GAN-Based Dual Image Super Resolution for Satellite Imagery Decreasing Radiometric Uncertainty

Michael Greza [1], Indraditya Bhattacharya [2], Ludwig Hoegner [3], Boris Jutzi [4]

[1] Technical University of Munich, Photogrammetry & Remote Sensing, Munich, Germany – michael.greza@tum.de
[2] Technical University of Munich, Photogrammetry & Remote Sensing, Munich, Germany – indraditya.bhattacharyya@tum.de
[3] University of Applied Sciences Munich, IAMLIS, Munich, Germany – ludwig.hoegner@tum.de
[4] Technical University of Munich, Photogrammetry & Remote Sensing, Munich, Germany – boris.jutzi@tum.de

Keywords: Satellite Image, Super Resolution, GAN, Radiometry, Mixed Pixels, CubeSat.

Abstract

Super resolution for satellite imagery possesses specific challenges due to its unique feature geometry compared to classic computer vision. Specifically, satellite images contain a high amount of relatively small, distributed high-frequency features that are hard to preserve when sampling up to a higher resolution. General adversarial networks (GANs) are suitable for super resolution tasks but need special attention concerning the geometric and radiometric accuracy of the results. We propose a network for versatile satellite imagery super resolution (VSISR) that focuses on high-frequency detail preservation and radiometric consistency. As a novelty, it is able to utilize a reference high resolution image during inference to lower hallucinations without altering the source image’s radiometric characteristics.
A GAN that is already handling well high frequencies in standard computer vision cases is adapted for satellite imagery and supplemented with a mixed pixel approach for data augmentation. Training on a diverse RGB dataset from four satellite missions results in a versatile super resolution model that is optimized to preserve radiometric features and minimize hallucinations. Compared to other neural networks for satellite image super resolution in their respective datasets, VSISR performs in the middle regarding mean PSNR (25.30 dB) and SSIM (0.8098). Nevertheless, it is the first time that, using mixed pixel training on a comparably small dataset, this versatility concerning the satellite data source is achieved while maintaining their unique radiometric traits.

1. Introduction

The growing number of commercial small-satellite missions with a focus on Earth observation and remote sensing creates unique chances and challenges concerning data processing. On the one hand, the availability of data increases drastically, enabling the development of new methods and products such as deep-learning-based processing and daily updated plant health maps. On the other hand, new methods are needed to be able to process these vast amounts of data, concerning both data preprocessing, e.g. up to including Level-2 processing Sentinel-2 terminology (Meygret et al., 2009), and processing for products. As a reference for current data amounts, Planet Labs, as of September 2023, downlinked over 30 terabytes of data from circa 200 satellites daily (Breaking Defense, 2023). For this, a fully automated processing chain is imperative. Apart from commercial platforms, data from classic, publicly funded multispectral Earth observation satellites like Sentinel-2 or Landsat is partly available as open data. They bring along more advantages such as additional spectral channels in higher wavelengths but lack in geometrical resolution compared to many commercial options. To mitigate this detriment, it is possible to use super resolution deep learning approaches which were introduced in recent years. One popular option to approach deep learning for super resolution is via the usage of general adversarial networks (GANs) as proposed by (Goodfellow et al., 2014). Super resolution for satellite imagery possesses specific challenges due to its unique feature geometry when compared to classic computer vision. Specifically, satellite images contain a high amount of relatively small, distributed high-frequency features that are hard to preserve when sampling up to a higher resolution, especially in urban regions. This encompasses, among many others, building agglomerations in cities as a common example. These high-frequency regions are challenging for GANs as they are prone to hallucinate false details into the image.

We propose a network for versatile satellite imagery super resolution (VSISR) that focuses on high-frequency detail preservation through a mixed pixel approach and Dual Image Super Resolution. For this, a GAN is adapted that already focuses on high frequencies in standard computer vision cases and subsequently a model for satellite imagery is created with its respective dataset. Initially, the network is designed to scale up PAN-sharpened Landsat imagery with a ground sampling distance (GSD) of 15 m with Sentinel-2 reference imagery. The datasets are then augmented with scenes captured by further satellites, resulting in a versatile super resolution model that is optimized to preserve radiometric features and minimize hallucinations. An overview of VSISR’s network structure is given in Section 2, followed by a placement into the frame of the research project and its satellite mission in Section 3. Section 4 presents the data that was used within the development of VSISR. The results of the training are displayed in Section 5, rounded off with a discussion in Section 6.

2. Neural Network

In this section, an overview of the network and its training modalities is given. This is followed by its architecture, and a description of the combined adaptive loss in use together with the new mixed-pixel approach.

2.1 Overview and Dual Image Super Resolution

The basic architecture of VSISR is taken from SRGAN for photo-realistic single image super-resolution (Chan et al., 2017). The network’s structure is altered to improve performance on

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-155-2024 | © Author(s) 2024. CC BY 4.0 License.

satellite data and address its inherent abundance of small but equally important features. GANs use hallucinations of the generator to create new images. This is on the one hand a vital, intended effect but on the other hand may create details that are unwanted as they are artificial and not present in reality. Reliable results are important for most use cases utilizing satellite imagery. Apart from network structure, the loss function is adapted to improve the results of the combined adaptive loss on remote sensing imagery, and the concept of mixed pixel training is newly introduced, both mitigating unwanted hallucinations. To further improve on this, the network is also fed a high-resolution reference image during inference. As the whole Earth is covered by high-resolution imagery daily, it is possible to find a reference counterpart for almost every low-resolution scene captured, that is taken within a temporal offset of one day. This ensures low differences in contents apart from traffic movements. As these reference images are captured with a different sensor platform and under different conditions, the network is trained to ensure radiometric consistency with the original low-resolution input. We title this tandem system of original input image and reference scene in inference Dual Image Super Resolution.

