A COMPARISON OF DEEP LEARNING-BASED SUPER-RESOLUTION FRAMEWORKS FOR SENTINEL-2 IMAGERY IN URBAN AREAS

S. T. Seydi [1], H. Arefi [1, *]

[1] i3mainz, Institute for Spatial Information and Surveying Technology, School of Technology, Mainz University of Applied Sciences, D-55118 Mainz, Germany - (seyd.seydi, hossein.arefi)@hs-mainz.de

KEY WORDS: Sentinel-2, Deep Learning, Super-Resolution, SRGAN, Urban Area, Very High Resolution.

ABSTRACT:

The high-resolution images are in demand for many applications in the monitoring of urban areas. The advent of remote sensing satellites such as Sentinel-2 has made data more accessible as it provides free multispectral imagery. However, the spatial resolution of these images is not sufficient for many of the tasks. With the advent of deep learning techniques, significant progress has been made in the field of super-resolution, which has shown promising results in the improvement of the spatial resolution of satellite images. In this study, we compare four the most common deep learning-based models for the super-resolution of Sentinel-2 imagery in dense urban areas using aerial images. These methods are including enhanced deep super-resolution network (EDSR), super-resolution generative adversarial networks (ESRGAN), residual feature distillation network (RFDN), and Super-Resolution Convolutional Neural Network (SRCNN). To determine the effectiveness of the models in improving image resolution, they were evaluated using visual quality and quantitative metrics. The super-resolution results show that deep learning-based models have high potential for the generation of the high-resolution dataset from Sentinel-2 imagery in urban areas. The RFDN outperformed other deep learning-based models that achieved the peak signal-to-noise ratio (PSNR) more than 17.8.

1. INTRODUCTION

Recent advances in satellite technology have enabled a detailed understanding of human activity on the earth's surface. The high-resolution imagery can provide valuable information for a wide range of applications, such as the monitoring of urban areas (Duncan and Boruff, 2023). However, high-resolution images are rarely acquired over much of the planet's surface, especially in developing countries where they are desperately needed, and are unaffordable to purchase in large quantities (Latte and Lejeune, 2020). Fortunately, Sentinel-2 satellite imagery is now being provided by the European Space Agency (ESA) with coverage on a global scale (Razzak et al., 2023). The Sentinel-2 satellite imagery suffers from a low spatial resolution due to design considerations and the limitations of the sensor hardware, which makes the extraction of small objects a major challenge (Feng et al., 2023). This limitation of Sentinel-2 imagery occurs in the extraction of urban elements such as buildings and roads.

Recently, several procedures have been developed for enhancing spatial resolution of Sentinel-2 imagery based on single image. These models can be categorized in three main groups (Chen et al., 2022; Michel et al., 2022): (1) Interpolation-based methods (.i.e. Bicubic) are known for their speed and ease of implementation, but their results are limited in terms of recovering more detail. (2) Reconstruction-based models assume that the lower-resolution image is the result of degradation of the higher-resolution image through a series of known degradations. However, as the scaling factor increases, the reconstruction degrades, and their performance is severely limited by the scaling factor. (3) Deep learning based models that are the most popular frameworks for single image super-resolution.

In recent years, many deep learning based have been developed for image super-resolution. For example, Lim et al. (2017) has developed an enhanced deep super-resolution network (EDSR) for single image SR. This model uses the residual scaling factor to handle a wide range of SR factors, making it useful in many tasks. Furthermore, Wang et al. (2018) designed an enhanced super-resolution generative adversarial networks (ESRGAN) model for SR. This model is based on the traditional GAN model but uses the residual-in-residual dense block as the fundamental network framework to give more realistic results. Additionally, Dong et al. (2015) designed a lightweight structure for single image super resolution based on convolutional neural network (SRCNN). The SRCNN has three main operation that are included: (1) patch extraction and representation, (2) non-linear mapping, and (3) reconstruction. Liu et al. (2020) proposed a super-resolution framework based on a feature distillation connection called the residual feature distillation network (RFDN). In this framework, the feature distillation lead to learn more discriminative feature representations.

