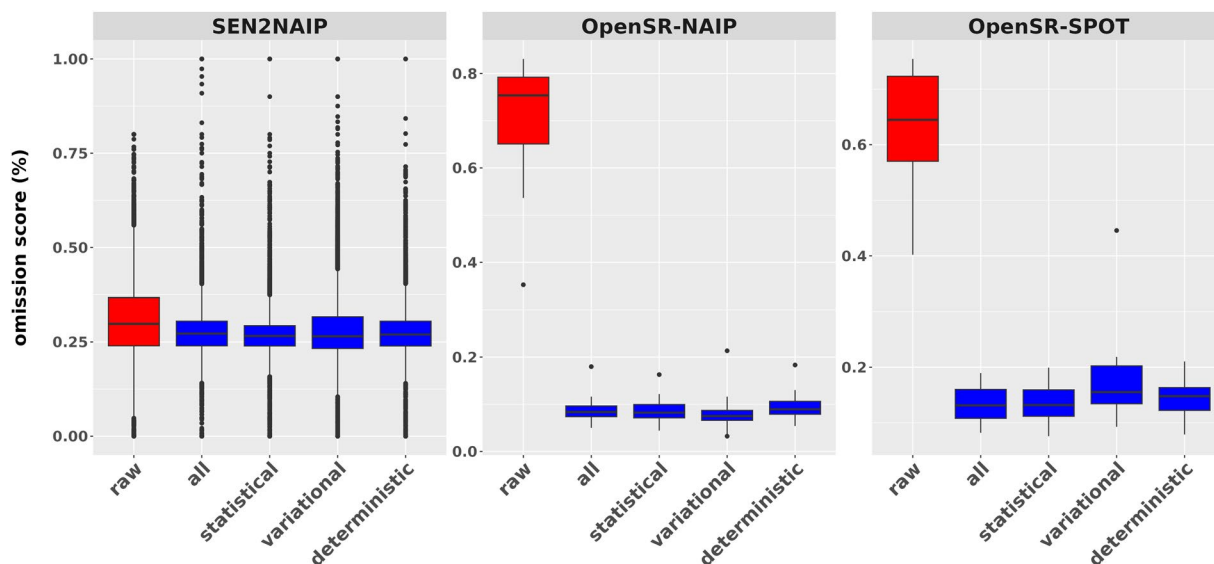


Fig. 11 Super-Resolution performance on SEN2NAIP (synthetic), OpenSR-NAIP (cross-sensor), and OpenSR-SPOT (cross-sensor) datasets. The red boxplot is the baseline showing the score value if no harmonization model is applied.