2.2 GAN Architecture

The generator is created with convolutional layers, PReLU layers, and batch normalization. Combined with these, it utilizes multiple upscaling and residual blocks (He et al., 2016), enabling it to reconstruct high-quality images from low-resolution inputs while mitigating the risk of vanishing or exploding gradients. Figure 1 displays a condensed version of the architecture of the generator model. It contains the 16 residual blocks and the upsampling block. The generator encompasses 1,453,955 parameters, thereof 1,449,731 trainable.

The discriminator encompasses seven discriminator blocks containing convolutional layers, batch normalization and LeakyReLU to enhance its feature-discrimination capabilities and to handle sparse gradients. It includes dropout and L2 regularization, which helps to prevent overfitting and supports the model’s ability to remain generalizable across various sets of satellite images. The result of its final layer is binary, deciding if the image presented by the generator is a real image or an artificial one. Figure 2 displays a condensed version of the architecture of the discriminator model. It encompasses 107,455,297 parameters, thereof 107,451,585 trainable.

2.3 Combined Adaptive Loss and Mixed Pixel Approach

To create the combined adaptive loss, a binary cross-entropy loss function is used in the discriminator in conjunction with the mean square error of extracted VGG19 (Simonyan and Zisserman, 2015) features. The content loss is calculated between the ground truth high resolution sample and the generated super resolution image. The aim of integrating content loss is to suppress undesirable hallucinations of the GAN.

Mixed pixel training is introduced to dedicatedly support the network in spectral unmixing. When upscaling an, e.g., 10 m GSD Sentinel-2 pixel to 5 m, the radiometric content of four 5 m GSD pixels is contained in the bounds of the Sentinel-2 pixel. As VSISR is intended to not only work geometrically accurate but also radiometrically accurate, artificial mixed pixel images are created utilizing the generator of the model to provide a further layer of data augmentation. Utilizing these artificial images, the discriminator is trained with three sets of high-resolution images. The first one containing only real images and the second one containing only completely fake mixed pixel imagery. The third set contains blended images where the ratio of artificial mixed pixels within one image is controlled by an alpha value variable. For the current version of VSISR, 10 % of the pixels within a scene are artificially created mixed pixels.

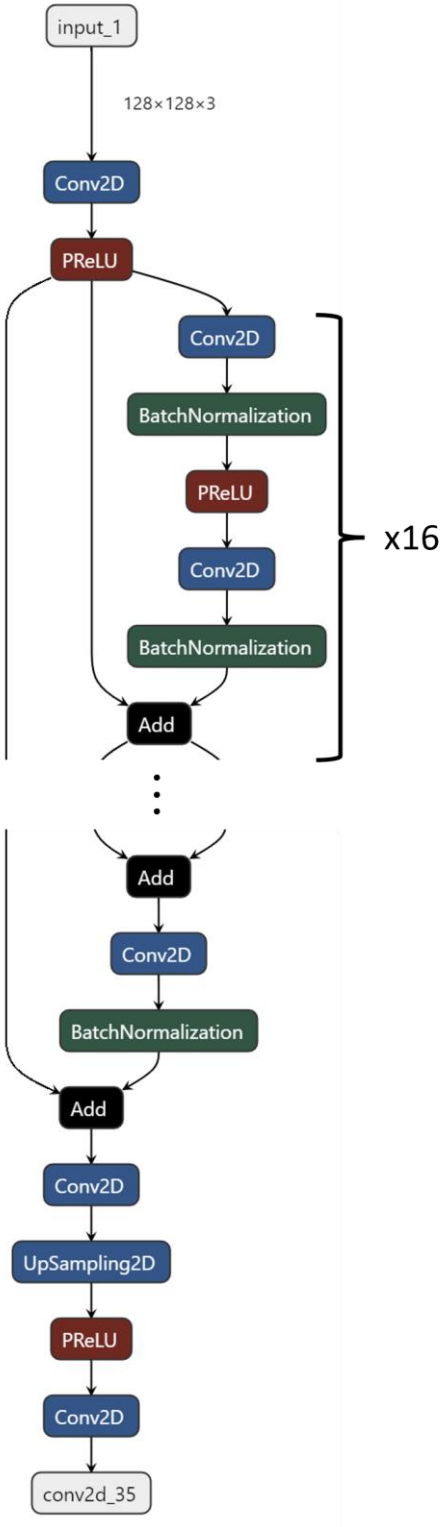


Figure 1. Architecture of the generator. Its main body consists of 16 residual blocks and one upscaling block that quadruples the number of pixels. This is equivalent to halving the GSD.

3. Project Frame & Processing Chain Relevance

Development of this network is part of the CuBy project. Within this project, a multispectral Earth observation satellite mission is being developed. The goal of the mission is to create a satellite system that is capable of monitoring the territory of the Bavarian state with a high temporal, geometrical and spectral resolution in a cost-efficient manner. Its main applications are agricultural and silvicultural biomonitoring. The first demonstrator mission phase encompasses five 6U CubeSats similar to the model of Figure 3 in a sun-synchronous three-day-repeat orbit with a planned local time of ascending node at 9-10 UTC. This ensures favorable and comparable acquisition circumstances with a high temporal resolution. Nominal orbit altitude is 470 km above the target region, resulting in a ground sampling distance of 4 m. The payload is a 12-bit multispectral CMOS pushbroom line scanner with seven narrow spectral channels and one PAN channel. PAN-sharpening is not available as the geometric resolution of the PAN channel is the same as the resolution of the other channels. The selection of the spectral filters is made to acquire imagery in wavelengths comparable to Sentinel-2, excluding SWIR. To enable a nominal mission duration of five years, the satellites are equipped with a propulsion system.

