How Does Super-Resolution for Satellite Imagery Affect Different Types of Land Cover? Sentinel-2 Case

Anna Malczewska and Maciej Wielgosz

***Abstract*—In the dynamic field of satellite imagery, the significance of super-resolution (SR) techniques, grounded on advanced deep learning methods, is paramount. A thorough understanding and remediation of the distinct challenges posed by various land cover types for image resolution enhancement form the essence of this research. This work diligently employs two unique neural networks, SRCNN and SwinIR Transformer, to scrutinize their varying impacts on a range of land cover types, ensuring a detailed and comprehensive exploration. This study transcends the mere enhancement of the Sentinel-2 dataset's resolution from 20 m/pix to 10 m/pix. It ambitiously seeks to excavate the intricate trends inherent to different land cover types and their corresponding interactions with SR processes. The application of neural networks on 255 × 254 pixel patches, covering six dominant types—forests, large fields, small fields, urban, sub-urban, and mixed—highlights substantial variations in metrics, underlining the individual interactions of each land cover type with SR techniques. A comprehensive accuracy assessment is meticulously conducted, employing an array of metrics and frequency domains to shed light on the nuanced differences and provide vital insights for optimizing each land cover type's SR approaches. Notably, the PSNR metric reveals significant disparities, particularly in the "forest" and "urban" categories for both SRCNN and SwinIR. According to the PSNR metric, the "forest" class yielded the best results with 66.06 for SRCNN and 67.00 for SwinIR, while the "urban" class marked the lowest with 55.09 and 57.02, respectively, reinforcing the critical nature of this study.**

***Index Terms*—Image enhancement, image super-resolution (SR), optical image processing, quality control, remote sensing.**

Nomenclature

Abbreviations

The following abbreviations are used in this manuscript:

AMFFN	Adaptive multiscale feature fusion network.
BOA	Bottom-of-atmosphere.
CNN	Convolutional neural network.
DMCN	Deep memory connected network.
DMOS	Differential average subjective score.
EDSR	Enhanced deep super-resolution network.
ERGAS	Erreur Relative Globale Adimensionnelle de Synthese.
ESA	European Space Agency.
HR	High resolution.
IBP	Iteration back projection.
LPIPS	Learned perceptual image patch similarity.
LR	Low-resolution.
MOS	Mean opinion score.
NASA	National Aeronautics and Space Administration.
NIQE	Natural image quality evaluator.
OA	Overall accuracy.
PNN	Pansharpening convolutional neural network.
PCA	Principal component analysis.
POCS	Projection onto convex sets.
PPV	Positive predictive value.
PSNR	Peak signal-to-noise ratio.
QNR	Quality no-reference.
RMSE	Root-mean-square error.
SAM	Spectral Angle Mapper.
SCC	Spatial correlation coefficient.
SR	Super-resolution.
SRCNN	Super-resolution convolutional neural network.
SSIM	Structural Similarity Index.
SVD	Singular Value decomposition.
SVR	Support vector regression.
SwinIR	Image restoration using Swin Transformer.
TOA	Top-of-atmosphere.
TPR	True positive rate.
TTSR	Texture transformer network for image super-resolution.
UIQI	Universal Image Quality Index.
USGS	United States Geological Survey.
VDSR	Very deep super-resolution network.

Manuscript received 9 June 2023; revised 5 August 2023 and 4 October 2023; accepted 14 October 2023. Date of publication 31 October 2023; date of current version 23 November 2023. *(Corresponding author: Anna Malczewska.)*

Anna Malczewska is with the Department of Photogrammetry, Remote Sensing of Environment and Spatial Engineering, AGH University of Krakow, 30-059 Krakow, Poland (e-mail: zadlo@agh.edu.pl).

Maciej Wielgosz is with the Faculty of Computer Science, Electronics and Telecommunications, AGH University of Krakow, 30-059 Krakow, Poland (e-mail: wielgosz@agh.edu.pl).

Digital Object Identifier 10.1109/JSTARS.2023.3328997

I. Introduction

MULTISPECTRAL satellite images are widely used in many issues related to land and coastal environment monitoring, such as determining the level of afforestation and built-up area or in emergencies by detecting fires as well determining the effects of floods. Several sensors can provide global

acquisition of multispectral images with high revisit frequency. Satellite data can be divided due to spatial resolution into three groups: low/medium (over 60 m/pix), high (60–10 m/pix), and very high (less than 10 m/pix) resolution. The two largest sources of free-of-charge images referred to high-resolution (HR) are Copernicus Program provided by European Space Agency (ESA) and Landsat Program led by NASA and USGS.

Sentinel-2 (ESA) images, which are used in this experiment having pixel size from 60 to 10 m depend on the band. For many purposes, this range is enough but for some analysis, it is very appreciated to have a higher resolution. Data fusion, pan-sharpening, and image super-resolution (SR) aims to reconstruct low-resolution (LR) image to HR image. Machine learning (ML) and deep learning (DL) methods achieve better results in almost every performance metric than previous "classical" methods based on pansharpening algorithms [1].

Due to the specificity of satellite images, it is necessary to adjust the methods. It is not obvious that DL models creating and learning on standard images can directly translate to remote sensing images. Compare to standard images, differences are caused by data specifications and can be seen in many aspects, such as the number of spectral channels, data recording type, or even difficulties in obtaining datasets. Satellite images are collected under different weather conditions, such as variable lighting, clouds, and fog. One more difference between SR on standard images compared to satellite images is the source of a LR dataset. For the standard approach, it is very often to make LR images by degradation of HR images. In that case, there is a possibility that the neural network model learns how to enhance image resolution by correcting the degradation methods [1]. Satellite images are collected with different resolutions so there is no risk that the model learns how to correct the degradation process. The challenging aspect is that very often LR images are collected by different sensors than HR images. To ensure accuracy, it is crucial that images should be recorded with the closest possible time and cover identical geographic regions. Moreover, satellite data after SR are very often a subject of further processing, such as image segmentation, Land Use Land Cover classification, or object detection. Considering that, the SR process in the case of satellite images should not only improve the visual impression of the image, but also keep spectral and spatial information to prevent quality for further processing. There are many different accuracy metrics that focus on checking for similarities between images. In this article, we checked their suitability for the specificity of satellite data.

The experiment presented in this article focuses mainly on developing a comprehensive quality evaluation method for satellite image SR, and assessing how different land cover types, such as forests, large fields, small fields, urban, sub-urban, and mixed areas, are affected by the process. The intention of this research was not to compare a multitude of SR models, but to deeply understand the impact of SR on different types of land cover. We hope that this approach allows to understand how SR neural networks can be used for actual problems with specific land cover types. For SR, we chose two neural networks with different levels of complexity and theoretical backgrounds:

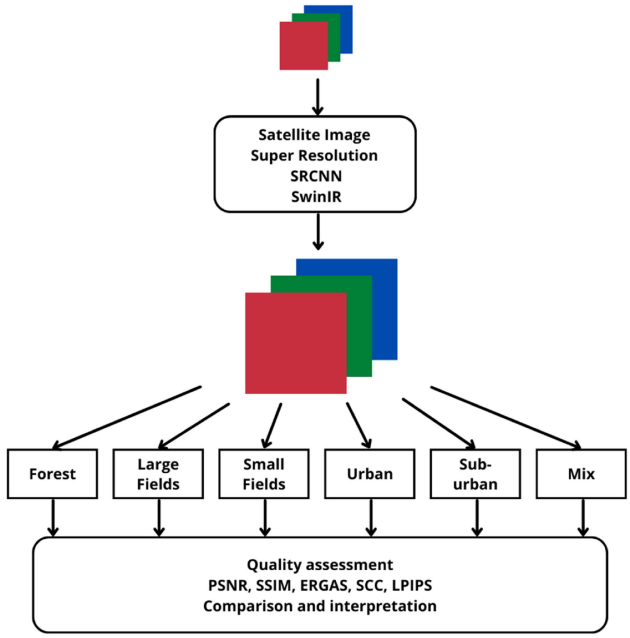

Fig. 1. Simplify flowchart of the experiment.

SRCNN [2] and SwinIR [3]. SRCNN is one of the pioneering convolutional neural network (CNN) models for SR and has demonstrated its robustness in handling Sentinel-2 images [4]. On the other hand, SwinIR, with its more recent development and cutting-edge approach, offered a new perspective in exploring the SR of satellite images. The evaluation for both methods was done by using selected metrics and approaches. The attention was paid to the diversity of results in relation to different types of land cover. In addition, we tested which of image similarity comparison metrics can be helpful in unsupervised classification of images.

The general workflow is presented on the flowchart in Fig. 1.

For this experiment, we used Sentinel-2 images, R, G, and B channels (B2, B3, B4) and enhanced resolution by 2. These channels are consistently distributed by the ESA at resolutions of 10 and 20 m, therefore they provide a stable and standardized base for our experiment. Moreover, we decided to limit the number of channels and level of SR to $\times 2$ in order to maintain the compatibility and integrity of the model SwinIR, which was fine-tuned. The $\times 2$ upscaling case is also extensively studied and used in literature and real world applications. We used this basic dataset in order to focus on developing an approach, which can be adopted to include different spectral bands and higher resolution input from various sensors.

This research endeavors to unravel the following queries: How does SR influence various land cover types? Does its impact remain uniform across diverse methods? Can we leverage this knowledge to enhance neural network architectures for specific applications?

This article primarily contributes the following.

1) Development of a technique utilizing DL models to individually handle diverse land cover types, focusing particularly on the crucial, smaller sections of images.
2) An in-depth comparative analysis of the SR outcomes for six different land cover types, thoroughly assessing the influence of SR on each.
3) Introduction of a pioneering, unsupervised protocol for the classification of different land cover types.
4) Customization of the DL models to work efficiently with the provided dataset, with the source code made available at the repository [5].

It is possible to improve quality of small selected patches in the images by using the proposed method which is composed of set of models, portfolio of metrics, and the interpretation schemes of their values. The key notion is that the application of the DL model should be done selectively (not globally) with and appropriate set of interdependent metrics and their dedicated angle of analyze which we propose in this work.

The rest of article is organized as follows. Section II summarizes in general ML and DL methods used in remote sensing, as well-detailed SR methods used on satellite images. Section III presents our approach, dataset, metrics, and SR methods. Section IV gives results, both SR and accuracy assessment connected to land cover types, as well unsupervised classification. Section V includes a discussion of the results. Finally, Section VI concludes this article.

II. Related Works

A. ML and DL in Remote Sensing

ML and DL have revolutionized remote sensing by enabling more accurate, efficient, and automated analysis of Earth Observation data. The capabilities of ML and DL models have demonstrated their superiority in handling the complexity and scale of remote sensing data, ultimately leading to more advanced applications in environmental monitoring. A comprehensive review, progress of the methods, and the potential of the ML and DL were widely discussed in [6], [7], [8], [9], and [10]. Frequently developed topic in the field of ML and DL in Earth Observation is image segmentation, which is very often called Land Use Land Cover classification. Review articles and meta-analyses have been provided in [11], [12], [13], and [14]. New methods and highly advanced models for image segmentation have been included in [15] and [16]. The application of satellite images for monitoring urban areas are presented in papers [17] and [18], and for agricultural area in [19], [20], and [21]. A big challenge and area of research is also the classification based on hyperspectral images due to their spectral feature representations. The using of DL in the classification of hyperspectral data (HSI) is presented in [22] and [23]. Due to their complexity, hyperspectral images are demanding to obtain. DL methods are also used to prepare such data in the HSI reconstruction phase in snapshot-hyperspectral compressive imaging (SCI) systems. An example of such a solution, also using remote sensing data, is presented in [24]. Other applications of ML and DL models in remote sensing are image registration [25], object detection [26], [27], change detection [28], [29], [30], and SR, which is detailed described in the next section.

B. Super-Resolution

The comprehensive review of application of SR models based on DL to solve remote sensing task is presented by Wang et al. [31] and Wang et al. [32]. Some of the multispectral images consist of bands with different resolutions within one product. For example, Sentinel-2 only has 2, 3, 4, and 8 channels in 10 m/pix, the rest is 20 m/pix or 60 m/pix. One of the SR approach for satellite images has a goal to enhance lower-resolution bands to the highest available in the product. Several works are based on convolutional layers.

