

Letter

Generation of Super-Resolution Benchmark Dataset for Compact Advanced Satellite 500 Imagery and Proof of Concept Results

Yonghyun Kim ^{1)*} · Jisang Park²⁾ · Daesub Yoon²⁾

Abstract: In the last decade, artificial intelligence’s dramatic advancement with the development of various deep learning techniques has significantly contributed to remote sensing fields and satellite image applications. Among many prominent areas, super-resolution research has seen substantial growth with the release of several benchmark datasets and the rise of generative adversarial network-based studies. However, most previously published remote sensing benchmark datasets represent spatial resolution within approximately 10 meters, imposing limitations when directly applying for super-resolution of small objects with *cm* unit spatial resolution. Furthermore, if the dataset lacks a global spatial distribution and is specialized in particular land covers, the consequent lack of feature diversity can directly impact the quantitative performance and prevent the formation of robust foundation models. To overcome these issues, this paper proposes a method to generate benchmark datasets by simulating the modulation transfer functions of the sensor. The proposed approach leverages the simulation method with a solid theoretical foundation, notably recognized in image fusion. Additionally, the generated benchmark dataset is applied to state-of-the-art super-resolution base models for quantitative and visual analysis and discusses the shortcomings of the existing datasets. Through these efforts, we anticipate that the proposed benchmark dataset will facilitate various super-resolution research shortly in Korea.

Keywords: Super-resolution, Benchmark dataset, CAS500, Deep learning, Foundation model, Spatial dependence, Spatial heterogeneity

1. Introduction

In recent years, breakthrough research has been carried out in remote sensing fields with remarkable advancement in the non-linear representation of neural

networks and the spatial feature learning capabilities of the convolution neural network (Ma et al., 2019; Zhu et al., 2017). More recently, each topic, such as semantic segmentation, change detection, and image captioning, has been integrated with large language

Received August 1, 2023; Revised August 14, 2023; Accepted August 22, 2023; Published online August 31, 2023

¹⁾ Senior Researcher, Mobility UX Research Section, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

²⁾ Principal Researcher, Mobility UX Research Section, Electronics and Telecommunications Research Institute, Daejeon, Republic of Korea

* Corresponding author: Yonghyun Kim (yhkeen@etri.re.kr)

Copyright © 2023 by The Korean Society of Remote Sensing. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

models for downstream tasks like question answering (Wen et al., 2023). However, it is worth noting that satellite image super-resolution has not been as actively pursued compared to its computer vision field and other tasks in the remote sensing domain. The primary reason for this is the lack of suitable datasets for analysis, particularly datasets with submeter or *cm* unit spatial resolution that are globally distributed, which are limited in availability. Another reason could be the practical challenge of obtaining a dataset that spans the entire range of imagery pyramids, from drone-to-aerial-to-satellite, considering spatial resolutions ranging from centimeters to tens of meters or even higher degradation.

Image super-resolution categories could be divided into single- and multi-frame image super-resolution (Chen et al., 2022). These categories are further distinguished by their approaches, which can be classified as supervised and unsupervised (Wang et al., 2022a). The distinct difference lies in how they utilize data during training. Supervised methods train the model using pairs of high-resolution and corresponding low-resolution images. On the other hand, unsupervised methods attempt to restore high-resolution images using only their inherent features without needing paired high-resolution data. In other words, both supervised and unsupervised approaches require reference data for quantitative and visual evaluations, and the importance of high-quality data becomes even more pronounced when developing large-scale foundation models such as the segment anything model (Kirillov et al., 2023). While one could envision a scenario where drone or aerial images are acquired during the satellite passing time, in this case, the significant disparity in image geometries would render them unsuitable for super-resolution research. In this letter, we propose a refined pipeline to address these challenges by leveraging images of extremely high resolution. If such images are available, our method, based on modulation transfer function (MTF)

simulation, allows us to create a benchmark dataset by degrading spatial resolution; thus, this method can generate datasets with spatial resolution ranging from a few centimeters to several tens of meters while maintaining the identical image geometry as the ultra-high resolution images. This significant advantage makes the benchmark dataset suitable for super-resolution research. In addition, by applying state-of-the-art base algorithms and conducting both visual and quantitative analyses, we introduce the value of the proposed data generation pipeline and discuss its implications.