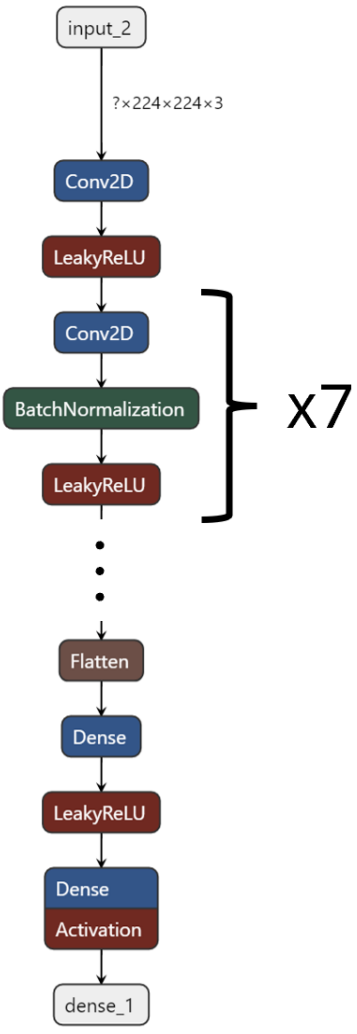


Figure 2. Architecture of the discriminator. Its main body consists of seven discriminator blocks and a dense block for classifying the results of the generator.

Due to the large amount of data that is produced by the system with every repeat cycle and its unique challenges resulting from the usage of a line imager, a fully automatic and robust processing chain needs to be developed. (Bauer et al., 2023) and (Greza et al., 2023) provide a mission overview, including a description of the image processing chain. A central aspect of a later stage in processing is the ability to detect building centers as tie points for a bundle block adjustment system that is comparable to (Spiegel, 2007) and in subsequent steps an orthra membrane adjustment described in (Roschlaub et al., 2023) and (Roschlaub et al., 2024). The building detection is automated through a deep neural network (Roschlaub, 2020). As it is vital to support the network’s ability to achieve as few false negative detections as possible, VSISR contributes to more clearly detectable buildings by improving the geometric resolution and the sharpness of the imagery. A ground sampling distance of four meters is suitable for building center detection in general, but centers of smaller buildings may be misplaced or even remain undetected without preprocessing.

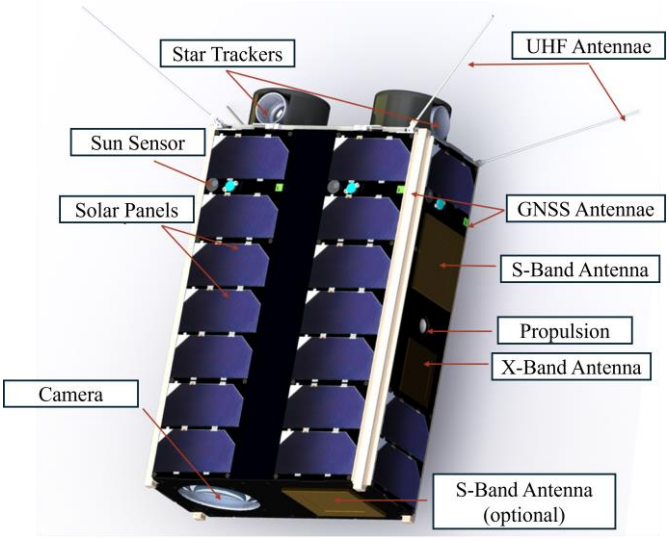


Figure 3. Illustration of a CuBy satellite. In total, five satellites will be launched within the first mission phase. The multispectral linescanner payload is located at the bottom of the satellite.

4. Dataset

This section presents an introduction to the composition of the dataset, illustrated by examples of the images. Further, quantitative, and qualitative measures are presented.

4.1 Overview

To train VSISR, image pairs of high- and low-resolution versions of the same scene are created. Only scenes with a temporal delay of one day at maximum between capturing the high- and low-resolution image are selected to ensure a high congruence between both sets. Additionally, only cloud-free scenes are chosen for the creation of the data set. Concerning geometry, one can assume that no drastic geometric changes between to scenes occur that are captured one day apart. The most prominent geometric changes in day-to-day scenes are cars that are relocated. This cannot be addressed in the current version of our network but is also assumed as out of scope for Landsat or Sentinel-2 scenes, and even PlanetScope (Planet Labs PBC, 2018) imagery. Stark radiometric changes can occur between two captures within a day, e.g., weather or illumination

changes. More drastic scene differences like green grass getting covered in snow are more seldom but may occur. As the current network is a proof of concept, the amount of these hard to handle cases is further reduced by using only imagery that is captured during the summer terms.

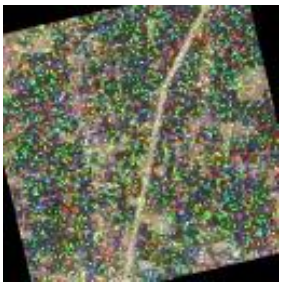


Figure 4. Example for a noisy, rotated image with mixed pixels.