Pansharpening CNN method (PNN) trained on GeoEye-1, WorldView-2, and Ikonos data was present in [33]. Two more versions of improved PNN were presented in [34] and [35]. Another pansharpening network (PanNet) [36] was created based on ResNet [37]. Modifications in network architecture added by authors could capture spectral and spatial preservation. CNN-based fusion method on Sentinel-2 data was presented and developed in [38], [39], and [40]. A more global approach was presented by Lanaras et al. [41]. In this work, not only 20 m Sentinel-2 bands were enhanced to 10 m, but also 60 m. Network architecture was inspired by enhanced deep super-resolution (EDSR) [42] and ResNet [37]. VDSR [43] was adapted to the remote sensing purpose [44] and uses residual factors, which stabilize learning and predict only high-frequency components, while low-frequency components are transferred directly from the input.

Researchers also focused on solutions which could super-resolve, e.g., Sentinel-2 bands to more than 10 m. To use ground truth data from the different sensors, it is necessary to take into account some specific characterization of images. First, very important is to find registrations which are temporarily close to each other to minimize changes that could appear. Second, the spectral bands of each data source should be as similar as it is possible. Also, prepossessing such a histogram matching could be relevant. The EDSR-based model and its combinations are widely used [45], [46]. RapidEye and PlantScope satellite images were used to obtain Senitnel-2 RGB and Nir bands in 5 and 2.5 m pixel resolution, respectively. A solution based on a residual convolutional network was also done [47]. One more paper present model based on residual network with spectral attention mechanism [48]. The model was trained on Senitnel-2 and PlanetScope images.

In DMCN, authors proposed different memory connections to combine residual information with output: local and global memory connection [49]. This solution was tested on three remote sensing HR datasets. Authors of adaptive multiscale feature fusion network (AMFFN) used several adaptive multiscale feature extraction modules to obtain a SR network, which is more independent of scale, pixel size, and source of the photo. Currently, one of the most promising methods used in DL is based on transformers. An example of texture transformer network for image super-resolution (TTSR) shows that the network learns to search for relevant textures from reference images for a target

region on the LR image [50]. Some of the solutions connect CNN and transformer components to obtain higher results and keep the advantages of these two methods [51], [3]. HyperTransformer was designed for pansharpening so used PAN as HR channel and hyperspectral channels as LR [52].

Hyperspectral images are an interesting area where SR methods are developed. As in the case of multispectral images, three resolution enhancement methods are specified: pansharpening-based approaches, DL-based approaches, and factorization-based approaches. HSI are often characterized by high spectral resolution but lower spatial resolution. Obtaining pairs of HR and LR hyperspectral images is often difficult, so researchers are focused on fusing HR multispectral data (MSI) with lower-resolution hyperspectral data (LR-HSI). Examples of such DL-based studies are: [53], [54]. For the fusion of HR-MSI and LR-HSI pairs, factorization-based approaches have been widely used. In this group of methods, the decomposition aspect is important, which allows to remove redundant information and translate high-dimensional data into low-dimensional. Chen et al. [55] presented factor-smoothed tensor ring (TR) decomposition method, which allow to explore the high spatial-spectral correlation of HR-HSI. This method is effective to reconstruct an HR-HSI from a pair of LR-HSI and HR-MSI of the same scene.

C. Quality Evaluation

Assessing the quality of super-resolved images plays an important role in understanding outputs and developing effective methods. Especially in cases when super-resolved images are used in analysis such as segmentation, LULC classification or object detection. Images quality could be assessed in two ways: subjective and objective. The subjective evaluation is highly connected to human perception and impression. The criteria can be a natural-looking and realistic image, easily recognizable details, and no artefacts and distortions. In remote sensing, it is very important in cases when images will be used in photointerpretation making by a human. An example of the subjective method is MOS [56] or DMOS. The objective evaluation methods focus on a numerical calculation using mathematical, statistical, and more often neural networks algorithm. In most cases, objective methods are based on a similarity comparison between a super-resolved image and a ground true HR image. Examples of widely used objective methods are: peak signal-to-noise ratio (PNSR) [57], structural similarity (SSIM) [57], and NIQE [58]. For remote sensing data commonly used are: *Erreur Relative Globale Adimensionnelle de Synthese* (ERGAS) [59], Spectral Angle Mapper (SAM) [60], Spatial Correlation Coefficient (SCC) [61], and UIQI [62]. A relatively new metric based on the DL approach is learned perceptual image patch similarity (LPIPS) [63]. There are also some metrics to assess the quality in case of the absence of ground truth. In remote sensing metric QNR [63] is used, but to calculate it HR panchromatic band is needed.

D. Application of Super-Resolved Satellite Images

Research on the SR impact on LULC classification or object detection of satellite images are less undertaken than the development of SR methods themselves. Nevertheless, there are articles that describe such results. Effects of atmospheric correction and pansharpening on LULC classification were tested on WorldView-2 data [64]. Successfully found that classification accuracy increased after using pansharpening PCI-PanSharp algorithm. Two classification methods were used: object-based support vector machine(OB-SVM) and pixel-based maximum likelihood classifier (PB-MLC). Studies based on the classical pan-sharpening approach have also been performed on Landsat-8 images [65]. Authors shown that improving spatial resolution is highly beneficial for crop type differentiation using both object- and pixel-based approach of LULC classification. Similar conclusions were obtained by checking the neural SRCNN on Landsat-7 images [66]. The enhancement of resolution of image reduce the number of misclassified pixels compare to results obtained on original images.

III. Materials and Methods

A. Data Specification and Preparation

The Copernicus program run by ESA provides free-of-charge access to the earth observation data. Among many missions, Sentinel-2 is mainly focused on monitoring land, seaside, and islands area, except Antarctica. Data are collected by two identical satellites with the same sun-synchronise, quasi-circular, near-polar, low-earth orbit but with a phase difference of 180°. The time to revisit the same place for each one is 10 days so both give images up to 5 days. Each scene contains 13 multispectral channels presenting an electromagnetic spectrum from visible to short-wave infrared. Images are acquired with a radiometric resolution of 12 b/pix, encoded and distributed in 16 b/pix JPEG2000 format. Each scene covers an area of 100 km $\times$ 100 km with an image size of 10 980 $\times$ 10 980 pixels. Data provided by ESA are available in spatial resolutions of 10 m, 20 m, and 60 m with the highest resolution different for each set of bands. There are 4 bands (visible and near-infrared) in 10 m, 6 bands (near-infrared and short-wave infrared) in 20 m and 3 bands in 60 m. Data are available in two types of correction: TOA (Level-C1) and BOA (Level-A2), which is now called "surface reflectance." We used Level-2 A data that are orthorectified and BOA-calibrated reflectance. Level-2 A products for R, G, and B channels are distributed by ESA in all resolutions: 10, 20, and 60 m. For this experiment, we used channels, which represent red, green, and blue colors (B2, B3, and B4) with 20 m spatial resolution as LR images and 10 m spatial resolution as HR images. Since both 10 and 20 m resolutions are available, there was no need to perform data degradation to obtain LR images. The dataset was created with different types of land cover from different parts of the world. The arrangement of the scenes is shown in Fig. 2. The list of tiles of scenes used for this experiment is shown in the Appendix in Table XIV.

Basic preparation for dataset was as follows: we divided each scene into smaller patches of size of 2040 $\times$ 1524 pix for HR images and 1020 $\times$ 762 pix for LR images. As we want to keep parts of every scene in each training, validation, and testing set so each one was randomly split into each set. The detailed number

Fig. 2. Location of the Sentinel-2 images used in the experiment. Image source: OpenStreetMap.

TABLE I
TRAINING, VALIDATION, AND TESTING NUMBER OF IMAGES (2040 × 1524 HR, 1020 × 762 LR)

Model	Training	Validation	Testing
SRCNN	950	90	10
SwinIR	950	90	10

of images for both networks are shown in the Table I and describe in Sections SRCNN III-B1 and SwinIR III-B2.

B. SR Neural Networks Used for Experiment

For this experiment, two SR models of neural network were used: SRCNN [2] with modifications and the model based on Swin Transformer (SwinIR) [3]. Both methods differ in the level of complexity, number of parameters, and in the general way of upsampling images. For this experiment, we upscale images by factor 2.

1) SRCNN: SRCNN was one of the first DL architectures created for SR. It consist of convolutional layers with relatively large filter sizes to account for bigger perception field. We based on publicly available Python repository implemented in PyTorch [2].

We have introduced modifications, which allowed to adopt the model to the data of our use. Consequently, the model was adapted to the three channels of satellite images with a radiometric resolution of 16-bit. We skip conversion to YCbCr color space in favor of working on each of the three channels. Most DL networks for image processing are designed to work on 8-bit images, but we keep the original for our dataset's 16-bit radiometric resolution. We also do not need to create LR dataset by degradation because we used 20 m bands as LR. We slightly change the architecture by adding residual connection after convolutional layers. The architecture of the modified SRCNN network is shown in Fig. 3.

Originally, the network was trained on a 91-image dataset broken down into smaller patches. The benchmark dataset contains 24 800 patches in size 33 × 33 pix created with stride 14. In our case, we used for training 950 images in size 2040 × 1524 pix for HR images and 1020 × 762 pix for LR images, which were divided into 299 250 patches in size 96 × 96 of

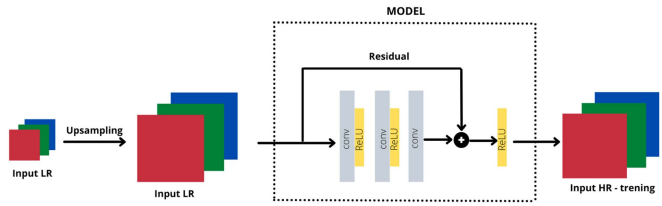

Fig. 3. Scheme of the SRCNN network architecture with changes.

TABLE II
TRAINING PARAMETERS OF SRCNN

Parameter	Value
learning rate	10e-4
batch size	128
epochs	200
loss	MSE

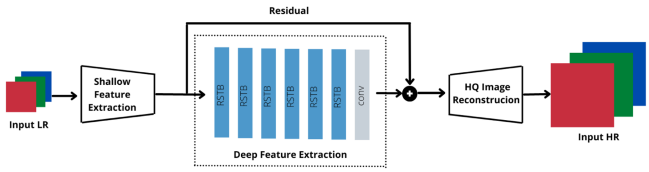

Fig. 4. Scheme of the SwinIR network architecture with Residual Swin Transformer Block (RSTB).

HR and 48 × 48 for LR. For validation, we used 90 images in size 2040 × 1524 pix for HR images and 1020 × 762 pix for LR images and of testing ten images in the same size. The parameters of the model were as follows: three convolutional layers with kernel size (9-5-5), number of filters 64 and 32, and the activation function was ReLU.

The training parameters of the model are given in Table II.

2) SwinIR: The second model which we adapt to our experiment was SwinIR, the network for image restoration using Swin Transformer [3]. The network consists of three parts: shallow feature extraction, deep feature extraction, and high-quality image reconstruction. Shallow feature extraction focus on low-frequencies parts and consists of 3 × 3 convolutional layers. The deep feature extraction model focus on recovering lost high frequencies and is composed of several residual Swin Transformer layers block (RSTB). High-quality image reconstruction part aggregates feature extraction models and use a subpixel convolutional layer to upsample the feature. The architecture of the network we used is shown in Fig. 4.

For our purpose, we fine-tune the network which is available in the publicly Python repository on GitHub [3], implemented in PyTorch. We did not change the architecture of the model. We used the parameters, which are described below. The patch size of the training dataset was 96 × 96 pix for HR and 48 × 48 pix for LR. The training parameters of the model are give in Table III.

C. Data Preparation for Detailed Quality Assessment

Data for detailed quality assessment of the different types of land cover were prepared by selection from validation and test

TABLE III
TRAINING PARAMETERS OF SWINIR

Parameter	Value
learning rate	2*10e-4
batch size	64
window size	8
iterations	60000
loss	L1

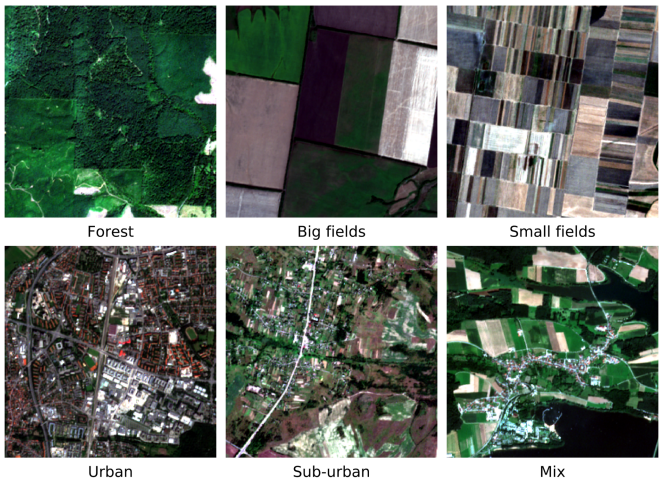

Fig. 5. Examples of six classes with different land cover.

images. Each patch was divided into small subpatch in size 255 × 254 pix so we obtained 4800 subpatches. The concept involved creating patches of a size small enough to represent a single land cover type. Some of the patches represent homogeneous nature objects like deserts or steppes. We are not taken into account this type of class in detailed quality assessment. From this huge dataset, we selected around 65–100 images for each class: forest, large fields, small fields, urban, sub-urban, and class, which represent a mix of types in one image (forest, fields, buildings, roads, and rivers). Examples of every type of land cover class are shown in Fig. 5.