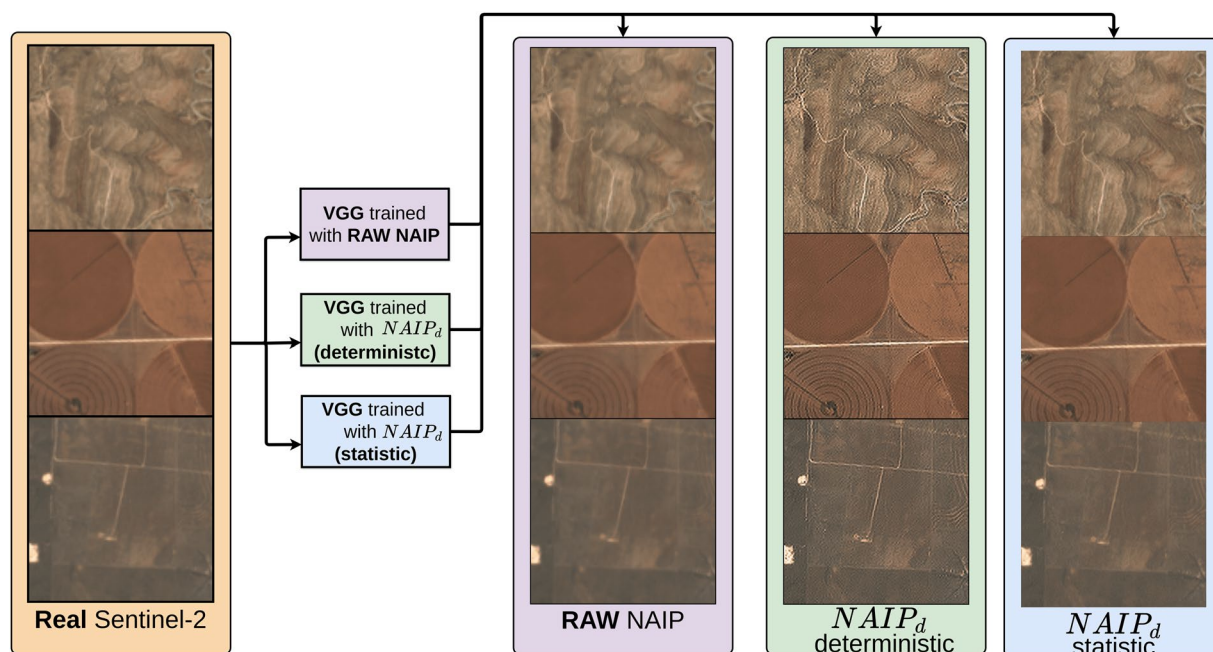


Fig. 12 Comparison of degradation models on SR networks: The first column shows the S2 image, followed by the results of the SR network trained with the raw $NAIP$ images. The subsequent columns display results using the deterministic degradation model and the statistical degradation approach, respectively.

After visually testing the networks on several S2 images, we found that the three proposed harmonization methods performed well in most cases, yielding similar results. However, the statistical method encountered some issues with certain images. For example, as shown in Fig. 12, the deterministic model introduced high-frequency details with greater intensity in S2 images compared to the statistical approach. The variational method and “all” approaches produced results comparable to the deterministic model.

This experiment is a proof-of-concept, focusing not on the quality of high-frequency details added but on how the models respond to S2 imagery. These findings emphasize the importance of aligning the degradation model with the characteristics of real-world sensors to ensure reliable SR outputs during inference. Readers should be aware that these outcomes are based on a simple SR network and may change with more advanced models. We plan to test these degradation models with larger SR networks and alternative architectures, such as

transformers³⁰ or recurrent linear units³¹. This study underscores the need for more research on developing more efficient degradation models that accurately mimic the characteristics of one sensor using another. Doing so could unlock the full potential of SR networks for real-world imagery beyond just NAIP data.

Usage Notes

This study introduces SEN2NAIP, a comprehensive dataset designed for realistic S2 super-resolution. It comprises images from NAIP and S2, covering the entire contiguous United States. The dataset's total volume is 167.21 GB. To facilitate easy access, we offer a Python script that streamlines the process of batch downloading the dataset³². Users should be aware that the majority of NAIP data is collected from June to August and only over the United States, which may influence model performance on imagery from other seasons and regions, especially in areas with snow cover.

Code availability

We have made the degradation models publicly available, as well as the SR models used for technical validation. Additionally, we provide code to extend the dataset for specific land use cases if needed. For more information, please refer to our GitHub repository (<https://github.com/esaOpenSR/opensr-degradation/>).

Received: 9 May 2024; Accepted: 2 December 2024;