For this first iteration of the super resolution network, only RGB spectral channels are used. For later versions, refitting the architecture with a multispectral approach for common wavelengths between Sentinel-2 and PlanetScope sensors is intended. Data augmentation comprised different degrees of rotations and Gaussian noise as depicted in Figure 4 as well as mixed pixels. The low-resolution imagery set is comprised of Landsat data at a 15 m PAN-sharpened GSD which was then supplemented with Sentinel-2 scenes with a 10 m GSD. The input shape of the low-resolution set is 128 x 128 pixels. The high-resolution imagery set consists of Sentinel-2 scenes with a 10 m GSD. To further improve the network’s ability to recreate finer details, PlanetScope scenes and SPOT 6 imagery of highly detailed urban regions are added to the high-resolution data pool. An overview of the data sources which are used for training are shown in Table 1. Figure 5 shows examples for every satellite that is used for training the network.

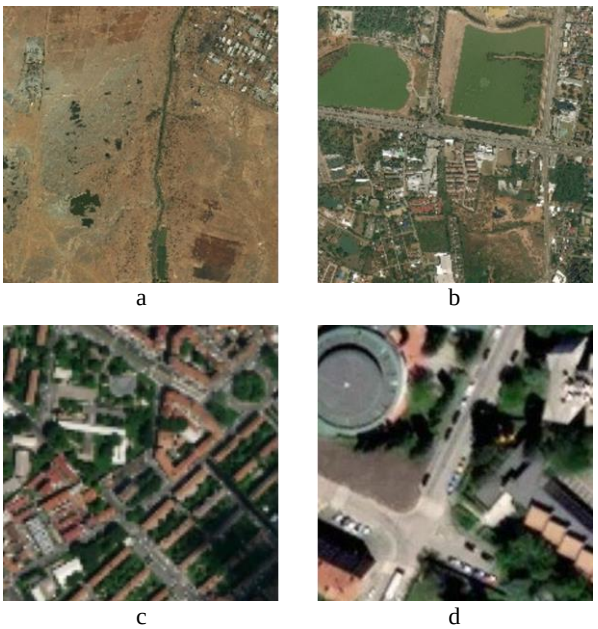


Figure 5. Examples for satellite image scenes. A Landsat scene with a PAN-sharpened GSD of 15 m is shown in a. A Sentinel-2 scene with a GSD of 10 m is shown in b. A PlanetScope scene with a GSD of 4 m is shown in c. A SPOT 6 scene with a PAN-sharpened GSD of 1.5 m is shown in d. (© Planet Labs PBC 2024 & © Airbus DS 2023).

The images are patched and rescaled for VSISR. The input shape of the high-resolution patches is 224 x 224 pixels. This limitation is induced by the VGG architecture in use. The resulting output shape for the super resolution image is 256 x 256 pixels, resulting in a quadrupled pixel count compared to the original low-resolution version, equal to halving the GSD. In total, 4510 images are used to train VSISR with a train/test split of 67/33. A validation set of 80 is added for evaluation.

Satellite Mission	GSD [m]
Landsat	15
Sentinel-2	10
PlanetScope	4
SPOT 6	1.5

Table 1. Satellite missions in the dataset with respective GSDs.

4.2 Metrics

Apart from typical machine learning optimization metrics like loss, the results are inspected visually, too. This is supported by creating normalized difference images, as displayed in Figure 6, helping to quickly identify problematic regions. These images are calculated from the original high-resolution sample and the generated super resolution image. A black difference image shows perfect similarity, a white difference image shows no similarity. To compare the results to other neural networks for super resolution on satellite imagery, Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) are calculated. The PSNR is calculated from the maximum signal and the mean squared error on a logarithmic scale. The higher the result, the higher the similarity. SSIM is applied as a measure of image similarity with results between -1 and 1. For a perfectly matching image, SSIM equals 1 while -1 reflects the perfect negative of the image. A value of 0 is a sign of no similarity.

Figure 6. Exemplary normalized radiometric difference image. Brighter spots correspond to higher radiometric differences between super resolution and the original high-resolution scene.

5. Results

First, an overview of the results of different iterations of VSISR is given in this section. This, inter alia, includes training strategies and their effects, and machine learning metrics. In 5.2, examples are given for the different combinations of satellites in training.

5.1 Overview

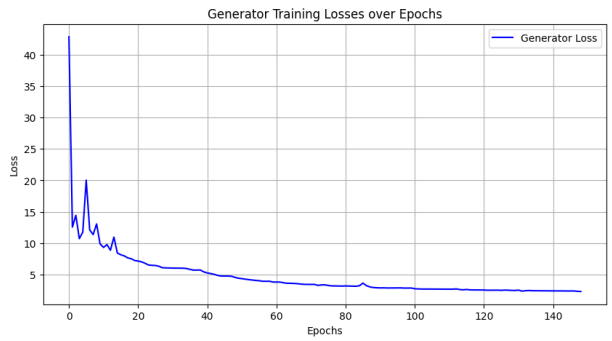
In the beginning, hallucinations in random places within the images occur which appear due to overfitting. This is resolved by introducing L2 regularization. Adjusting the loss function also improves radiometric results. The biggest change in radiometric similarity is achieved with the introduction of

mixed pixels. A reference image without mixed pixels is depicted in Figure 7. Utilizing mixed pixels results in a greater radiometric similarity between super resolution and ground truth and reduces artifacts in the generated image.