D. Quality Assessment Approach

We basically wanted to determine the impact of SR for different types of land cover using known methods of quality evaluation. We used an approach based primarily on numerical calculation, histogram comparison as well other methods based on neural network and discrete Fourier transform. To compare super-resolved images to the original one we choose metrics as follows. *HR* is an HR ground truth image and *SR* is super-resolved image, which we want to compare to HR.

Root mean square error

$$\mathrm{RMSE} = \sqrt{\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[\mathrm{HR}(i,j) - \mathrm{SR}(i,j)]^2} \quad (1)$$

with *m*, and *n* being the size of the image.

Peak signal to noise ratio (PSNR) [57] is the ratio between the maximum possible power and the power of corrupting noise and is based on mean squared error (mse)

$$\mathrm{PSNR} = 10 * \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{mse}}\right) \quad (2)$$

with MAX being the maximum possible pixel value of the image, for this case is 2^{16}.

SSIM [57] takes into account the influence of three components: structural information, illumination, and contrast

$$\mathrm{SSIM} = \frac{(2\mu_{\mathrm{HR}}\mu_{\mathrm{SR}} + c_1)(2\sigma_{\mathrm{HRSR}} + c_2)}{(\mu_{\mathrm{HR}}^2 + \mu_{\mathrm{SR}}^2 + c_1)(\sigma_{\mathrm{HR}}^2 + \sigma_{\mathrm{SR}}^2 + c_2)} \quad (3)$$

with μ_{HR} being the pixel sample mean of HR; μ_{SR} being the pixel sample mean of SR; σ_{HR}^2 being the variance of HR; σ_{SR}^2 being the variance of SR; σ_{HRSR} being the covariance of HR and SR. $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$, where L is the dynamic range of pixel value (in this case is $2^{16} - 1$), and $k_1 = 0.01$ and $k_2 = 0.03$ by default.

Less used in general but more common in assessing the quality of reconstructing satellite images are *SAM* [60], which shows the deviation between true and estimated spectral signatures and was calculated for each pixel and then averaged over the whole image. The metric ignores absolute brigantines and measures how faithful the spectral distribution of a pixel is reconstructed

$$\mathrm{SAM} = \cos^{-1}\left(\frac{\sum_{i=1}^{nb}\mathrm{HR}_i\mathrm{SR}_i}{\sum_{i=1}^{nb}\mathrm{HR}_i^2\sum_{i=1}^{nb}\mathrm{SR}_i^2}\right) \quad (4)$$

with nb is number of bands in compared image.

SCC [61] is a correlation coefficient between a high-pass filtered SR image and a high-pass filtered HR original image. This metric is especially for images or part of the images concentrated in the high-frequency domain. The mask of high-pass filter, which was used, is given as follows:

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}. \quad (5)$$

Erreur Relative Globale Adimensionnelle de Synthese (ERGAS) [59] shows the overall error of the product, which is independent of the units and instruments, independent of the number of spectral bands under consideration, and independent of the ratio of the scales

$$\mathrm{ERGAS} = 100\frac{h}{l}\sqrt{\frac{1}{nb}\sum_{i=1}^{nb}\left(\frac{\mathrm{RMSE}(\mathrm{HRSR}_i)^2}{\mathrm{HRmean}_i^2}\right)} \quad (6)$$

with nb being number of bands in compared image, and *h/l* is a ratio between sizes of the images (in that case 1/2, because we super-resolved images by 2).

LPIPS [63] is a deep network-based solution to assess the similarity of the super-resolved image to the original HR image suchlike human perception.

For each class, we made a visual interpretation for color composition that is easier to interpret [67]. We also compare histograms for original and super-resolved images for each band for selected images. The shape of both expects to be close to each other. In order to obtain an unambiguous measure that will allow to compare the discrepancies between the histograms, we

calculated *Jensen–Shannon distance (JS)* as a square root of the JS divergence. JS is a method of measuring the similarity between two probability distributions and is based on the Kullback–Leibler divergence. The JS divergence is symmetric and has a finite value

$$a = \frac{1}{2}\sum_{x\in X} H_{\mathrm{HR}}(x)\log_2 \frac{H_{\mathrm{HR}}(x)}{\frac{1}{2}(H_{\mathrm{HR}}(x)+H_{\mathrm{SR}}(x))},$$

$$b = \frac{1}{2}\sum_{x\in X} H_{\mathrm{SR}}(x)\log_2 \frac{H_{\mathrm{SR}}(x)}{\frac{1}{2}(H_{\mathrm{HR}}(x)+H_{\mathrm{SR}}(x))},$$

$$\mathrm{JS}(H_{\mathrm{HR}}, H_{\mathrm{SR}}) = \sqrt{a+b} \tag{7}$$

where H_{HR} and H_{SR} are histograms of HR and SR, respectively.

Moreover, we made a *Discrete Fourier Transform* comparison that is well represented by graphs. We calculate DFT over each image, which represent the energy distribution for each ones. Graphs show how the difference between images change from the low to the high-frequency domain. Most methods of SR do well with low frequencies but high frequencies are something that need to be improve.

E. Clustering Image Patches Based on SR Results

The additional experiment in this study focused on checking the reliability of clustering of image patches representing land cover types based on SR results. We prepared the classification of images into six classes (forest, large fields, small fields, urban, sub-urban, and mix), as described in Section III-C, and calculated metrics for SwinIR SR result, as described in Section III-D. After those processes, we clustered data into six clusters based on the metrics results using the k-means method. We made two variants, the first was based on all metrics (RMSE, PSNR, SSIM, ERGAS, SCC, SAM, and LPIPS), and the second based on PSNR and SCC. In both variants, metrics scores were standarized. For ablation studies, clustering was performed also on RGB HR images. For this purpose we used first 16 components of singular value decomposition (SVD) and k-means classifier. Steps of the process are illustrated in Fig. 6.

To evaluate clustering results, we used a confusion matrix and parameters as follows.

Purity—the sum of the highest score for each cluster to the total number of samples

$$\frac{1}{N}\sum_{i=1}^{k} \max_j |c_i \cap t_j|. \tag{8}$$

OA—the number of correctly classified samples to the total number of samples.

$$\frac{\sum_{i=1}^{n} \mathrm{TP}_i}{\sum_{i=1}^{k}(\mathrm{TP}_i + \mathrm{TN}_i + \mathrm{FP}_i + \mathrm{FN}_i)}. \tag{9}$$

True positive rate (TPR)—the number of correctly classified samples in one class to the total number of samples in this class.

$$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}. \tag{10}$$

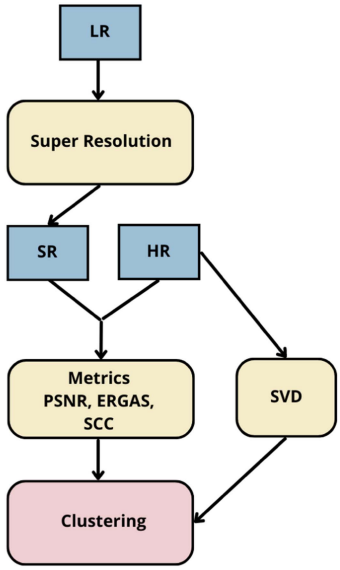

Fig. 6. Images clustering procedure.

TABLE IV
PSNR METRIC SCORE FOR SWINIR FINE-TUNING AND TRAINING FROM SCRATCH, TESTING, AND VALIDATION DATASET

	Iterations	Fine-tuning	From scratch
Test	60000	63.3443	55.9138
Validation	60000	64.2460	57.4382

PPV—the number of correctly classified samples in one class to the total number of samples in one cluster.

$$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}. \tag{11}$$

with k—number of clusters, N—number of samples, c_i—cluster i, t_j class j, TP—true positive, TN—true negative, FP—false positive, FN—false negative.

IV. EXPERIMENTS

In this section, we present the results of the image SR based on selected methods, image reconstruction performance and quality assessment, results of SR for each of the selected classes, and results for unsupervised classification.

A. Results for SR Methods

In order to assess training results for the SwinIR model, we tested different training settings. At first, we used shared weights and without training use it with validation and testing sets to calculate metric scores. Then, based on shared weights, we fine-tune the model up to 90 000 iterations. We also trained from scratch up to 140 000 iterations. We choose weights for the model after fine-tuning up to 60 000 iterations. The results were stable, higher, and less time-consuming compared to training from scratch. The scores of PSNR of training options for the testing and validation dataset are presented in Table IV.

In the case of the SRCNN model, we used training parameters described in Section III-B1. Table V shows general quality

TABLE V
METRICS SCORE FOR VALIDATION AND TESTING SETS FOR SWINIR AND SRCNN WITH BICUBIC INTERPOLATION BASELINE

Method	Set	PSNR	PSNR	ERGAS	SCC
Bicubic	Test	60.8751	0.9988	3.1801	0.5442
	Validation	60.5010	0.9940	3.5294	0.4869
SRCNN	Test	61.8675	0.9990	2.7954	0.5598
	Validation	61.2626	0.9940	3.3553	0.4924
SwinIR	Test	63.3443	0.9993	2.2551	0.6585
	Validation	64.2460	0.9994	1.6911	0.6293

TABLE VI
AVERAGE JS DISTANCE BETWEEN HISTOGRAMS OF HR AND SR FOR SIX DIFFERENT LAND COVER TYPES FOR BICUBIC INTERPOLATION, SRCNN AND SWINIR

Land cover class	Bicubic	SRCNN	SwinIR
FOREST	0.01167	0.01056	0.00900
LARGE FIELDS	0.00980	0.00853	0.00644
SMALL FIELDS	0.01608	0.01413	0.01082
URBAN	0.04708	0.04393	0.03438
SUB-URBAN	0.02080	0.01891	0.01516
MIX	0.01891	0.01706	0.01360

assessment results for model options, we used in the further analysis: SRCNN and SwinIR with fine-tuning up to 60 000 iterations and for baseline, we used bicubic interpolation. Table V presents average metrics scores (PSNR, SSIM, ERGAS, and SCC) for validation and testing sets.

B. *Quality Assessment for Different Land Cover Types*

The visual effects of image SR for selected images for each land cover class and for the three used methods are shown in Fig. 7. From these figures, we can observe that compare to bicubic interpolation, both DL methods give more sharpen elements boundaries. For SwinIR, it is possible to see some thin and small elements such as dirt roads. Also, the texture for example on the fields are better preserved in SwinIR results. Absolute differences between the HR image and SR results for each class and method for large fields are presented in Fig. 8 and the rest of land type are presented on the figures in the Appendix: forest (see Fig. 11), small fields (see Fig. 12), urban (see Fig. 13), sub-urban (see Fig. 14), and mix (see Fig. 15). The results are presented for each channel and with the same scale. The absolute differences decrease with the level of complexity of the neural network, which is not surprising such that the stronger the network, the better the results. The largest differences occur for the edges of objects; therefore, the more differentiated the terrain, the more higher results appear like in "urban" or "sub-urban" examples. Moreover, the higher values of differences are for RED channel in every land cover types.

Histogram comparisons was done for each classes and methods. The results of sample for each classes are shown in figures in the Appendix: forest (see Fig. 16), large fields (see Fig. 17), small fields (see Fig. 18), urban (see Fig. 19), sub-urban (see Fig. 20), and mix area (see Fig. 21). In order to measure the similarity of HR histogram to SR histogram, the JS divergence was used. Histograms for each samples in each land cover class was calculated, and then, the average for each class has been determined. The results are shown on Table VI.