2. Super-Resolution Benchmark Dataset

2.1. Open Dataset

As seen in various deep learning research, the significance and necessity of well-crafted and sophisticated datasets have increased, just as the development of more advanced and deeper networks has explainability (Hasanpour Zaryabi et al., 2022). Several publicly available datasets exist in the computer vision fields of super-resolution research to meet these needs. However, in the remote sensing field, the availability of publicly accessible datasets is limited, and most datasets exhibit spatial resolutions within approximately 10 meters (Wang et al., 2022a). As a result, there are significant constraints for conducting super-resolution research at a submeter or *cm* unit spatial resolution. In the field of remote sensing, recent trends in super-resolution research and datasets can be referred to the well-constructed study by Wang et al. (2022b), and as for the latest developments in real-world single image super-resolution and datasets, Chen et al. (2022) serve as a valuable reference.

2.2. Benchmark Dataset for CAS500-1

The primary targeting sensor of this study is the Compact Advanced Satellite (CAS)500-1 satellite of South Korea; however, the proposed generation

Table 1. Specification of CAS500-1

Specification	
Spatial resolution	Panchromatic: 0.5 m Multispectral: 2.0 m
Spectral range	Panchromatic: 450–900 nm Multispectral: 450–900 nm
Radiometric resolution	12 bits
Swath	≥ 12 km
Orbit	Circular sun-synchronous ascending orbit
Launch date	March 22, 2021

process of the super-resolution image pyramidal dataset can be equally applied to other satellites, such as the Korean multi-purpose satellite constellation. The CAS500-1 was launched in March 2021 with a focus on efficient land management, cartography, disaster response, and 3D digital twin applications, and the main specifications of CAS500-1 are summarized in Table 1 (Yoon et al., 2020). The CAS500-1 mission, which has approximated 528 km orbital altitude, was initiated as part of standardizing the production of medium-sized satellites weighing about 500 kg, and the sensor comprises the panchromatic band with a 0.5-meter spatial resolution and the multispectral bands with a 2.0-meter spatial resolution. The multispectral bands, as shown in Table 1, consist of four channels RGB, near-infrared (NIR), and the life cycle of the mission is four years.

3. A New Benchmark Dataset Using MTF Simulation

3.1. Pyramidal Image Generation Using Airborne Imagery

The image pyramid or pyramid representation is a structure of multiscale signal representation applied in signal processing and computer vision (Adelson et al., 1984). The primary assumption of this image pyramid is that the signal or image is subject to repeated

smoothing and subsequent sampling. In generating an image pyramid, low-pass filters are commonly used, and the MTF-based Gaussian low-pass filters have been widely adopted in image fusion fields due to their ability to mimic the physical characteristics of satellite sensors effectively (Kallel, 2014; Kim et al., 2017; Massip et al., 2012). In other words, targeting high- and low-resolution image pairs or super-resolution pyramidal benchmark datasets can be constructed by applying low-pass filtering that effectively reflects the physical characteristics. As a result, this image pyramid serves as valuable reference data in the research of super-resolution imaging. In simple terms, we utilized 12 cm resolution aerial imagery provided by the National Geographic Information Institute (NGII), and these datasets mean that we can make training data that can enhance or super-resolve the 50 cm spatial resolution of CAS500-1 imagery by nearly four times to approximately 10 cm.

3.2. Generation of Benchmark Dataset Using MTF Simulation

Typical super-resolution processes and optical degradation models are considered inverse relationships (Chen et al., 2022; Wang et al., 2022b); thus, low-resolution image I^{LR} is normally assumed to be a degraded version from a corresponding high-resolution image I^{HR} , which can be represented as:

$$I^{LR} = [I^{HR} \otimes M] \downarrow_r + n \quad (1)$$

where \otimes denotes the optical degradation models defined by the MTF kernels M and \downarrow_r represents the downsampling operator. Mathematically, n means the additive noise typically assumed to be Gaussian noise. As shown in Eq. (1), in generating a high-quality super-resolution benchmark dataset, the most crucial aspect lies in effectively emulating the low-pass filters using the MTF kernels M . This paper simulates the low-pass filter using the MTF cutoff frequency, resulting in a Gaussian bell shape.