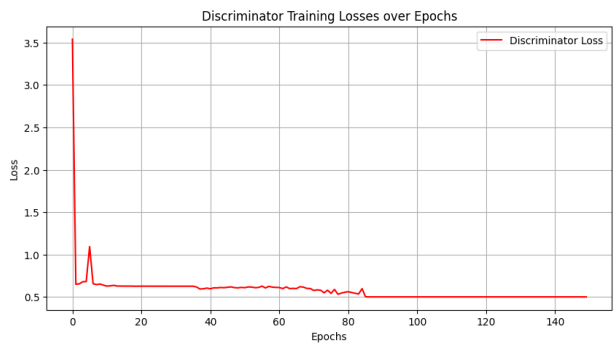
a b

Figure 7. Without mixed pixels, more artifacts occur, and overall radiometry of the super resolution image a is dissimilar to the high-resolution ground truth b.

To optimize the results geometrically, the SPOT 6 imagery is introduced to the training set. With this, the results again greatly improve in quality. Its main effect are the reduction of structural hallucinations and further improvements to radiometric versatility and accuracy. Discriminator loss remains constant after epoch 85 as shown in Figure 8. As generator loss still decreases, training is continued. After more than 150 epochs, no significant gains in accuracy can be achieved, therefore training is stopped with epoch 150 to prevent overfitting. The training passes take 7-8 hours on an Nvidia V100 with similar times on an A100 for a batch size of one as typical for GANs.

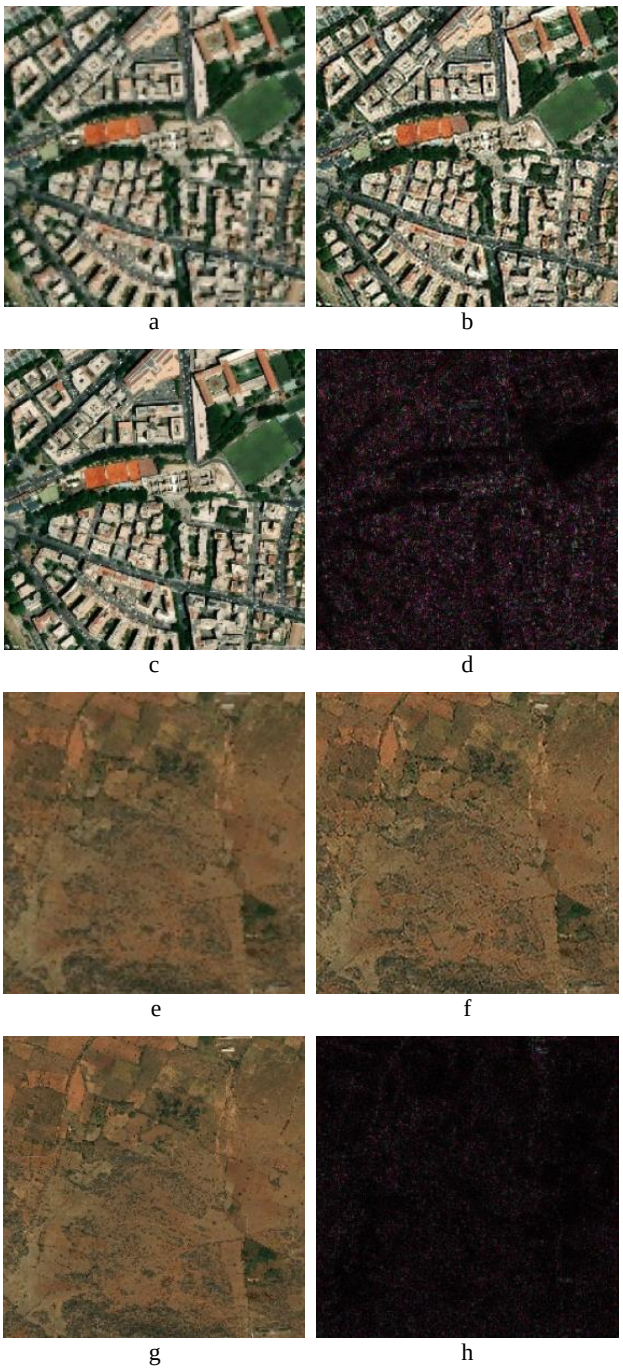


a



b

Figure 8. The loss curve for the generator is shown in a. As generator loss still decreases after epoch 85, training continues. Discriminator loss remains constant after epoch 85 as shown in b.



a b c d e f g h

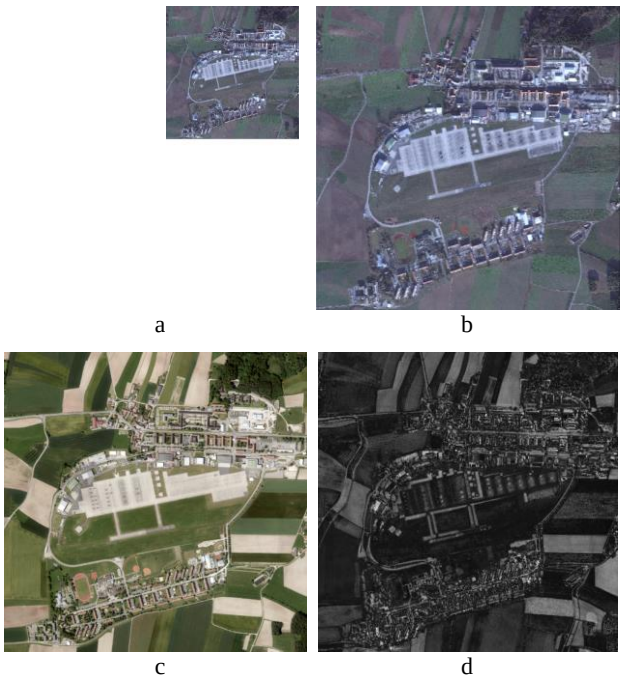
Figure 9. Results for scaling a Landsat scene a with a Spot 6 scene c. The super resolution is displayed in b with its respective difference image d (PSNR: 24.69, SSIM: 0.7875). Results for scaling a Landsat scene e with a Sentinel-2 scene g. The super resolution is displayed in f with its respective difference image h (PSNR: 24.54, SSIM: 0.6857). The contrast between image h and the comparably low SSIM is notable. (© Airbus DS 2023).