The main contribution lies in its evaluation and comparison of the performance of four of the most common deep learning-based SR frameworks, namely SRCNN, SRCNN, SRGAN, and RFDN, for improving the spatial resolution of Sentinel-2 imagery in urban areas and advancing research in this field. In addition, the traditional Bicubic model is implemented for more comparison. In addition, the results of SR are assessed in terms of quantitative metrics such as root mean square error (RMSE), mean absolute error (MAE), peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and visual quality using a real high-resolution dataset.

* Corresponding author

https://doi.org/10.5194/isprs-annals-X-1-W1-2023-1021-2023

2. STUDY AREA AND SATELLITE IMAGES

2.1 Study Area

Figure 1 shows the geographical location of our study area, which is located in the dense city of Berlin, Germany. The area is characterised by a high density of buildings with different types of roof structures. It is also home to a wide variety of plant and tree species. Based on the information provided in Figure 1, we divided the study area into separate locations for the test and training/validation datasets.

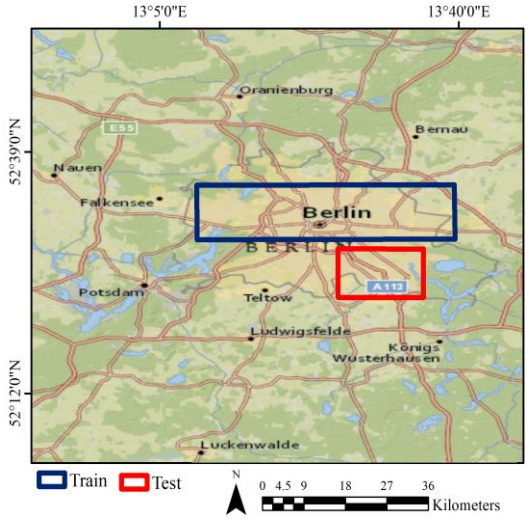

Figure 1. The location of the sample dataset in the study area.

2.2 Satellite Dataset

This study uses the Sentinel-2 imagery, as low-resolution image for SR. Sentinel-2 is a constellation of two identical satellites, Sentinel-2A and Sentinel-2B, developed by the European Space Agency. Sentinel-2 observes the Earth's surface in 13 spectral bands, with a spatial resolution of 10 to 60 metres. In addition, we have used the four visible bands (red, green, blue and near-infrared) of the Sentinel-2 Level 2A product, which provides the surface reflectance with a spatial resolution of 10 (metres).
The Sentinel 2 data sets were produced by mosaicking a number of scenes. The VHR dataset was also acquired in different epochs.
The high-resolution image used in this study was obtained from State government of Brandenburg (LGB), which takes aerial photographs covering the state of Brandenburg, Germany. The data set consists of four spectral bands (red, green, blue and near infrared). The details of incorporated datasets are represented by Table 1.

Method	VHR	Sentinel-2
Spectral Bands	4	4
Spatial Resolution	2.5 (m)	10 (m)
Radiometric Resolution	8-bit	12-bit
Acquisition Date	2023-01	2023-01

Table 1. Description of incorporated dataset.

There are a total of 567 patches in the sample dataset, with 465 of them assigned to the training dataset and the remaining 102 patches assigned to the validation dataset. In addition, it is worth noting that the test dataset was not used during the training process.

3. METHODOLOGY

Figure 2 shows the general framework of super-resolution using four deep learning models (ESRGAN, SRCNN, SRGAN, and RFDN). In addition, the Bicubic model will be used for comparison with models based on deep learning. As can be seen, super-resolution can be applied in two main steps: (1) sample data extraction and model parameter tuning, and (2) prediction and accuracy evaluation using the prediction model.
The Sentinel-2 and high-resolution images are divided into 128×128 and 512×512 patches, respectively. Furthermore, the SRCNN model receives input as 512×512 and provides output as 512×512. The sample dataset is then divided into training and validation datasets. The model parameters are initialized using the He-Normal initializer (He et al., 2015). The model is then trained on the training data set and evaluated on the validation data set using the loss function (mean square error). The error of the network is fed to an Adam (Kingma and Ba, 2014) optimizer for adjustment of the error in the whole network. This process continues until the stop condition is reached. The predictive model is then applied to the test dataset and the output of the model is evaluated by comparison with the original high resolution dataset.