For each land cover class, we calculate metrics: RMSE, PSNR, SSIM, ERGAS, SCC, SAM, and LPIPS. The results of average for each class are presented for bicubic interpolation, SRCNN and SwinIR, respectively, in Tables VII–IX. In each of the methods, the highest results for almost all metrics were obtained by areas that are represented by homogeneous land cover type within one image, i.e., forests and large fields. The lowest values for almost all metrics were recorded in fragments showing urban areas. These regions present higher complexity exhibiting exceptionally diverse and intricate features. These are often the target areas for SR, as enhancing resolution in such regions allows for better differentiation of finer details. The exception is the SCC metric, where the best scores were achieved by small fields. This metric shows the correlation between two images after applying high-pass filters, so it focuses on the areas contained in the high-frequency domain. Small fields as well as buildings and most part of the mix area are images in the high-frequency domain. The SCC metric and its impact and application are described in more detail in Section V.

Correlation charts for metrics were created to visualize and compare the results. The whole pair plots comparing accuracy assessment metrics for bicubic interpolation, SRCNN, and SwinIR method are in Figs. 22–24 of the Appendix. This type of representation of SR metric scores is easy to interpret and understand the relationship between metrics. Some of the pairs of metrics are very well correlated, such as PSNR to RMSE and PSNR to SSIM or ERGAS to SAM. The general tendency is similar for each method. The distinction of land cover classes is most visible for the scatter plot of SCC to PSNR and SCC to ERGAS. Both examples for bicubic, SRCNN, and SwinIR are shown in Fig. 9. For PSNR and SCC, higher score is better, and for ERGAS, the lower score is better. For each land cover type, the characteristics of the metrics are slightly different. For areas more homogeneous within one class, such as forests and large fields, the variability of the results is smaller. Classes representing more differential land cover have a much larger variability. The distribution of classes on the charts for each method looks similar, but the lowest result for SCC differs by about 0.1 and the best by more than 0.15.

Fig. 10 shows plots of the results of Discrete Fourier Transform for bicubic, SRCNN, and SwinIR SR compare to HR images. The plots are made for the samples of each land cover class shown in Fig. 7. Fourier transform converts a signal from the time domain to the frequency domain. In the case of images, the higher the frequency, the more diverse the area are shown in the images. DFT-based graphs allow to visualize the difficulty of the terrain and the amount of detail. It can be said that if the outcome of the SR method closely resembles the HR image, it indicates a similar level of detail is present in the super-resolved image. The frequency distribution is significantly different for each of the land cover examples. "Urban" and "sub-urban" area have higher values in general and especially in high-frequency parts. For bicubic interpolation, the scores starts to be different from HR reference around the frequency 10–15. For SRCNN and SwinIR, this score is between 20 and 30, and for SwinIR, the amplitude is much more similar to HR than SRCNN. The SRCNN DFT results for 15–25 frequency sometimes are

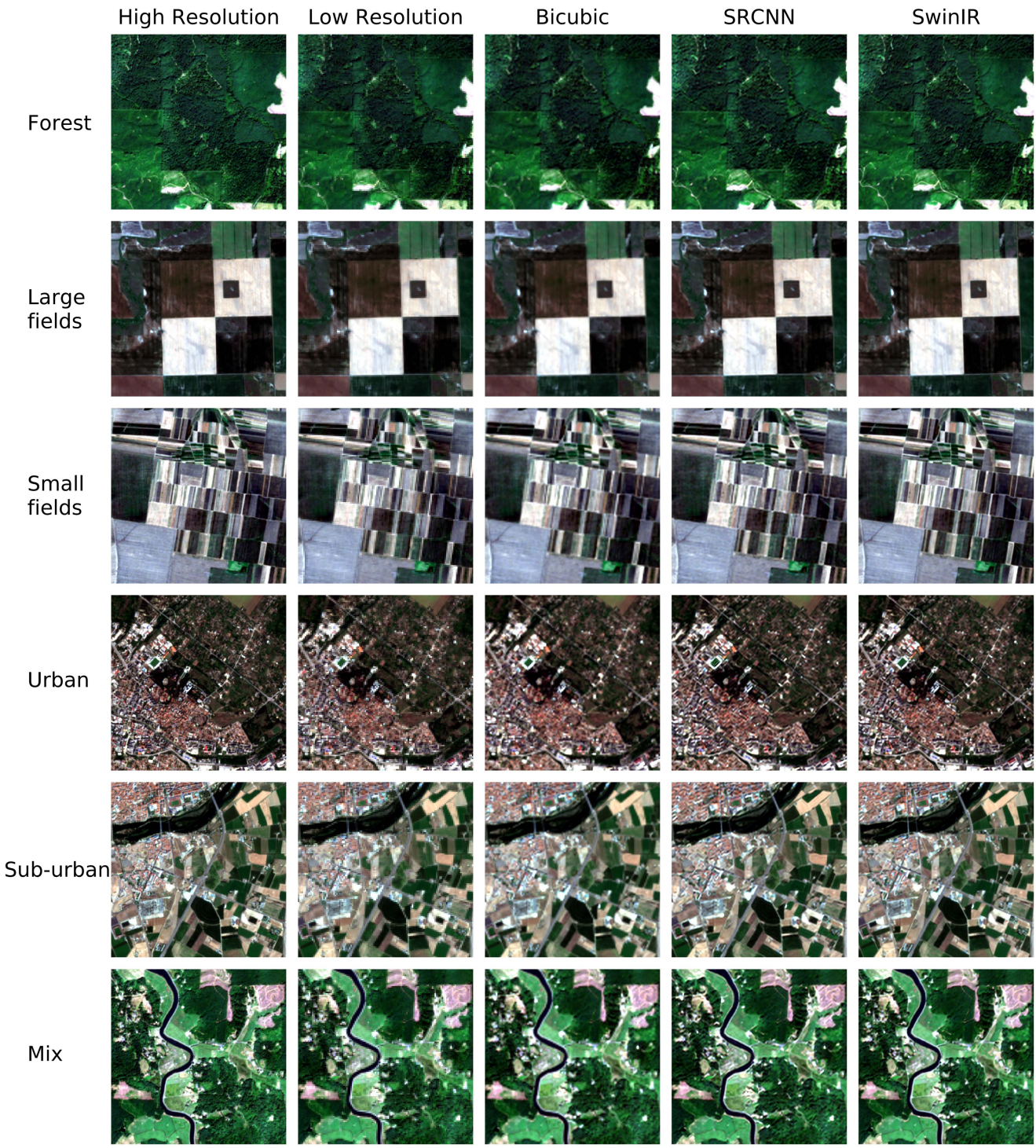

Fig. 7. Visual comparison between results from three used methods.

TABLE VII
AVERAGE METRICS SCORES FOR SIX DIFFERENT LAND COVER TYPES FOR BICUBIC INTERPOLATION

Land cover class	RMSE	PSNR	SSIM	ERGAS	SCC	SAM	LPIPS
FOREST	37.7820	65.0694	0.99962	1.82338	0.49342	0.0341	0.00327
LARGE FIELDS	40.5878	64.4101	0.99957	1.4158	0.53975	0.0270	0.00560
SMALL FIELDS	67.0783	60.1096	0.99884	2.37410	0.57948	0.0442	0.01399
URBAN	136.7438	54.1319	0.99583	8.36839	0.55913	0.1445	0.03643
SUB-URBAN	77.9604	59.1799	0.99830	3.29538	0.55623	0.0604	0.01881
MIX	69.9807	59.8949	0.99873	3.17469	0.55021	0.0570	0.01566

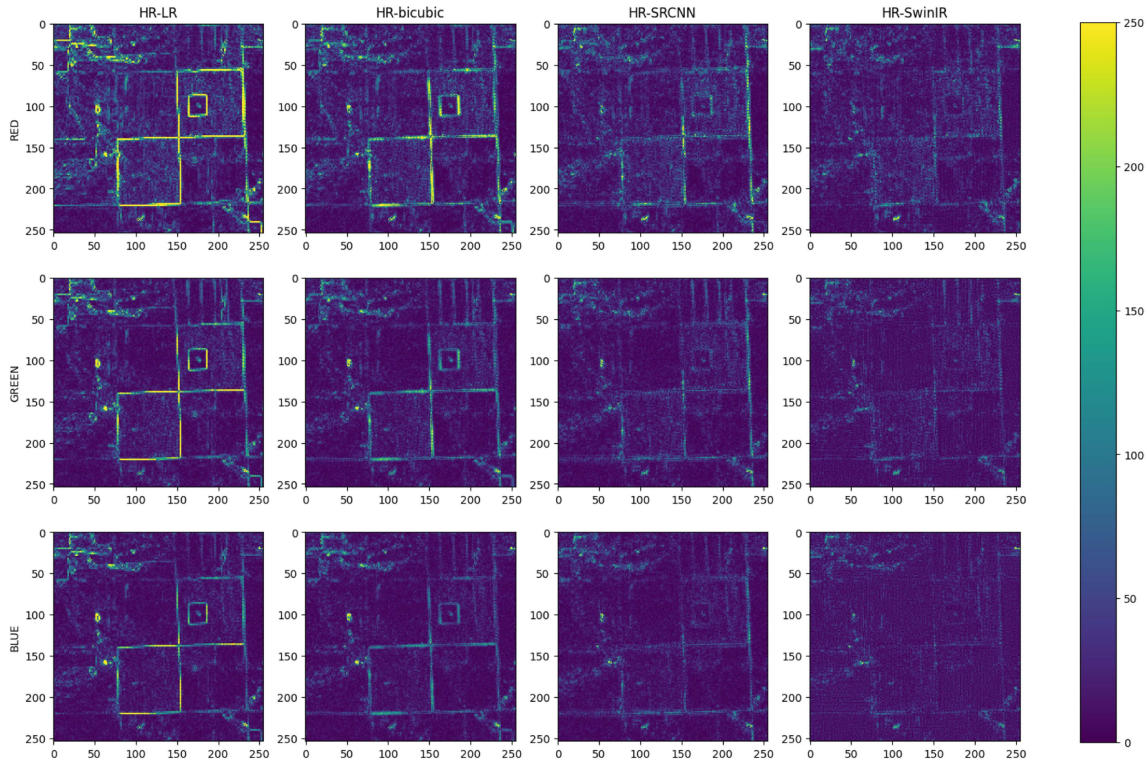

Fig. 8. Absolute difference between ground truth (HR) and super-resolved images for sample of large fields. The image shows absolute distance for each method (bicubic, SRCNN, SwinIR) for R, G, and B channels.

TABLE VIII
AVERAGE METRICS SCORES FOR SIX DIFFERENT LAND COVER TYPES FOR SRCNN METHOD

Land cover class	RMSE	PSNR	SSIM	ERGAS	SCC	SAM	LPIPS
FOREST	33.5448	66.0755	0.99970	1.61115	0.49565	0.0302	0.00037
LARGE FIELDS	35.1183	65.6695	0.99968	1.2335	0.55418	0.0237	0.00094
SMALL FIELDS	59.2355	61.1621	0.99909	2.09399	0.61824	0.0392	0.00208
URBAN	122.0948	55.0412	0.99662	7.44431	0.57739	0.1290	0.01076
SUB-URBAN	70.3356	60.0928	0.99859	2.94409	0.57860	0.0542	0.00433
MIX	62.2796	60.9455	0.99898	2.81460	0.57386	0.0507	0.00279

TABLE IX
AVERAGE METRICS SCORES FOR SIX DIFFERENT LAND COVER TYPES FOR SWINIR METHOD

Land cover class	RMSE	PSNR	SSIM	ERGAS	SCC	SAM	LPIPS
FOREST	29.9384	67.0068	0.99975	1.43976	0.59114	0.0253	0.00014
LARGE FIELDS	27.9537	67.5418	0.99979	1.0066	0.65154	0.0176	0.00019
SMALL FIELDS	47.6023	63.0183	0.99941	1.66742	0.73284	0.0304	0.00086
URBAN	96.7749	57.0150	0.99784	5.87170	0.69524	0.1011	0.00456
SUB-URBAN	57.7096	61.7718	0.99903	2.40368	0.68635	0.0434	0.00189
MIX	51.2757	62.6995	0.99927	2.29433	0.67969	0.0403	0.00133

higher than HR reference, which is not desirable because having a one type of terrain more detailed than in fact it is, is not desire.