While optimizing the network configuration, a mostly white, textured noise pattern occurs in several super resolution images. Comparisons between these occurrences show that they are induced in areas where larger white patches (e.g. metal rooftops on warehouses) are present in the low-resolution images. Additionally, it occurs in regions of high-frequency white areas where features are hard to distinguish within the low-resolution image (e.g. parallel striped markings). A distinct source for this

effect is not identified. These noise pattern problems are solved by introducing the high-resolution SPOT imagery in combination with the introduction of a gradient penalty.

5.2 Experiments

Table 2 gives an overview of the PSNR and SSIM values of each satellite combination. Figures 9 and 11 display examples of super resolution products along the low- and high-resolution data and their respective difference images. While testing on the validation set, i.e. data the network does not see during training, a mean PSNR of 25.30 dB and a mean SSIM of 0.8098 is achieved. Figure 10 illustrates VSISR’s 4x upscaling capability, general adaptiveness and radiometric consistency of the resulting super resolution image. In this case, a significant time delay between low resolution and reference image is chosen. Further, Worldview-3 imagery was not part of the training, test or validation dataset. Additional results are displayed in the appendix.

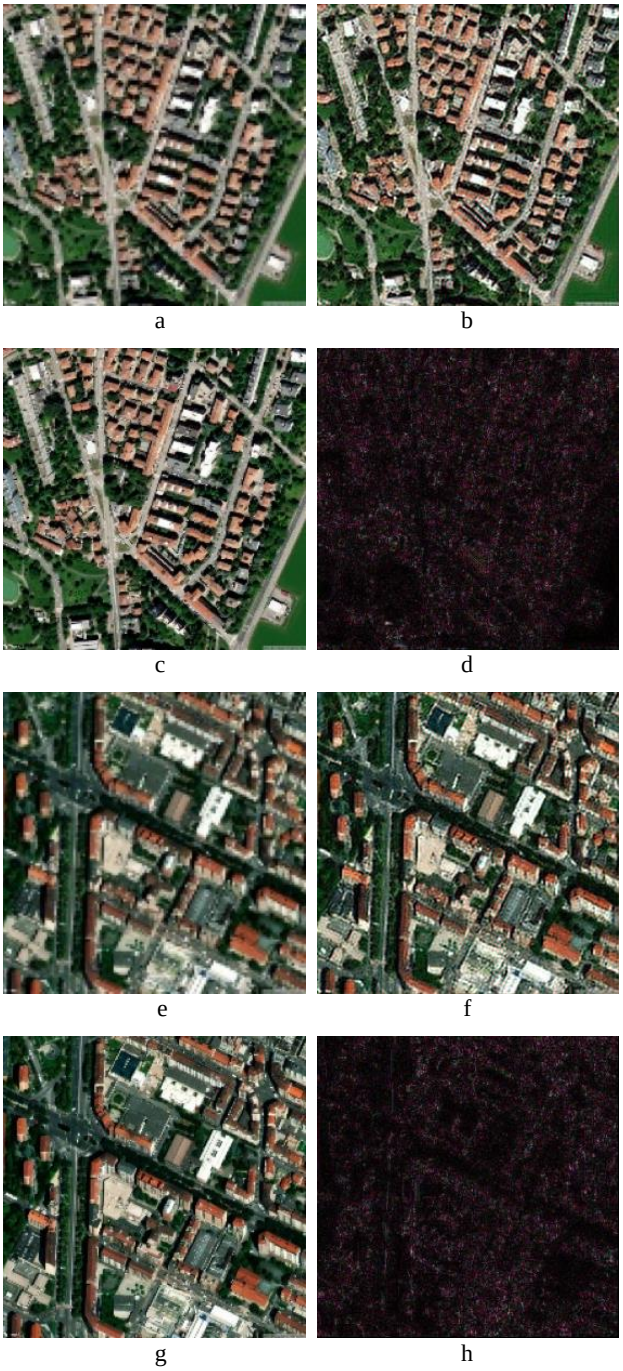


a b c d

Figure 10. Results for scaling a Worldview-3 scene a with a reference PlanetScope scene c. In this case, the time offset between input and reference image is chosen deliberately high. The super resolution is displayed in b. It is to note that VSISR managed to maintain the radiometric characteristics even with an intentionally contrasting reference image. This results in a more distinct difference map d compared to other cases (PSNR: 27.19, SSIM: 0.8123). (© TPMO, 2018)