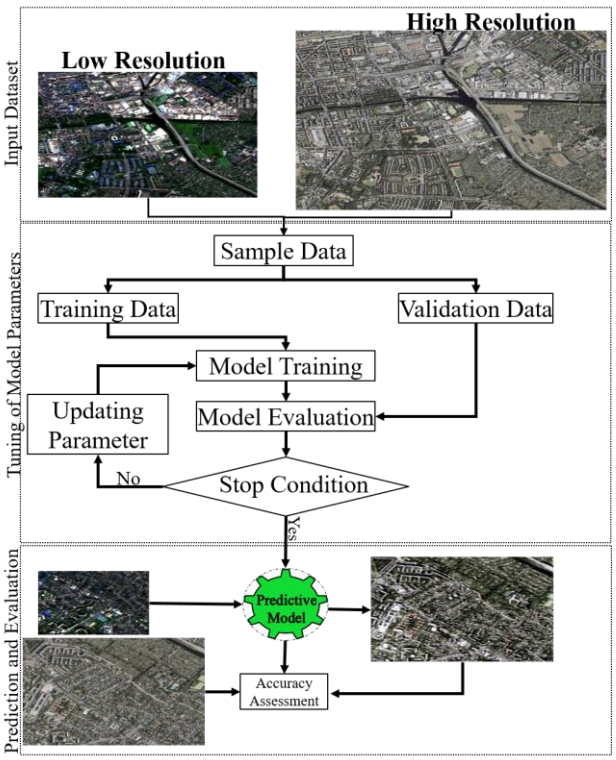

Figure 2. Overview of super-resolution through deep learning models.

3.1 ESRGAN

The ESRGAN is based on SRGAN model that has two main modification that are included removing all batch normalization layers and replacing the original basic block with the Residual-in-Residual Dense Block, which combines a multi-level residual network and dense connections (Figure 3). The generative network uses for generating high resolution image from low resolution image. Furthermore, the discriminative network try to discriminative which image is true or fake.

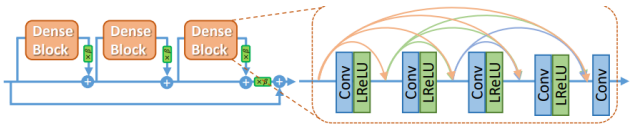

Figure 3. The structure of Residual in Residual Dense Block in ESRGAN model (Wang et al., 2018).

3.2 EDSR

The main architecture of the EDSR model is shown in Figure 4. The architecture of the EDSR is made up of several residual blocks. Furthermore, in the structure of EDSR, the batch normalisation layers are removed.

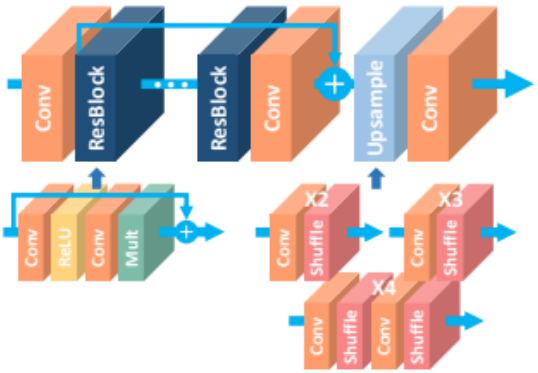

Figure 4. The architecture of the EDSR model (Lim et al., 2017).

3.3 SRCNN

The SRCNN is a simple CNN architecture that consists of three components (Figure 5). The patch extraction layer, which extracts the patches from the input data set and uses convolutional filters to represent them. The nonlinear mapping layer, which consists of 1×1 convolutional filters. The final reconstruction layer reconstructs the high-resolution image.

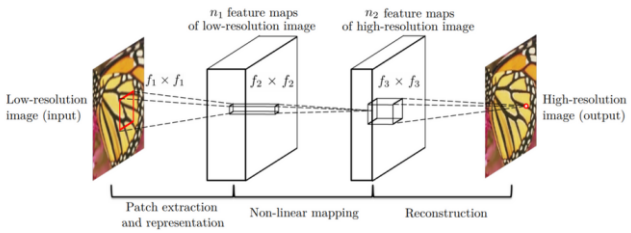

Figure 5. The architecture of the SRCNN model (Dong et al., 2015).