C. Result of Clustering

After clustering data samples by all metrics scores, we assigned predicted clusters to actual classes. While for the classes: "forest," "large fields," "small fields," and "urban," it is quite possible to select a class, for classes "sub-urban" and "mix," it is not unambiguous. The clustering results based on all quality assessment metrics are shown on the confusion matrix in Table X. Each row of the matrix represents the instances in a predicted cluster, while each column represents the instances in an actual class. SCC metric is sensitive to the type of land cover, as it has been shown in Fig. 9; therefore, we decided to use it with PSNR as the basis for the clustering. The results are shown on the confusion matrix IN Table XI. In this variant, the above-mentioned clusters could easily be matched to actual classes but for "sub-urban" and "mix" it was slightly more straightforward. The TPR for each classes is similar or higher, especially for

TABLE X
CONFUSION MATRIX FOR CLUSTERING RESULTS BASED ON ALL METRICS

	FOREST	LARGE FIELDS	SMALL FIELDS	URBAN	SUB-URBAN	MIX	sum	PPV
"0"	96	48	1	0	2	1	148	0.6486
"1"	11	53	30	1	47	59	201	0.2637
"2"	0	0	70	6	30	33	139	0.5036
"3"	0	0	0	41	11	4	56	0.7321
"4"	0	0	0	5	3	0	8	0.3750
"5"	8	0	0	12	10	11	41	0.2683
sum	115	101	101	65	103	108	593	
TPR	0.8348	0.5248	0.6931	0.6308	0.0291	0.1019		

Classes are in the columns and clusters are in the rows.

TABLE XI
CONFUSION MATRIX FOR CLUSTERING RESULTS BASED ON PSNR AND SCC

	FOREST	LARGE FIELDS	SMALL FIELDS	URBAN	SUB-URBAN	MIX	sum	PPV
"0"	79	0	0	0	0	1	80	0.9875
"1"	34	69	4	0	5	10	122	0.5656
"2"	0	0	63	1	16	6	86	0.7326
"3"	0	0	1	45	18	11	75	0.6000
"4"	2	0	12	19	35	38	106	0.3302
"5"	0	32	21	0	29	42	124	0.3387
sum	115	101	101	65	103	108	593	
TPR	0.6870	0.6832	0.6238	0.6923	0.3398	0.3889		

Classes are in the columns and clusters are in the rows.

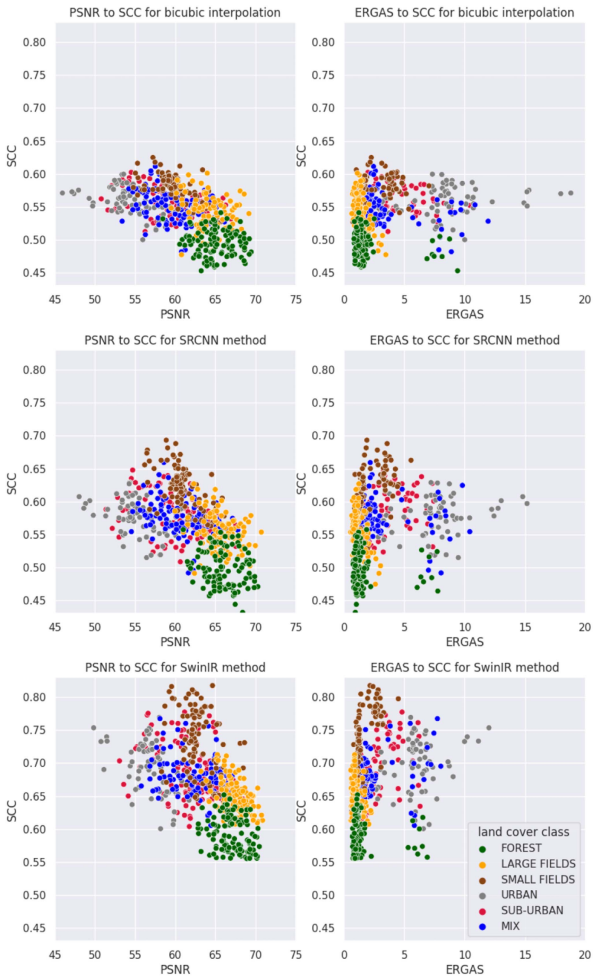

Fig. 9. PSNR and ERGAS to SCC for bicubic interpolation, SRCNN and SwinIR.

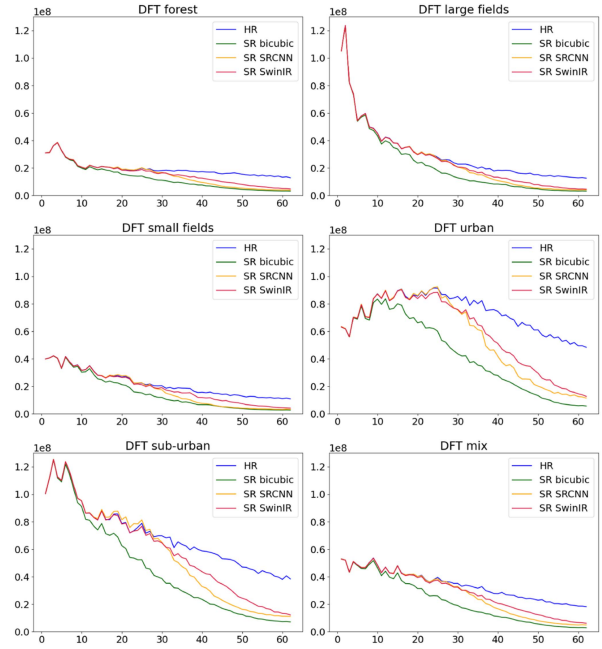

Fig. 10. Discrete Fourier Transform for bicubic, SRCNN and SwinIR SR compare to HR images. Plots presented for samples with each land cover classes. *X*-axis is a frequency, *Y*-axis is an amplitude.

the two worst-performing classes. Only the "forest" perform worse. Moreover, the number of samples in clusters is more consistent with the number within real classes. In the case of clustering based on RGB HR images, the results are presented on the confusion matrix in Table XII. Apart from classes "forest," "large fields," and "small fields," it was hard to assign clusters

TABLE XII
CONFUSION MATRIX FOR CLUSTERING RESULTS BASED RGB HR IMAGES, 16 COMPONENTS OF SVD

	FOREST	LARGE FIELDS	SMALL FIELDS	URBAN	SUB-URBAN	MIX	sum	PPV
"0"	96	10	1	6	18	26	157	0.6115
"1"	10	47	6	7	43	36	149	0.3154
"2"	0	14	55	0	10	11	90	0.6111
"3"	0	17	28	26	9	9	89	0.2921
"4"	9	8	4	25	18	19	83	0.2169
"5"	0	5	7	1	5	7	25	0.2800
sum	115	101	101	65	103	108	593	
TPR	0.8350	0.4653	0.5446	0.4000	0.1748	0.0648		

Classes are in the columns and clusters are in the rows.

TABLE XIII
PURITY AND OA FOR TWO VARIANTS OF CLUSTERING: BASED ON ALL METRICS AND PSNR WITH SCC AND THE ABLATION STUDY BASED ON RGB HR IMAGE, 16 COMPONENTS OF SVD

Purity	HR RGB	All metrics	PSNR SCC
"0"	0.61146	0.64865	0.98750
"1"	0.31544	0.29353	0.56557
"2"	0.61111	0.50360	0.73256
"3"	0.31461	0.73214	0.60000
"4"	0.30120	0.62500	0.35849
"5"	0.28000	0.29268	0.33871
OA	0.41989	0.46206	0.56155

representation. Table XIII shows purity for clustering for both variants based on SR results representing by metrics and for reference clustering on RGB HR images based on SVD components. Table XIII presents also OA score for each variant and reference clustering. Clustering scores are higher for variants based on SR metrics than on SVD components. Moreover, the variant based on PSRN and SCC performs better than the option based on all SR metrics.

V. DISCUSSION

In the conducted research, the application of SRDL methods was explored for the enhancement of satellite images, particularly Sentinel-2. The underlying motivation of the experiment was to rigorously analyze the outcomes attained using SR methods across diverse land cover types and critically assess the uniformity of these results over various SR techniques.

This investigation prominently centered around the R, G, and B channels, enhancing the resolution from 20 to 10 m/pix. This strategic choice was made in light of the availability of these channels in both 10 and 20 m resolutions, thereby allowing the harnessing of a substantial dataset for rigorous analysis. Furthermore, the utilization of data from the same satellite eliminated potential discrepancies related to acquisition times, sensor specifications, and camera calibration, ensuring the reliability and consistency of the results obtained. This careful approach assured that the discerned differences in results across diverse land cover types were unequivocally attributed to terrain specifics, eliminating the influence of external factors.

The focal point of the experiment was not merely the enhancement of the Sentinel-2 images' resolution. Rather, the research concentrated predominantly on delineating a robust methodology for assessing image quality post-SR. This meticulous approach allowed for a comprehensive insight into the utility of the results in the broader context of developing or fine-tuning neural networks for specialized requirements.

A. Discussion on the SR Results

The general results for the validation and testing datasets, as displayed in Table V, exhibit enhanced performance with more complex network architectures. The PSNR scores for bicubic, SRCNN, and SwinIR for the test dataset are sequentially 60.8751, 61.8816, and 63.3443, with higher values indicating superior performance. Concurrently, the ERGAS scores register at 3.1801, 2.7885, and 2.2551, where lower scores denote superior results. This trend in metric scores is evidently dependent on the complexity level of the DL methods.

Compared to bicubic interpolation, advanced methods such as SwinIR offer significant enhancements, especially in high-frequency regions. A detailed visual analysis reveals that SwinIR efficiently identifies diminutive elements, such as field paths or building boundaries in urban landscapes, surpassing other methods in performance.

Inspection of SR results across diverse land cover classes showcases a consistent trend: more homogeneous areas yield superior scores, regardless of the neural network model utilized. For instance, PSNR scores for the SwinIR method range from 57.0150 in urban areas to 67.0068 in forest regions. A corresponding pattern is observed for the ERGAS metric, emphasizing the significance of considering these variations for terrain-specific tasks.

Examining histogram comparisons and JS distance reveals a distinct ranking for land cover types, listed in decreasing order: large fields, forest, small fields, mix, sub-urban, and urban. Most discrepancies between post-SR and original HR images are located at object edges, as evident in absolute differences representations (see Figs. 11–15). Despite these observations, the SwinIR metric results consistently outperform bicubic interpolation and SRCNN, particularly in forest regions.

Interestingly, small-field areas, despite their variability, attain high scores, likely attributed to their organized patterns conducive for model training.

In contrast, the SCC metric for diverse land cover classes presents a unique scenario. Average SCC scores remain

consistent with other metrics, with the most complex method, SwinIR, achieving the highest score. Despite its effectiveness in comparing entire populations and distinguishing land cover types, SCC's sensitivity to terrain variance renders it less objective compared to other metrics. It should thus be employed cautiously, ensuring alignment with the specific aspect under evaluation. Preliminary tests indicate an overall clustering accuracy of 0.5616 based on SCC and PSNR, reinforcing the necessity of thoughtful metric selection for varied analyses and evaluations.

B. Discussion on Clustering Methods

The examination of the distribution of SR results concluded that, to a certain extent, metrics scores align with land cover types. It was observed that the combination of the PSNR and SCC metrics is adept at distinguishing various land cover types based on their corresponding values. Guided by this observation, an inquiry was made to determine the aptness of these metrics for unsupervised classification, aiming to initially discern the prevailing types of land cover in a dataset. This strategy proves invaluable when encountering datasets with unknown land cover types, enabling their characterization.

An ablation study was conducted, comparing the clustering results on the chosen metrics with the results of clustering HR images based on SVD components. Both purity and overall accuracy metrics reached their zenith when classifications were based on PSNR and SCC metrics. This affirms the effective reflection of land cover types by these metrics, demonstrating the potential of the presented approach especially when there is an imperative to characterize datasets for which SR is conducted with respect to land cover types [68], [69], [70].

The ensuing results from the abovementioned clustering process possess the potential for enhancement through the employment of kernel methods. These methods, by transitioning nonlinearly separable data into a dimensionally higher space, possibly refine the definition and thereby the assignment of predicted clusters to actual classes. The employment of Kernel k-means [71], Kernel spectral clustering [72], or Kernel PCA [73] could further augment the clustering process, fostering increased accuracy and robustness.

C. Discussion on Spectral Variability

The influence of spectral variability on the performance of our SR method was may be an important factor. Spectral variability pertains to the variations in spectral data due to factors, such as atmospheric conditions, sensor noise, and intrinsic variations in the earth's surface. While our primary focus in this study was the enhancement of spatial resolution, it is important to consider the potential impact of spectral variability on our SR results.

While our SR model is designed to be robust in the face of spatial variability across different land cover types, it assumes that the spectral characteristics of these types remain relatively constant. Any substantial variability in these characteristics could potentially influence the overall performance of our SR method.