Published online: 18 December 2024

References

- Masoud, K. M., Persello, C. & Tolpekin, V. A. Delineation of agricultural field boundaries from Sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks. *Remote sensing* **12**, 59 (2019).
- Zhang, T. *et al.* FSRSS-Net: High-resolution mapping of buildings from middle-resolution satellite images using a super-resolution semantic segmentation network. *Remote Sensing* **13**, 2290 (2021).
- Michel, J., Vinasco-Salinas, J., Inglada, J. & Hagolle, O. Sen2ven μ s, a dataset for the training of sentinel-2 super-resolution algorithms. *Data* **7**, 96 (2022).
- Dong, R., Mou, L., Zhang, L., Fu, H. & Zhu, X. X. Real-world remote sensing image super-resolution via a practical degradation model and a kernel-aware network. *ISPRS Journal of Photogrammetry and Remote Sensing* **191**, 155–170 (2022).
- Xia, G.-S. *et al.* Structural high-resolution satellite image indexing. In *ISPRS TC VII Symposium-100 Years ISPRS* **38**, 298–303 (2010).
- Cheng, G., Han, J. & Lu, X. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE* **105**, 1865–1883 (2017).
- Xia, G.-S. *et al.* AID: A benchmark data set for performance evaluation of aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing* **55**, 3965–3981 (2017).
- Cornebise, J., Oršolić, I. & Kalaitzis, F. Open high-resolution satellite imagery: The worldstrat dataset—with application to super-resolution. *Advances in Neural Information Processing Systems* **35**, 25979–25991 (2022).
- Wolters, P., Bastani, F. & Kembhavi, A. Zooming Out on Zooming In: Advancing Super-Resolution for Remote Sensing. *arXiv preprint arXiv:2311.18082* (2023).
- Razzak, M. T. *et al.* Multi-spectral multi-image super-resolution of Sentinel-2 with radiometric consistency losses and its effect on building delineation. *ISPRS Journal of Photogrammetry and Remote Sensing* **195**, 1–13 (2023).
- Cresson, R. SR4RS: A tool for super resolution of remote sensing images. (2022).
- Aybar, C., Montero, D., Donike, S., Kalaitzis, F. & Gómez-Chova, L. A Comprehensive Benchmark for Optical Remote Sensing Image Super-Resolution. *IEEE Geoscience and Remote Sensing Letters* **21**, 1–5 (2024).
- Chuvieco, E. *Fundamentals of satellite remote sensing: An environmental approach* (CRC press, 2020).
- Sabins Jr, F. F. & Ellis, J. M. *Remote sensing: Principles, interpretation, and applications* (Waveland Press, 2020).
- Camps-Valls, G., Tuia, D., Zhu, X. X. & Reichstein, M. *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences* (John Wiley & Sons, 2021).
- Baddeley, A., *et al.* Package 'spatstat'. *The Comprehensive R Archive Network* () 146 (2014).
- Aybar, C. *et al.* CloudSEN12, a global dataset for semantic understanding of cloud and cloud shadow in Sentinel-2. *Scientific data* **9**, 782 (2022).
- Aybar, C., Wu, Q., Bautista, L., Yali, R. & Barja, A. rgee: An R package for interacting with Google Earth Engine. *Journal of Open Source Software* **5**, 2272 (2020).
- Lindenberger, P., Sarlin, P.-E. & Pollefeys, M. LightGlue: Local Feature Matching at Light Speed. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 17627–17638 (2023).
- Tyszkiewicz, M., Fua, P. & Trulls, E. DISK: Learning local features with policy gradient. *Advances in Neural Information Processing Systems* **33**, 14254–14265 (2020).
- Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* **18**, 234–241 (Springer, 2015).
- Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114 (PMLR, 2019).
- Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging* **3**, 47–57 (2016).
- Doersch, C. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908* (2016).
- Zhou, Y. *et al.* When awgn-based denoiser meets real noises. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, 13074–13081 (2020).
- Radford, A. *et al.* Learning transferable visual models from natural language supervision. In *International conference on machine learning*, 8748–8763 (PMLR, 2021).
- Sculley, D. Web-scale k-means clustering. In *Proceedings of the 19th international conference on World wide web*, 1177–1178 (2010).
- Aybar, C. *et al.* SEN2NAIP: A large-scale dataset for Sentinel-2 Image Super-Resolution, <https://doi.org/10.57760/sciencedb.17395> (2024).
- Montero, D., Mahecha, M. D., Aybar, C., Mosig, C. & Wieneke, S. Facilitating advanced Sentinel-2 analysis through a simplified computation of Nadir BRDF Adjusted Reflectance. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-4/W12-2024*, 105–112 (2024).
- Vaswani, A. Attention is all you need. *Advances in Neural Information Processing Systems* (2017).
- Gu, A. & Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752* (2023).
- Aybar, C. *et al.* SEN2NAIP Dataset and scripts, <https://huggingface.co/datasets/isp-uv-es/SEN2NAIP> (2024).

Acknowledgements

The Sentinel-2 Level 2 A and NAIP data have been generously provided by ESA and USDA, respectively. This work has been funded by the European Space Agency (ESA, Φ -lab, OpenSR project). Cesar Aybar and Julio Contreras acknowledge support by CONCYTEC, Peru (“Proyectos de Investigación Básica – 2023-01” program, PE501083135-2023-PROCIENCIA). Luis Gómez-Chova acknowledges support from the Spanish Ministry of Science and Innovation (PID2019-109026RB-I00 and PID2023-148485OB-C21 funded by MCIN/AEI/10.13039/501100011033).

Author contributions

C.A.: Methodology, Data curation, Validation, Writing - original draft preparation. D.M.: Formal analysis, Validation, Data curation. J.C.: Data curation, Visualization. S.D.: Methodology, Formal analysis, Validation, Data curation. F.K.: Writing - Review & Editing, Resources, Funding acquisition, Project supervision. L.G.C.: Conceptualization, Supervision, Writing - Review & Editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-04214-y>.

Correspondence and requests for materials should be addressed to C.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024