3.4 RFDN

The architecture of the RFDN model is shown in Figure 6. In order to learn more discriminative feature representations, the RFDN used multiple feature distillation connections. In addition to this, the shallow residual block (SRB) is used in the main building block.

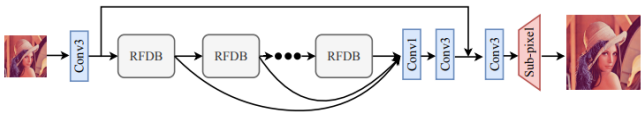

Figure 6. The architecture of the RFDN model (Liu et al., 2020).

4. RESULT

4.1 Implementation details

The deep learning models were developed using the Keras framework with the TensorFlow backend. These models were trained for 350 epochs with a learning rate of 0.001 and a batch size of 2.

4.2 Results

Results of the super-resolution are shown in Figure 7 for the test area. On this basis, all the deep learning models have improved the spatial resolution of the Sentinel-2 image. The main difference between the models is the preservation of the spectral information. The ESRGAN and the RFDN models preserved both the spectral and the spatial information, whereas the EDSR and the SRCNN models failed to preserve the spectral information.

The performance of four models was further investigated by adding a zoom range. Figure 8 is an illustration of the performance of the deep learning models in the zoom region. As can be seen, the EDSR has provided the weakest performance in the recovery of the details. The SRCNN has provided the performance in spatial information, but lacks the spectral information, which is more evident in red buildings. The spectral information was better preserved than the spatial information in the Bicubic model. The ESRGAN and RFDN have the better performance compared to the other two models. The edge of the building was also preserved by these models.