Notably, the integration of concepts to address spectral variability, as seen in the study [74], could potentially bolster our SR framework's robustness to such variabilities. However, such an integration would entail a shift in our problem formulation and is considered a promising direction for future work.

In light of this, we acknowledge that a thorough exploration of the impact of spectral variability on our SR results is an important avenue for future research. Understanding and accounting for spectral variability could bring about more accurate and robust SR results, contributing significantly to the field of remote sensing image analysis.

VI. Conclusion

This article presented a robust methodology for evaluating the quality of super-resolved satellite images across diverse land cover types. Utilizing SRCNN and SwinIR methods for SR, our comprehensive quality assessment approach encompasses multiple metrics, histogram comparison, visual interpretation, and distinct Fourier transform analysis. This research's novelty lies in discerning the differential impact of SR across varied land cover types, such as forest, large and small fields, urban, sub-urban, and mixed terrains. Uniform terrains, such as forests or large fields, consistently exhibited superior SR results across all metrics. Interestingly, the SCC metric showcased unique behavior, performing better with increased terrain variation, offering potential utility in land cover type clustering.

The insights gleaned from this study advocate for the tailored application of SR neural networks to enhance results for specific land cover types. For instance, networks trained on abundant examples of specific land covers, such as small fields in a particular country, may yield improved outcomes for tasks, such as semantic segmentation, field border detection, or precise plant species identification, even though their applicability might be limited to similar terrains.

Our future work aims to extend this research by conducting experiments with high- and LR images from diverse sensors, thereby validating our proposed method's generalizability and robustness across various sensing technologies. An ambitious goal includes augmenting the resolution of Sentinel-2 images from 10 m/pixel to a staggering 2.5 m/pixel, utilizing PlanetScope data as the HR benchmark. Our exploration will also encompass the inclusion of infrared channels to further enrich land cover classification and potentially enhance SR performance.

Further endeavors include fine-tuning networks for specific land covers, such as urban or small field images. This targeted approach will investigate potential improvements in metric scores for these specific types, possibly leading to heightened SR efficacy for such categories. The experiments conducted suggest the potential for enhanced SR applications, such as border detection for small fields, by fine-tuning networks using selected land cover type datasets. In future work, these hypotheses will be rigorously tested and validated, and the adaptation of neural network architectures for specific land cover types will be explored to further optimize SR outcomes for diverse terrains.

APPENDIX

TABLE XIV
LIST OF SENTINEL-2 SCENES USED IN THE TRAINING, VALIDATION, AND TESTING PROCESS WITH REGIONS AND THEIR CHARACTERISTICS

Sentinel-2 tile	Region	Characteristic of the region
T36MXD_20210312T074721	Tanzania, Kenya	vegetation, cities, mine, forests, water
T32UPV_20210814T102031	Germany	crops, cities, roads, few forests
T35TMJ_20210823T090601	Rumunia, Bułgaria	crops, cities, orange soil
T38SLE_20210829T074611	Iraq	deserts, cities, farmland, river
T43QEC_20220404T052651	India	crops, small fields, cities, orange soil
T55HDE_20220405T002101	Australia	crops and nature
T43PHR_20220428T050701	India	crops, nature, small fields, cities
T36TWL_20220521T083611	Türkiye	crops, nature, cities
T13SBD_20220608T175921	USA: Colorado	nature, mountains, villages
T15TXL_20220620T165901	USA: Wisconsin	nature, forests, fields, towns
T15UXP_20220620T165901	USA, Canada	nature, forests, water
T13REN_20220719T172911	Mexico	nature, desert, cities, river, round fields
T10TDN_20220726T185931	USA: Oregon	nature, crops, smaller towns
T36UWD_20220812T084611	Russia, Ukraine	forests, crops, towns
T38UQB_20220814T074621	Russia, Kazakhstan	crops, towns
T35UPS_20220818T090611	Ukraine	forests, crops, towns
T10UFG_20220818T190931	Canada	forests, crops, towns
T37UDB_20210511T083559	Russia	cities, crops, forests
T29TQF_20210824T110619	Spain	crops, towns, dry land
T36RUT_20210826T082559	Egypt	deserts
T36TVK_20210829T083559	Türkiye	desterts, cities
T48RVQ_20220307T033609	China	very diverse, cities, crops, forests
T20KKB_20220505T142729	Bolivia	mountains, dry land
T34LCJ_20220516T083559	Angola	forests, deserts
T35LNF_20220517T080609	Zambia	crops
T36JTP_20220518T073609	RSA	crops, vegetation
T35WMQ_20220629T100029	Finland	forests, water
T13TFM_20220809T174909	USA: North Dakota	crops
T39UWV_20220816T073619	Russia	large crops
T38VNJ_20220818T081609	Russia	forests

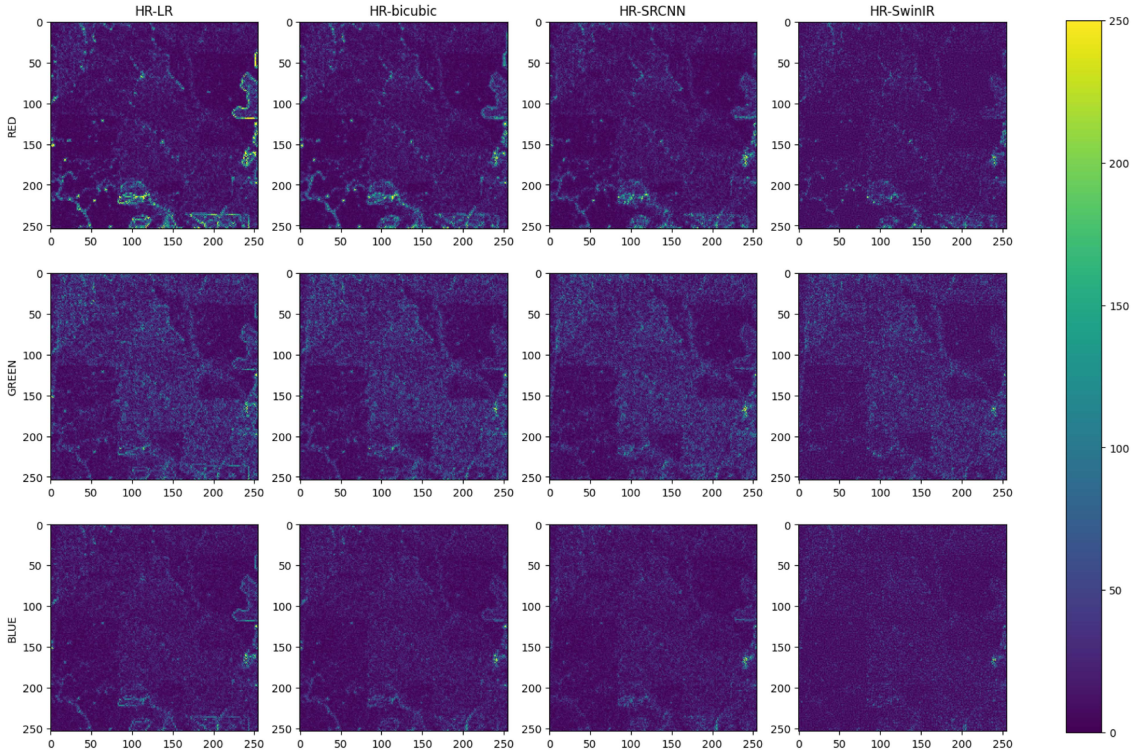

Fig. 11. Absolute difference between ground truth (HR) and super-resolved images for sample of forest. The image shows absolute distance for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels.

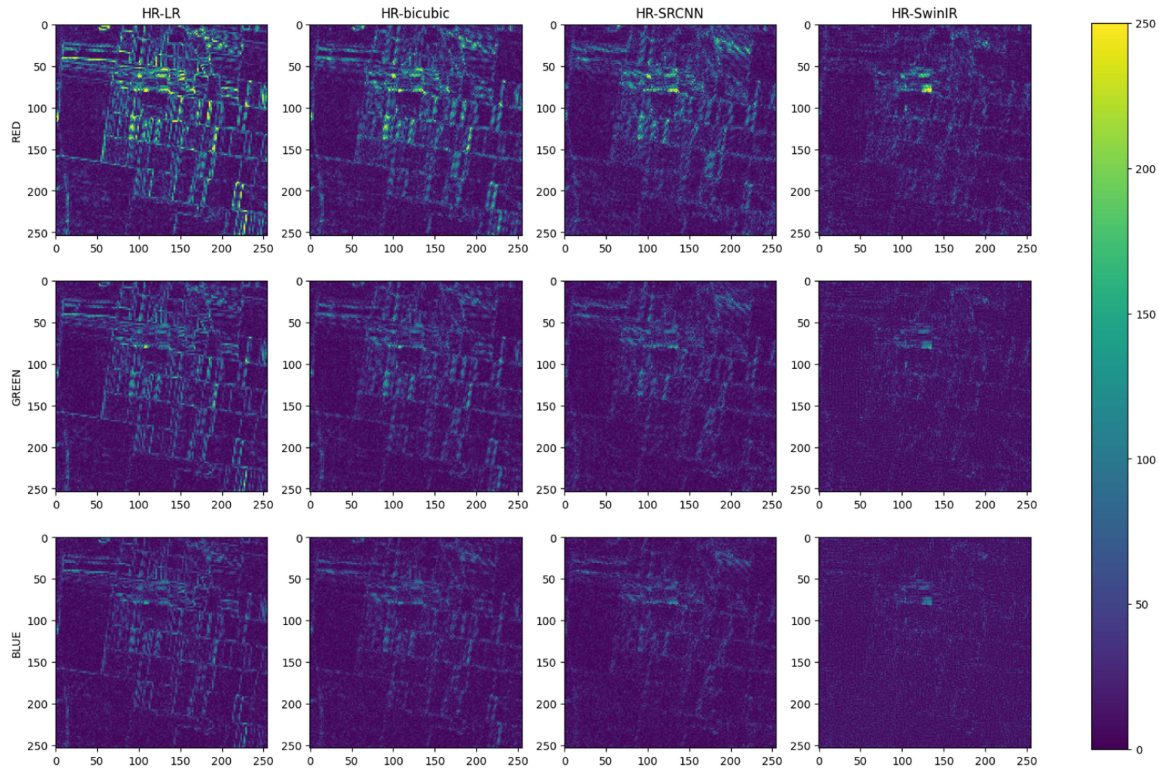

Fig. 12. Absolute difference between ground truth (HR) and super-resolved images for sample of small fields. The image shows absolute distance for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels.

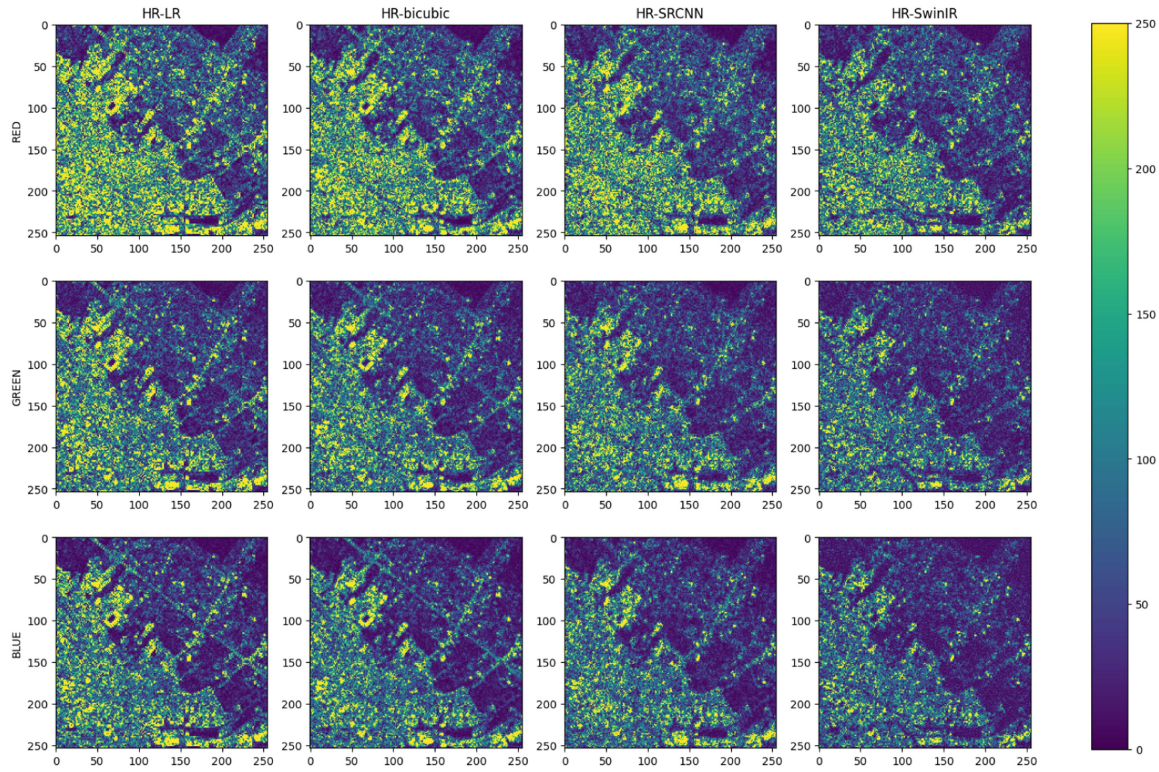

Fig. 13. Absolute difference between ground truth (HR) and super-resolved images for sample of urban area. The image shows absolute distance for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels.