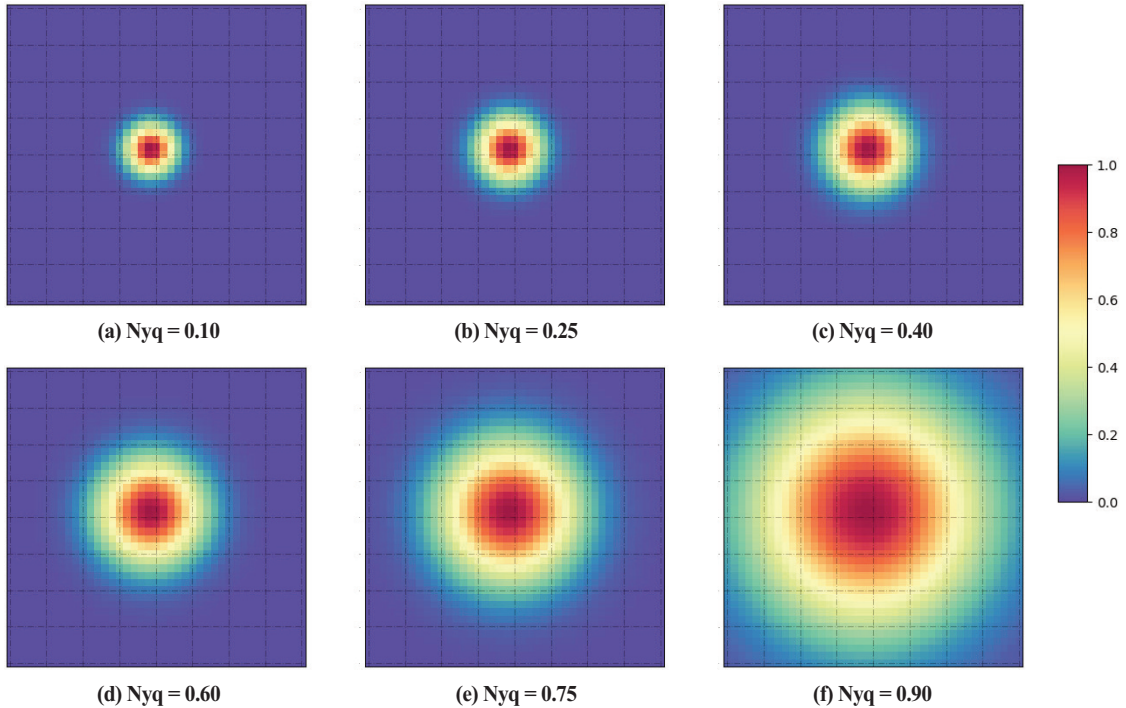


Fig. 1. Simulated 2D Gaussian kernels by mimicking the MTFs of the satellite sensor. The spatial resolution ratio is set to 4, and the Nyquist frequency ranges (a) 0.10, (b) 0.25, (c) 0.40, (d) 0.60, (e) 0.75, and (f) 0.90.

The MTF is the module of the Fourier transform of the point spread function, and the satellite manufacturer usually releases information about the MTF of the sensor (Ghassemian, 2016; Kallel, 2014). The multispectral band cutoff values generally range from 0.25 to 0.30 in actual satellite sensors. In this study, thus, the cutoff value of the CAS500-1 was set to 0.26 and assumed to be a bell-shaped Gaussian low-pass filter. The 2D Gaussian low-pass filter is shown in Fig. 1 with cutoff values at Nyquist (Nyq) frequency range from 0.1 to 0.9. As illustrated in Fig. 1, as the cutoff frequency decreases, a significant amount of high-frequency components or edge information in the imagery is attenuated or removed by the MTF filters. The primary research to utilize the MTF of satellite sensors can be found in Aiazzi et al. (2006) and Vivone et al. (2014).

4. Experiments

The airborne imagery utilized in this experiment was provided by the NGII, and the study area corresponds to map sheet number Seoul078-37608078 at a 1:5,000 digital map scale. The complete size of the imagery is denoted as height \times width, measuring 23,975 \times 19,250 pixels. For experimentation purposes, a subset of the image in the main entrance of Yonsei University was extracted at a size of 512 \times 512 pixels, as shown in Fig. 2. The CAS500-1 has four channels; however, we processed data by targeting only three RGB channels. The reason for this approach is that the input aerial imagery provided by the NGII consisted of only three RGB channels. It should be noted that even if the initial aerial imagery had four bands, including the NIR band, or if it was hyperspectral imagery, the processing steps would remain the same.