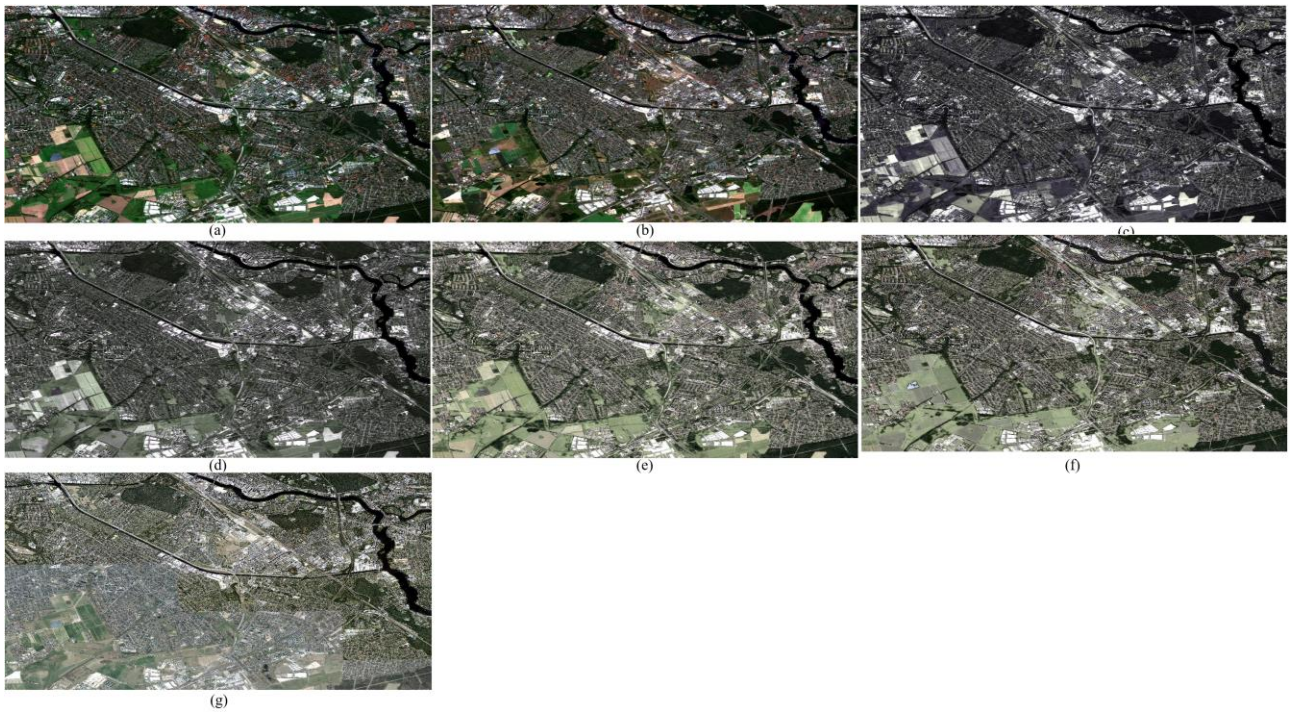

Figure 7. The result of super-resolution models for test area: (a) Sentinel-2, (b) EDSR, (c) SRCNN, (d) ESRGAN, (e) RFDN, and (f) Original VHR dataset.

The numerical results of the super-resolution are presented in Table 2. As can be seen, the RFDN method performs best overall, achieving the lowest values for MAE (25.64) and RMSE (32.53), indicating that the algorithm was able to accurately reproduce the spectral information of the high-resolution image from the low-resolution image. The RFDN method also achieves the highest PSNR (17.88), indicating that its high-resolution images are closest to the original low-resolution images. The RFDN method also has the highest SSIM value of 0.344, which is structurally similar to the original images. The ESRGAN method also performs well, achieving the second highest values for all four metrics, closely followed by the SRCNN method. The EDSR method performs worst overall, achieving the highest values for MAE and RMSE, and the lowest values for PSNR and SSIM. The Bicubic achieved the worst performance in SR comparison with other models.

Method	MAE	RMSE	PSNR	SSIM
Bicubic	100.730	109.308	7.35	0.025
EDSR	35.77	47.29	14.63	0.046
SRCNN	28.38	35.31	17.17	0.280
ESRGAN	26.13	32.90	17.78	0.330
RFDN	**25.64**	**32.53**	**17.88**	**0.344**

Table 2. Comparison of numerical results from SR for the test area.

Table 3 shows the comparison SR results with original Sentinel-2 images. Based on these numerical results, the Bicubic method preserved the most content of spectral information as it has provided MAE and RMSE lower than 41 and 42, respectively. Furthermore, among deep learning models RFDN has provided the best performance in keeping spectral information.

Method	MAE	RMSE	PSNR	SSIM
Bicubic	**41.05**	**42.40**	**10.51**	**0.143**
EDSR	79.00	88.87	10.51	0.143
SRCNN	69.08	70.25	10.51	0.143
ESRGAN	69.05	71.68	10.51	0.143
RFDN	64.99	68.20	10.51	0.143

Table 3. Comparison of numerical results from SR between results and original Sentinel-2 imagery for the test area.

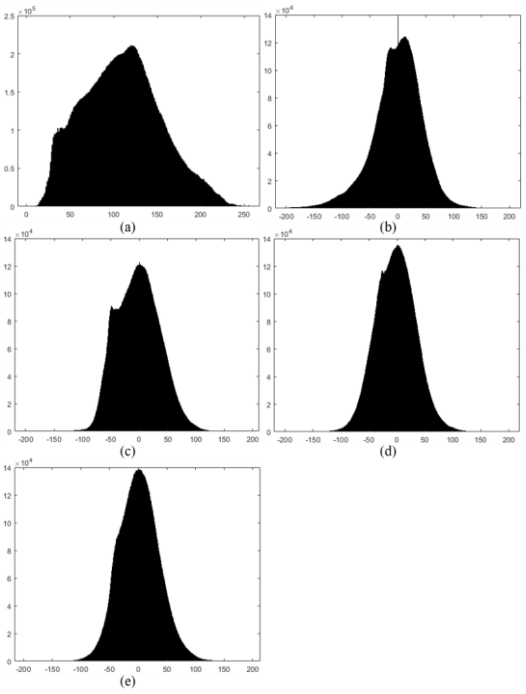

Figure 9. Plot of histogram of super-resolution results and high-resolution image for deep learning models: (a) Bicubic, (b) EDSR, (c) SRCNN, (d) ESRGAN, and (e) RFDN.