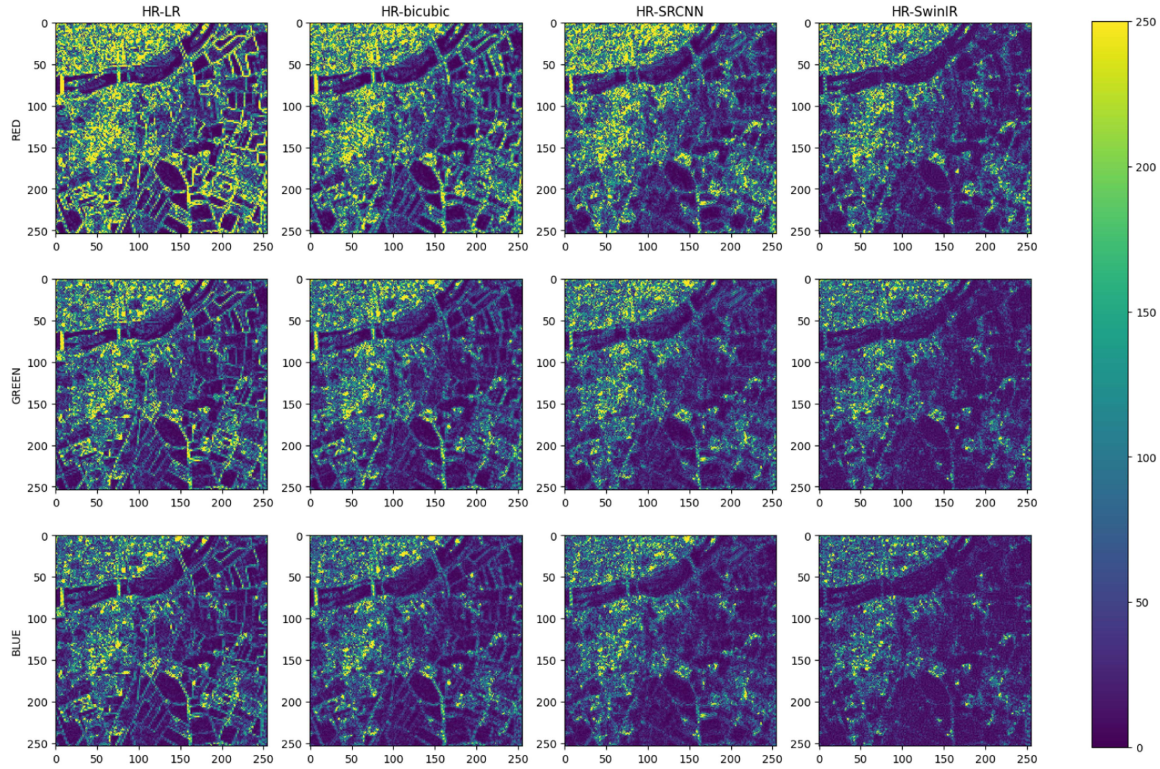

Fig. 14. Absolute difference between ground truth (HR) and super-resolved images for sample of sub-urban area. The image shows absolute distance for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels.

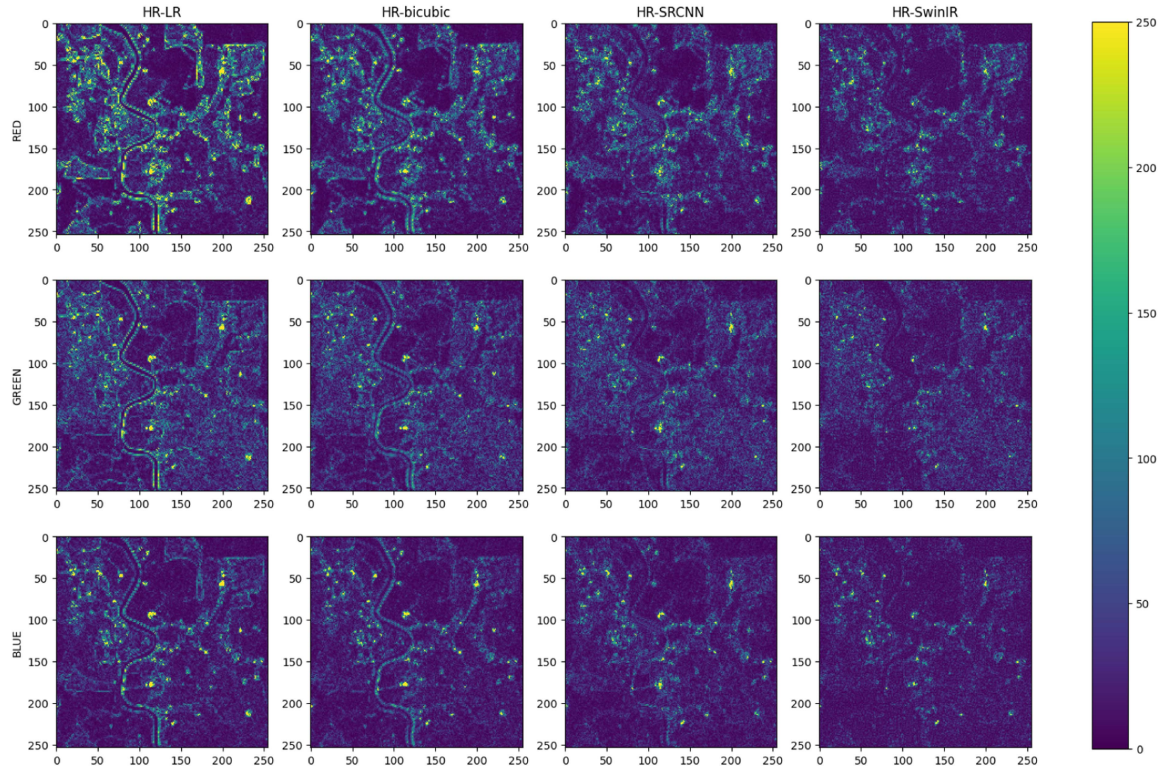

Fig. 15. Absolute difference between ground truth (HR) and super-resolved images for sample of mix area. The image shows absolute distance for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels.

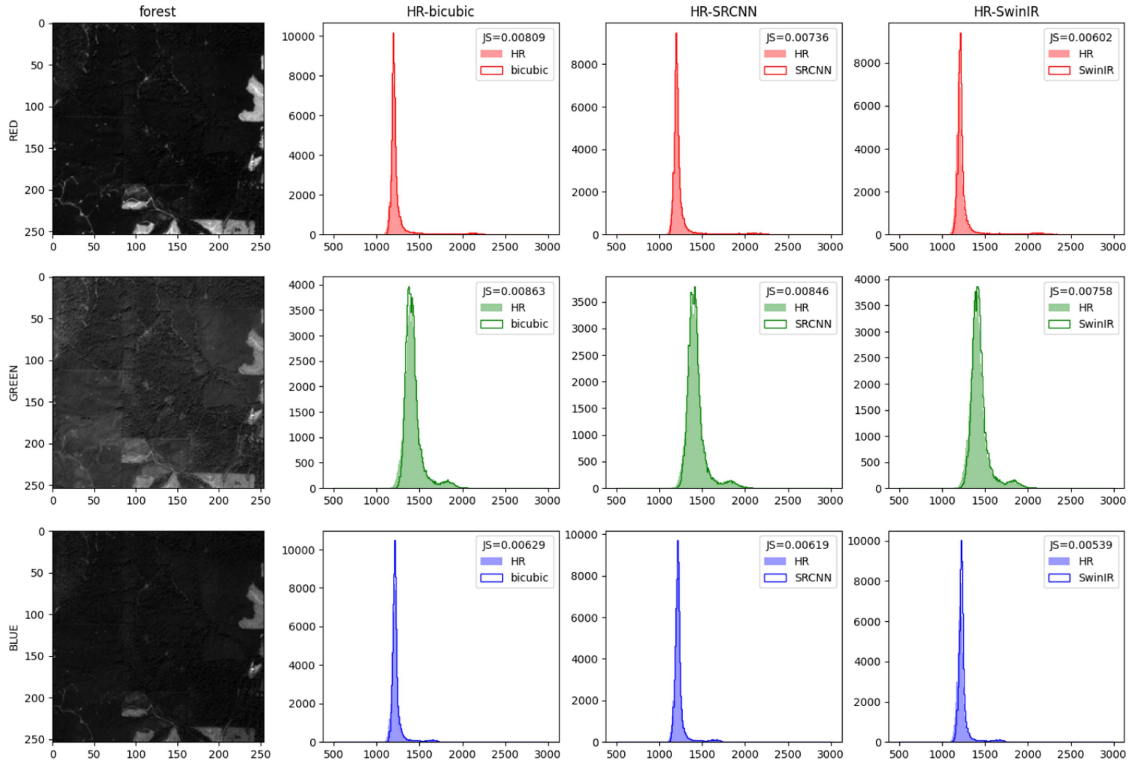

Fig. 16. Comparisons of histograms between ground truth (HR) and super-resolved images for sample of forest. The image shows histograms for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels and JS divergence score in the legend.

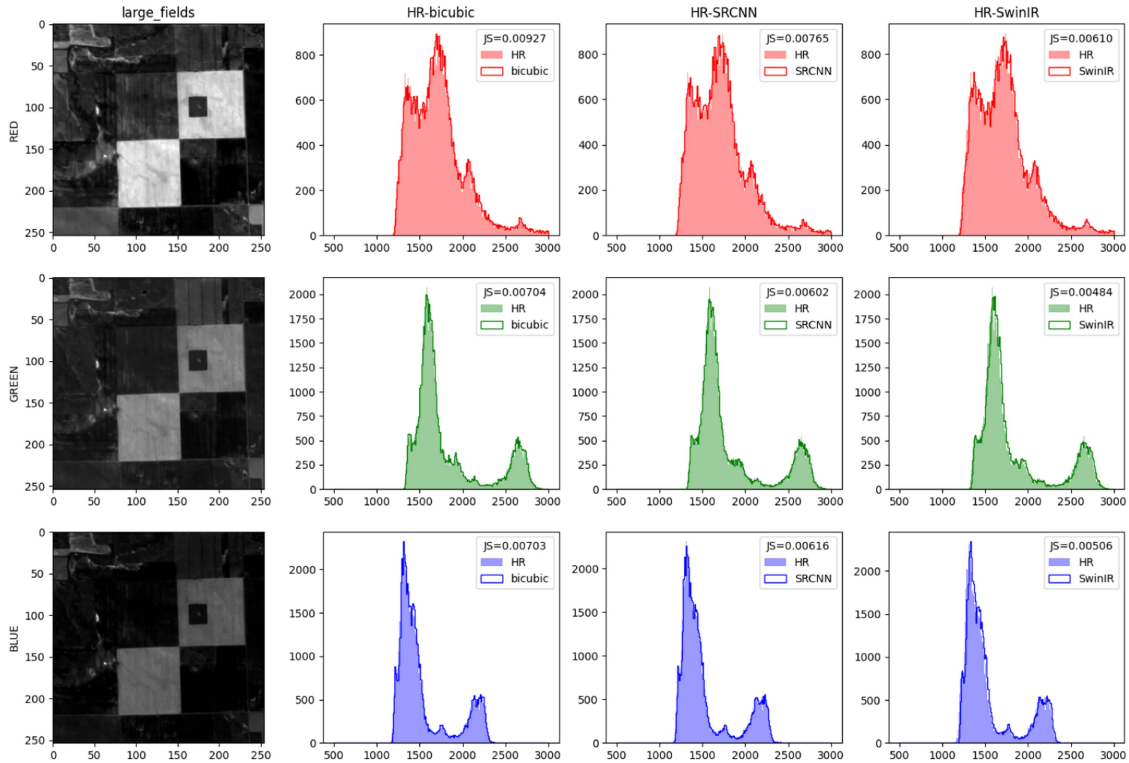

Fig. 17. Comparisons of histograms between ground truth (HR) and super-resolved images for sample of large fields. The image shows histograms for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels and JS divergence score in the legend.