4.1. Results and Analysis

The primary experimental goal of this study is to apply our benchmark dataset to well-established algorithms, often utilized as base models in natural image super-resolution research, and assess their suitability for satellite imagery super-resolution. In essence, our intention is not to imply the ineffectiveness of the deep learning networks within these base models. Instead, the focus

is on recognizing that the dataset used to train these base models lacks appropriate spatial features for satellite image super-resolution. For this reason, we utilized the super-resolution base methods, namely the Blind Super-Resolution Generative Adversarial Networks (BSRGAN), Swin transformer Image Restoration (SwinIR)-Large (Liang et al., 2021; Zhang et al., 2021). The former is a GAN-based super-resolution method, while the latter



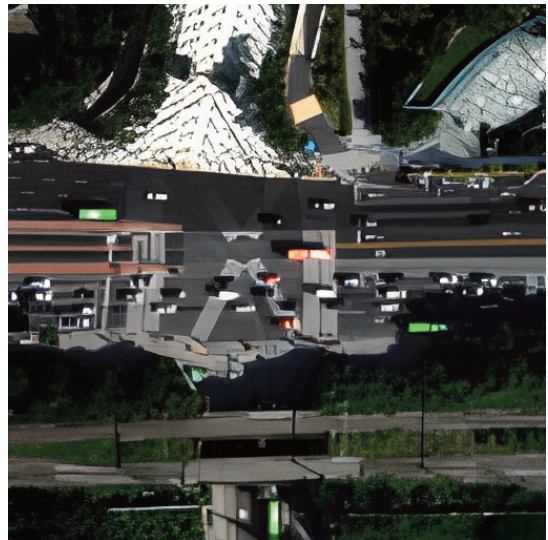
(a) Input airborne imagery (12 cm)



(b) MTF-simulated imagery (48 cm)



(c) BSRGAN (12 cm)



(d) Swin-Large (12 cm)

Fig. 2. The super-resolution results for proof of concept: (a) the input airborne imagery, (b) the MTF-simulated imagery, (c) the output of the BSRGAN model, and (d) the output of the Swin-Large model.

Table 2. Statistics of the super-resolved imagery

Method \ Statistic	PSNR	UIQI	RMSE
BSRGAN	16.84	0.38	36.70
SwinIR-Large	16.04	0.35	40.21

employs a vision transformer-based approach with shifted windows.

The experimental results are depicted in Fig. 2. The 12-cm spatial resolution image is shown in Fig. 2(a), and the MTF-based simulated 48-cm spatial resolution image, which is assumed identical to CAS500-1 data, is shown in Fig. 2(b). In addition, Fig. 2(a) indicates details like pedestrian crossings, speed limits, railway tracks, and other intricate landscapes. In contrast, the simulated low-resolution image in Fig. 2(b) lacks discernible traffic markings on the roads, and even the pedestrian crossings, indicative of image modulation, appear blurred. Fig. 2(c) presents the outcome of the BSRGAN model, where the overall results resemble an image segmentation output more than actual high-resolution imagery and lack sharpness. Similarly, in Fig. 2(d), the results of the Swin-Large model also exhibit a lack of image sharpness, akin to an image segmentation output, except for some apparent landscape features on the ground.

Furthermore, neither produced satisfactory results when quantitatively comparing the two super-resolution methods to the reference data (Table 2). These outcomes could potentially stem from using pre-trained models without fine-tuning in our investigation. However, the results signify that the algorithms serving as base models for super-resolution research have not learned suitable spatial features for deep learning analysis of remote sensing data, specifically aerial and satellite imagery. The quantitative evaluation of Table 2 reaffirms this by providing numerical evidence; the peak signal-to-noise ratio (PSNR) values were around 16 dB, and the root-mean-square error (RMSE) approximated 40 pixels. In 8-bit image super-resolution research,

achieving a PSNR value of around 30 dB is typically considered common. In other words, a PSNR value of approximately 16 dB and a universal image quality index (UIQI) value of about 0.35 are notably low, which aligns with the earlier-mentioned visual assessment of inadequate results. This consistency between the visual evaluation and quantitative outcomes underscores the need for further research utilizing the well-crafted benchmark dataset.

5. Conclusions

An ideal scenario would be to utilize a globally distributed benchmark dataset to train super-resolution deep learning models that capture worldwide land covers and urban forms. However, constructing a dataset encompassing different spatial resolutions considering various spatial features is a highly challenging endeavor. To address these limitations, this study employed MTF simulation of the CAS500-1 satellite to create a benchmark dataset. Subsequently, super-resolution base models were applied and compared to the reference data for visual evaluation in Fig. 2 and quantitative assessment in Table 2. However, the restored imagery exhibited a loss of sharpness and even showed segmentation-like outcomes, underscoring the need for further research.

Especially considering the spatial heterogeneity and dependence, which signify the varied configurations of urban environments, conducting simulation studies using real-world datasets rather than synthetic ones would substantially enhance the values, as demonstrated in this study. In conclusion, this study highlights the significance of the benchmark dataset used for training rather than solely pointing out limitations within the deep learning networks of the base model. Again, when considering spatial heterogeneity, using the MTF simulation employed in this study could create a benchmark dataset optimized for land covers in Korea.