The histogram plots in Figure 9 show the differences between the high-resolution image and the super-resolution results produced by each model. In particular, the histograms show the distribution of pixel differences between the original image and super-resolution results. The histograms show, as indicated by the lack of difference values around zero, that some objects were not detected by the EDSR and SRCNN models. Among four models, the RFDN has provided the best performance for super-resolution.

Computational cost refers to the time and resources required to train and run a model on a given hardware setup in the context of super-resolution deep learning models. Table 4 compares the number of parameters for each deep learning model, which is an indicator of model complexity and memory requirements. Among the four super-resolution models, SRCNN has the fewest parameters, followed by EDSR, RFDN and ESRGAN, which has the most.

Method	EDSR	SRCNN	ESRGAN	RFDN
Number of Parameters	1,518,147	**25,283**	3,100,867	1,221,763

Table 4. Comparison of numerical results from SR for the test area.

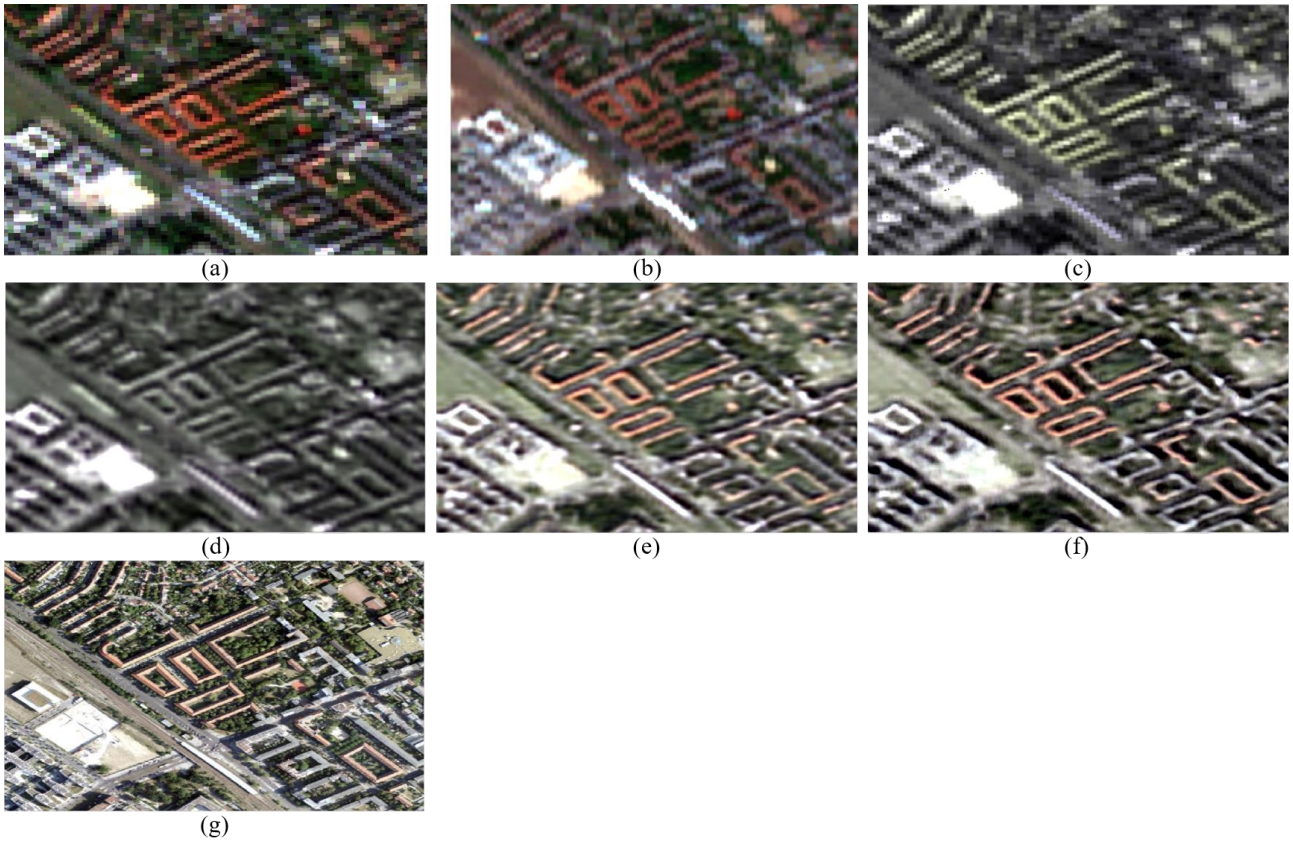

Figure 8. The zoom area of the super-resolution results: (a) Sentinel-2, (b) Bicubic, (c) EDSR, (d) SRCNN, (e) ESRGAN, (f) RFDN and (g) Original VHR dataset.