6. Discussion

Through iterating on network parameter selection and training regime, radiometric discrepancies are minimized. Still, as expected, even with optimization techniques focusing on high-frequency details, there remain issues in especially complex image regions. Common patterns like houses ore streets are reconstructed well, while very irregular patterns, for example unique or uncommon lane markings still pose problems for the network. The step up in geometric accuracy by introducing very high-resolution imagery brings the desired improvements. Still, residual differences between super resolution and high-resolution ground truth remain in complex scenes. Due to the wider distribution of features and their resulting lower frequency in rural areas, these are handled better by the network which is reflected in the difference map of Figure 9 h. It is to note that the low SSIM of the Landsat and Sentinel-2 sample of 0.6857 contrasts with the low values present in the difference image. Visually, the result is on par with other combinations.



a b c d e f g h

Figure 11. Results for scaling a Sentinel-2 scene a with a PlanetScope scene c. The super resolution is displayed in b with its respective difference image d (PSNR: 25.10, SSIM: 0.8146). Results for scaling a Sentinel-2 scene e with a Spot 6 scene g. The super resolution is displayed in f with its respective difference image h (PSNR: 25.12, SSIM: 0.7823). (© Planet Labs PBC 2024 & © Airbus DS 2023)

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-3-2024-155-2024 | © Author(s) 2024. CC BY 4.0 License.

Low Resolution	High Resolution	PSNR [dB]	SSIM
Landsat	SPOT	24.69	0.7875
Landsat	Sentinel-2	24.54	0.6857
Sentinel	Planet	25.10	0.8146
Sentinel	SPOT	25.12	0.7823
Worldview-3	Planet	25.12	0.7823
Mean on validation		25.30	0.8098

Table 2. PSNR and SSIM results for the satellite combinations shown in the examples. Additionally, the overall mean values on the validation set are shown. Matches are a coincidence.

Further tests are needed in this case to examine potential reasons for this effect and identify the component lowering SSIM.
Augmenting the data by applying different scaling factors is not considered for this dataset as the different ground truth sizes of the imagery already introduce a scale variation. A comparison of the effects of data scaling can be introduced to a second iteration of VSISR. The mean values for PSNR of 25.30 and SSIM of 0.8098 are favorable but do not outperform state of the art super resolution networks as displayed in Table 3. It is to note that these results might be misleading as the datasets are different. A direct comparison would only be possible when working on the same data. In general, VSISR generalizes well due to its diverse dataset. This is set out by utilizing data that is completely unknown to the network like in Figure 10.

Network	PSNR [dB]		SSIM	
	min.	max.		
SRGAN	21.48	29.18	0.60	0.83
ESRGAN	22.76	28.74	0.60	0.83
MCWESRGAN	25.05	31.10	0.70	0.92
SRCNN	22.87	25.4	0.61	0.82
FSRCNN	24.58	31.27	0.67	0.90
VSISR	25.30		0.81	

Table 3. Comparison of PSNR and SSIM results on two datasets of established super resolution methods by (Karwowska and Wierzbicki, 2023), supplemented with VSISR results.

7. Conclusion

VSISR already shows promising results in this proof-of-concept stage. Adapting the original concept of (Ledig et al., 2017) creates a versatile and capable GAN for super resolution of remote sensing scenes. Focusing on high-frequency details and keeping radiometric consistency are the prime ideas behind the adjustments that are made to the network and its dataset. The results are favorable, although further research is already laid out concerning dataset and network structure. Additionally broadening the dataset can improve on the network’s ability to generalize on even more diverse datasets and also the handling of high-frequency regions in the imagery. Compared to other neural networks for satellite image super resolution in their respective datasets, VSISR performs in the middle regarding PSNR and SSIM, but it is the first time that, using mixed pixel training, this versatility concerning the satellite data source is achieved with a comparably small dataset while maintaining their unique radiometric traits. Introducing very high-resolution satellite imagery produces the expected outcome to further help in learning to distinguish small features. New prospects are forthcoming as more very high-resolution systems become publicly available. Utilizing GSDs below one meter as provided by PlanetLab’s SkySat platform (Kiran et al., 2014) or even in the regime of 10 cm as proposed by (Tri, 2022) for commercial access in 2025 will further enhance super resolution capabilities. An important next step concerning the evaluation of VSISR’s performance is creating a dataset for comparing the building center detection network’s capability to reliably segment buildings with and without a prior treatment by VSISR. Additionally, VSISR will be adapted to process multispectral data.

Acknowledgement

The CuBy project is funded by the Bavarian Financial Ministry (Bayerisches Staatsministerium der Finanzen und für Heimat). The project partners are the TUM professorship of Photogrammetry & Remote Sensing. the Bavarian Surveying Administration (Landesamt für Digitalisierung, Breitband und Vermessung), Zentrum für Telematik Würzburg and the TUM chair of astronomical and physical geodesy.

References

Bauer, R., Pail, R., Schilling, K., Stilla, U., Schleder, D., Dotterweich, M., Roschlaub, R., Kleinschrodt, A., Draschka, L., Gruber, T., Zingerle, P., Greza, M., 2023. Bayerisches Satellitennetzwerk für Fernerkundung und Biomonitoring. *ZfV–Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 148, 4/2023, 219-229. doi.org/10.12902/zfv-0436-2023.

Breaking Defense, 2023. Earth’s a big world with secrets to reveal to those who can sense and make sense of it, breakingdefense.com/2023/09/earths-a-big-world-with-secrets-to-reveal-to-those-who-can-sense-and-make-sense-of-it (22 April 2024).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, doi.org/10.48550/arXiv.1406.2661.