In essence, what takes precedence is the meticulously crafted benchmark dataset, and the high-performance algorithm comes next, wherein performance comparison is based on these benchmark datasets. Through this approach, we anticipate the development of foundation models shortly for super-resolution imaging in the remote sensing field.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant R&D program of Digital Land Information Technology Development funded by the Ministry of Land, Infrastructure and Transport (MOLIT) (Grant RS-2022-00142501).

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

- Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., and Ogden, J. M., 1984. Pyramid methods in image processing. *RCA Engineer*, 29(6), 33–41.
- Aiazzi, B., Alparone, L., Baronti, S., Garzelli, A., and Selva, M., 2006. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogrammetric Engineering & Remote Sensing*, 72(5), 591–596. <https://doi.org/10.14358/PERS.72.5.591>
- Chen, H., He, X., Qing, L., Wu, Y., Ren, C., and Sheriff, R., 2022. Real-world single image super-resolution: A brief review. *Information Fusion*, 79, 124–145. <https://doi.org/10.1016/j.inffus.2021.09.005>
- Ghassemian, H., 2016. A review of remote sensing image fusion methods. *Information Fusion*, 32, 75–89. <https://doi.org/10.1016/j.inffus.2016.03.003>
- Hasanpour Zaryabi, E., Moradi, L., Kalantar, B., Ueda, N., and Halin, A. A., 2022. Unboxing the black box of attention mechanisms in remote sensing big data using XAI. *Remote Sensing*, 14(24), 6254. <https://doi.org/10.3390/rs14246254>
- Kallel, A., 2014. MTF-adjusted pansharpening approach based on coupled multiresolution decompositions. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6), 3124–3145. <https://doi.org/10.1109/TGRS.2014.2369056>
- Kim, Y., Kim, M., Choi, J., and Kim, Y., 2017. Image fusion of spectrally nonoverlapping imagery using SPCA and MTF-based filters. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2295–2299. <https://doi.org/10.1109/LGRS.2017.2762427>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L. et al., 2023. Segment anything. *arXiv preprint arXiv:2304.02643*. <https://doi.org/10.48550/arXiv.2304.02643>
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R., 2021. SwinIR: Image restoration using swin transformer. *arXiv preprint arXiv:2108.10257*. <https://doi.org/10.48550/arXiv.2108.10257>
- Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166–177. <https://doi.org/10.1016/j.isprsjprs.2019.04.015>
- Massip, P., Blanc, P., and Wald, L., 2012. A method to better account for modulation transfer functions in ARSIS-based pansharpening methods. *IEEE Transactions on Geoscience and Remote Sensing*, 50(3), 800–808. <https://doi.org/10.1109/TGRS.2012.0000000>

- 2011.2162244
- Vivone, G., Alparone, L., Chanussot, J., Dalla Mura, M., Garzelli, A., Licciardi, G. A. et al., 2014. A critical comparison among pansharpening algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5), 2565–2586. <https://doi.org/10.1109/TGRS.2014.2361734>
- Wang, P., Bayram, B., and Sertel, E., 2022a. 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., Yi, J., Guo, J., Song, Y., Lyn, J., Xu, J. et al., 2022b. A review of image super-resolution approaches based on deep learning and applications in remote sensing. *Remote Sensing*, 14(21), 5423. <https://doi.org/10.3390/rs14215423>
- Wen, C., Hu, Y., Li, X., Yuan, Z., and Zhu, X. X., 2023. Vision-language models in remote sensing: Current progress and future trends. *arXiv preprint arXiv:2305.05726*. <https://doi.org/10.48550/arXiv.2305.05726>
- Yoon, S. J., Son, J. H., Park, H., Seo, J., Lee, Y., Ban, S. et al., 2020. CAS 500-1/2 Image utilization technology and system development: achievement and contribution. *Korean Journal of Remote Sensing*, 36(5–2), 867–879. <https://doi.org/10.7780/kjrs.2020.36.5.2.2>
- Zhang, K., Liang, J., Van Gool, L., and Timofte, R., 2021. Designing a practical degradation model for deep blind image super-resolution. *arXiv preprint arXiv:2103.14006*. <https://doi.org/10.48550/arXiv.2103.14006>
- Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., and Fraundorfer, F., 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36. <https://doi.org/10.1109/MGRS.2017.2762307>