CONCLUSION

In this paper, we compared four deep learning models for super-resolution Sentinel-2 imagery using a single image. We evaluated the performance of the models in urban areas. First, we evaluated the results of SR based on comparison with VHR dataset. The numerical and visual analysis of the super-resolution results shows that the deep-based models were able to preserve both the spatial and the spectral information comparison with Bicubic model. Furthermore, the ESRGAN and RFDN models were able to preserve both the spatial and the spectral information, while the EDSR and SRCNN models were not able to preserve the spectral information. The RFDN method performed best overall, achieving the lowest values for MAE and RMSE, and the highest values for PSNR and SSIM among the deep learning models. Second, we compared the SR results with the original Sentinel-2 images as a measure of spectral information retention. Based on the numerical and quantitative results, the Bicubic model has a high degree of similarity with the original Sentinel-2 image. Thus, compared to other deep learning based models, the Bicubic model preserves more content of spectral information.

The RFDN with more parameters was more robust and produced higher quality super-resolution results compared to the other deep learning based models. However, it also required more computational resources to train and run. Overall, the RFDN model provided the best performance for super-resolution.

ACKNOWLEDGEMENTS

We would like to acknowledge the use of the "Data licence Germany – attribution – version 2.0 available at http://www.govdata.de/dl-de/by-2-0" in our project. This license enabled us to utilize and process the data and metadata provided under the conditions stated in the license. We confirm that we have fulfilled all the requirements stated in the license for non-

commercial use of the data. We are grateful to the provider for making this data available to us under this license.

REFERENCES

Data licence Germany – attribution—Version 2.0 available at http://www.govdata.de/dl-de/by-2-0

Chen, H., He, X., Qing, L., Wu, Y., Ren, C., Sheriff, R. E., and Zhu, C.: Real-world single image super-resolution: A brief review, Information Fusion, 79, 124-145, 2022.

Dong, C., Loy, C. C., He, K., and Tang, X.: Image super-resolution using deep convolutional networks, IEEE transactions on pattern analysis and machine intelligence, 38, 295-307, 2015.

Duncan, J. M. and Boruff, B.: Monitoring spatial patterns of urban vegetation: A comparison of contemporary high-resolution datasets, Landscape and Urban Planning, 233, 104671, 2023.

Feng, L., Xu, P., Tang, H., Liu, Z., and Hou, P.: National-scale mapping of building footprints using feature super-resolution semantic segmentation of Sentinel-2 images, GIScience & Remote Sensing, 60, 2196154, 2023.

He, K., Zhang, X., Ren, S., and Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, Proceedings of the IEEE international conference on computer vision, 1026-1034,

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Latte, N. and Lejeune, P.: PlanetScope radiometric normalization and sentinel-2 super-resolution (2.5 m): A straightforward spectral-spatial fusion of multi-satellite multi-sensor images using residual convolutional neural networks, Remote Sensing, 12, 2366, 2020.

Lim, B., Son, S., Kim, H., Nah, S., and Mu Lee, K.: Enhanced deep residual networks for single image super-resolution, Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 136-144,

Liu, J., Tang, J., and Wu, G.: Residual feature distillation network for lightweight image super-resolution, Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16, 41-55,

Michel, J., Vinasco-Salinas, J., Inglada, J., and Hagolle, O.: SEN2VENμS, a dataset for the training of Sentinel-2 super-resolution algorithms, Data, 7, 96, 2022.

Razzak, M. T., Mateo-García, G., Lecuyer, G., Gómez-Chova, L., Gal, Y., and Kalaitzis, F.: 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.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks, Proceedings of the European conference on computer vision (ECCV) workshops, 0-0,