Greza, M., Hoegner, L., Hirt, P., Roschlaub, R., Stilla, U., 2023. Satellite Network Bavaria – Mission and Data Processing. *Publikationen der Deutschen Gesellschaft für Photogrammetrie, Fernerkundung und Geoinformation (DGPF)*, 43. doi.org/174-182. 10.24407/KXP:184104914X.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition* 770-778. doi.org/10.48550/arXiv.1512.03385.

Karwowska, K., Wierzbicki, D., 2023. MCWESRGAN: Improving Enhanced Super-Resolution Generative Adversarial Network for Satellite Images, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, 9459-9479. doi.org/10.1109/JSTARS.2023.3322642.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W., 2017. Photo-realistic single image super-resolution using a generative adversarial network, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681-4690. doi.org/10.48550/arXiv.1609.04802.

Meygret, A., Baillarin, S., Gascon, F., Hillairet, E., Dechoz, C., Lacherade, S., Martimort, P., Spoto, F., Henry, P., Duca, R, 2009. SENTINEL-2 image quality and level 1 processing, *Proceedings of SPIE - The International Society for Optical Engineering*. doi.org/10.1117/12.826184.

Murthy, K., Shearn, M., Smiley, B., Chau, A., Levine, J., Robinson, M., 2014. SkySat-1: very high-resolution imagery from a small satellite, *Proc. SPIE 9241, Sensors, Systems, and Next-Generation Satellites* XVIII. doi.org/10.1117/12.2074163.

Planet Labs PBC, 2018. Planet Application Program Interface: In Space for Life on Earth. api.planet.com (11 March 2024).

Roschlaub, R., Möst, K., Krey, T., 2020. Automated classification of building roofs for the updating of 3D building models using heuristic methods. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88, 85-97. doi.org/10.1007/s41064-020-00099-9.

Roschlaub, R., Glock, C., Zerndl, M., Maier, A., Möst, K., Hirt, P., Greza, M., 2023. Grundlagen zur automatisierten Ermittlung hochgenauer Passpunkte für CubeSat-Satellitenbilder mittels Deep-Learning-gestützter Delaunay-Triangulation basierend auf Gebäudedaten des Liegenschaftskatasters. *ZfV–Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 148, 2/2023, 66-93. doi.org/10.12902/zfv-0424-2023.

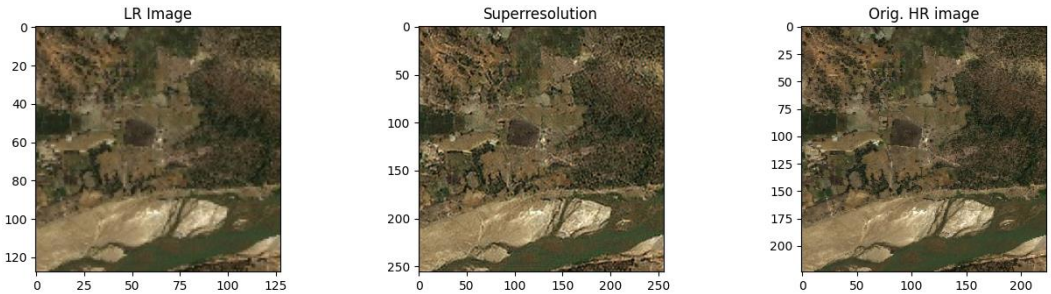
Roschlaub, R., Glock, C., Zerndl, M., Maier, A., Möst, K., Krey, T., 2024. Mosaikieren von Fernerkundungsdaten für im Verbund fliegende CubeSats. *ZfV-Zeitschrift für Geodäsie, Geoinformation und Landmanagement*, 149, 1/2024, 23-44. doi.org/10.12902/zfv-0456-2023.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. *Computer Vision and Pattern Recognition (cs.CV).* doi.org/10.48550/arXiv.1409.1556.

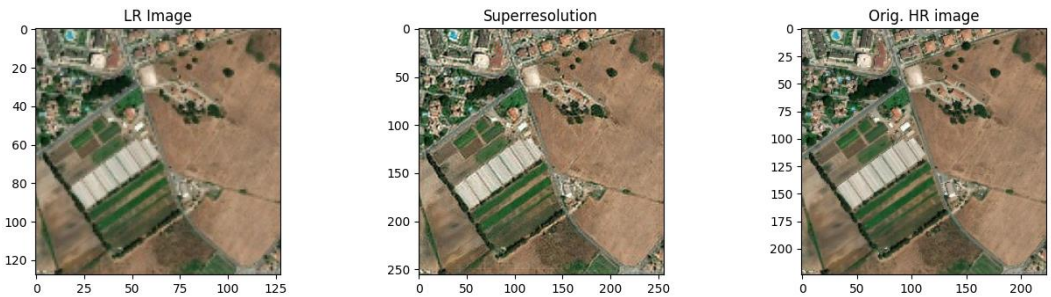
Spiegel, M., 2007. Improvement of interior and exterior orientation of the three line camera HRSC with a simultaneous adjustment. *International Archives of Photogrammetry and Remote Sensing*, *36*(3/W49B), 161-166.

Tri, W., 2022. A First Look at 10cm Satellite Imagery, albedo.com/post/albedo-simulated-imagery (22 April 2024).

Appendix



Exemplary result for Landsat imagery scaled to Sentinel-2.



Exemplary result for Sentinel-2 imagery scaled to PlanetScope. (© 2024 Planet Labs PBC)

Exemplary result for Worldview-3 imagery scaled to PlanetScope. (© TPMO, 2018)