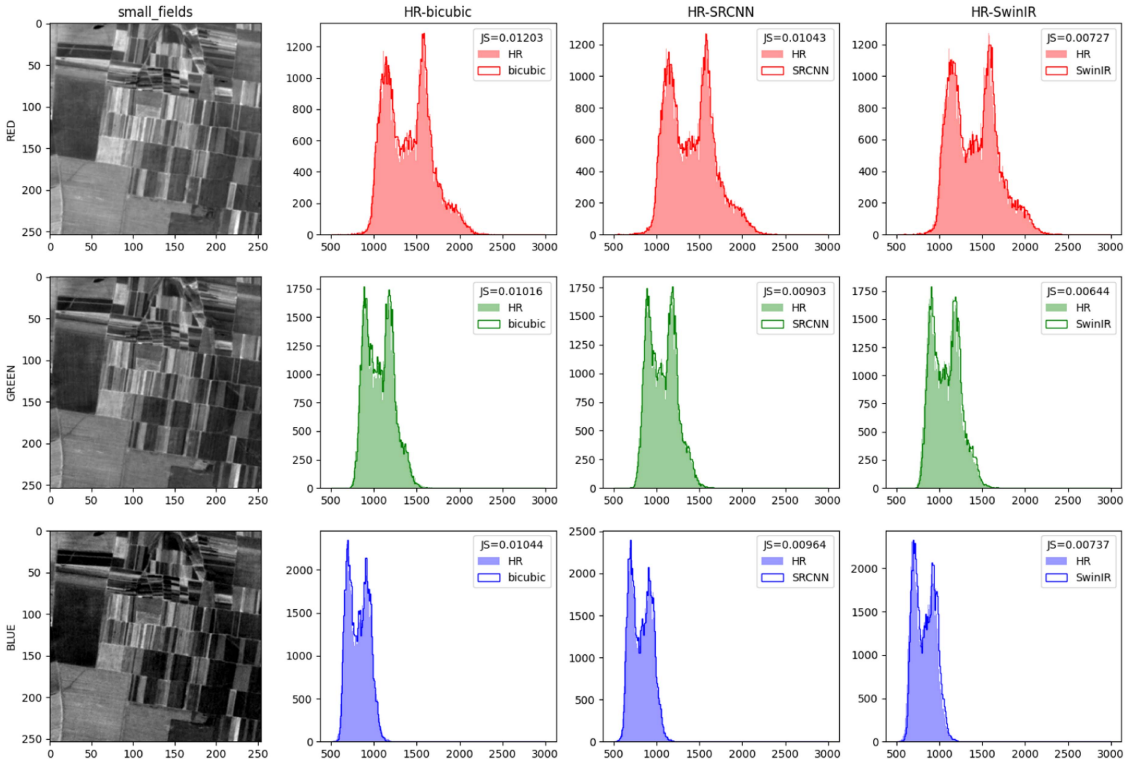

Fig. 18. Comparisons of histograms between ground truth (HR) and super-resolved images for sample of small fields. The image shows histograms for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels and JS divergence score in the legend.

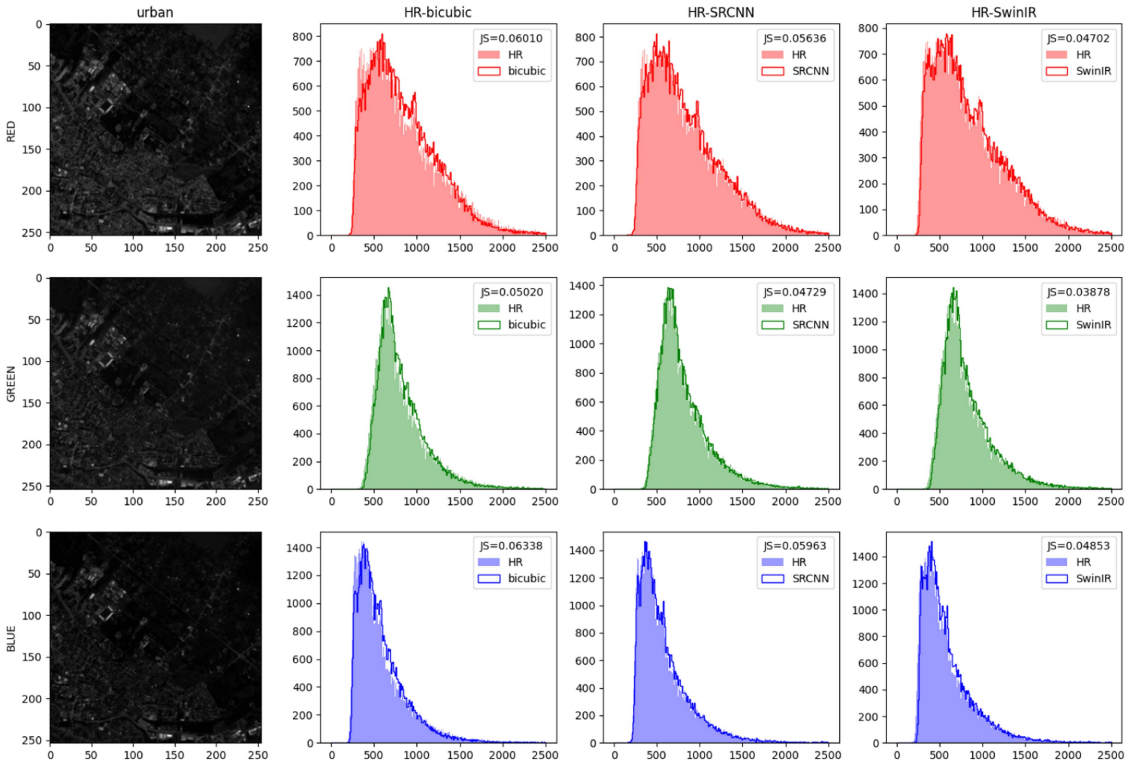

Fig. 19. Comparisons of histograms between ground truth (HR) and super-resolved images for sample of urban area. The image shows histograms for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels and JS divergence score in the legend.

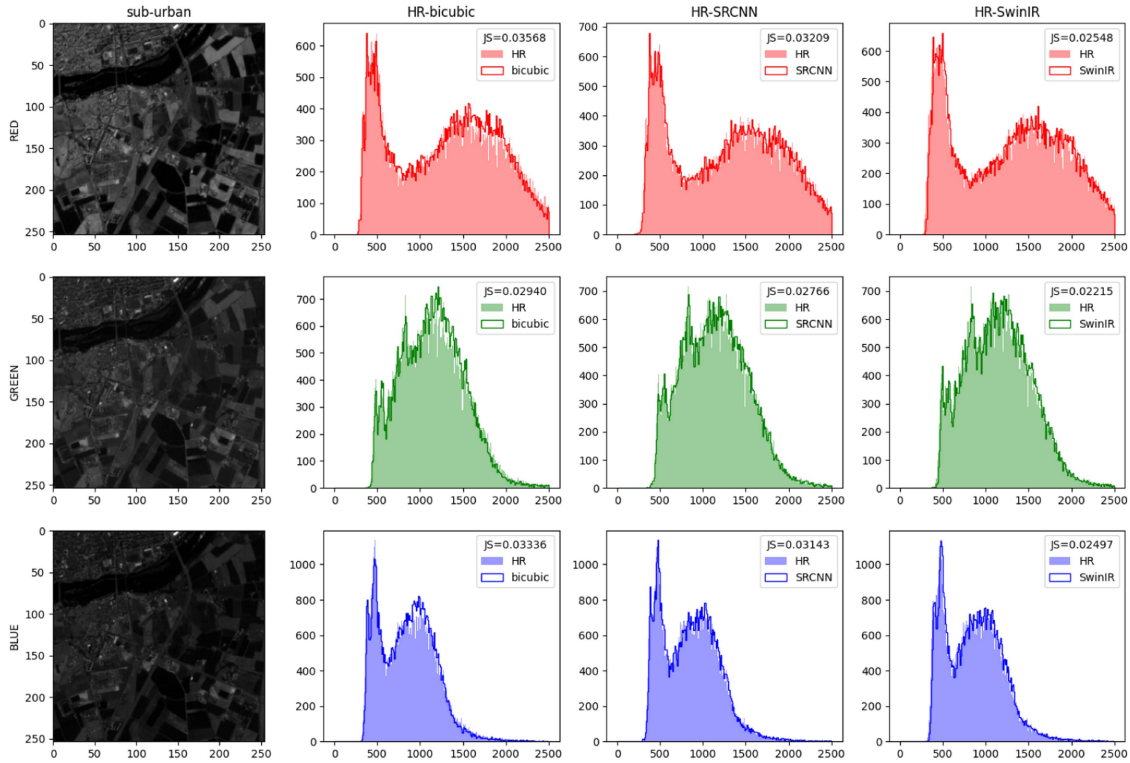

Fig. 20. Comparisons of histograms between ground truth (HR) and super-resolved images for sample of sub-urban area. The image shows histograms for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels and JS divergence score in the legend.

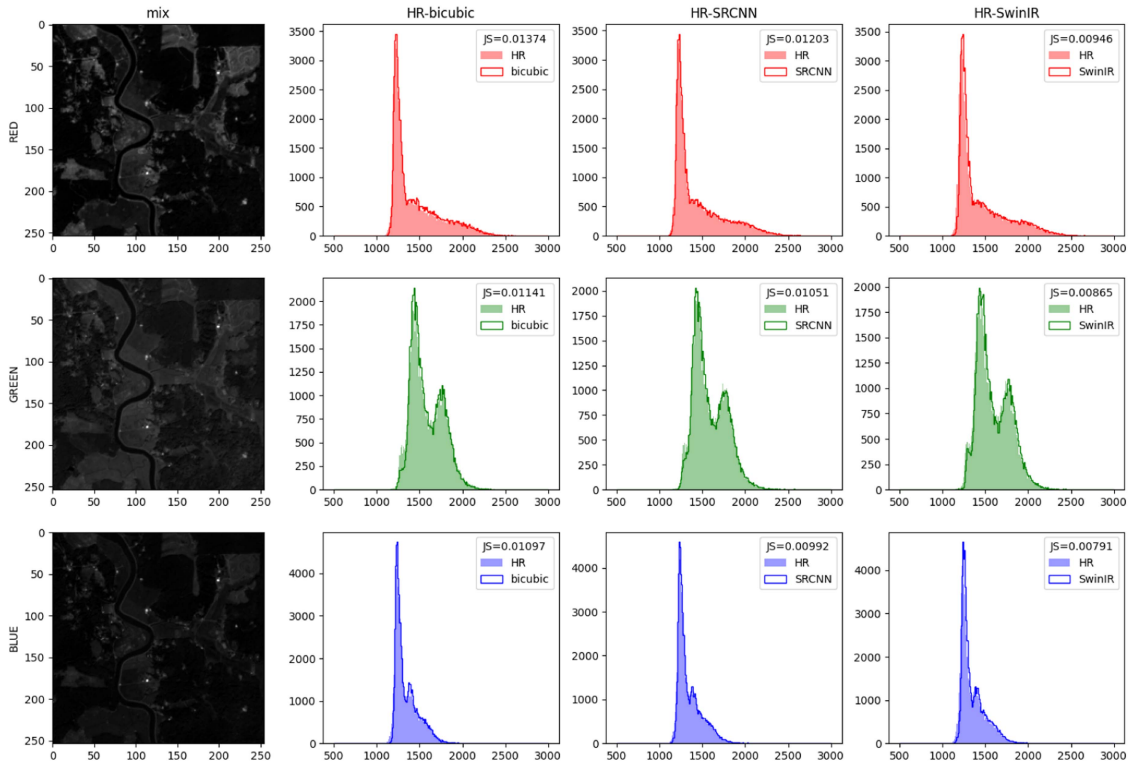

Fig. 21. Comparisons of histograms between ground truth (HR) and super-resolved images for sample of mix area. The image shows histograms for each method (bicubic, SRCNN, and SwinIR) for R, G, and B channels and JS divergence score in the legend.

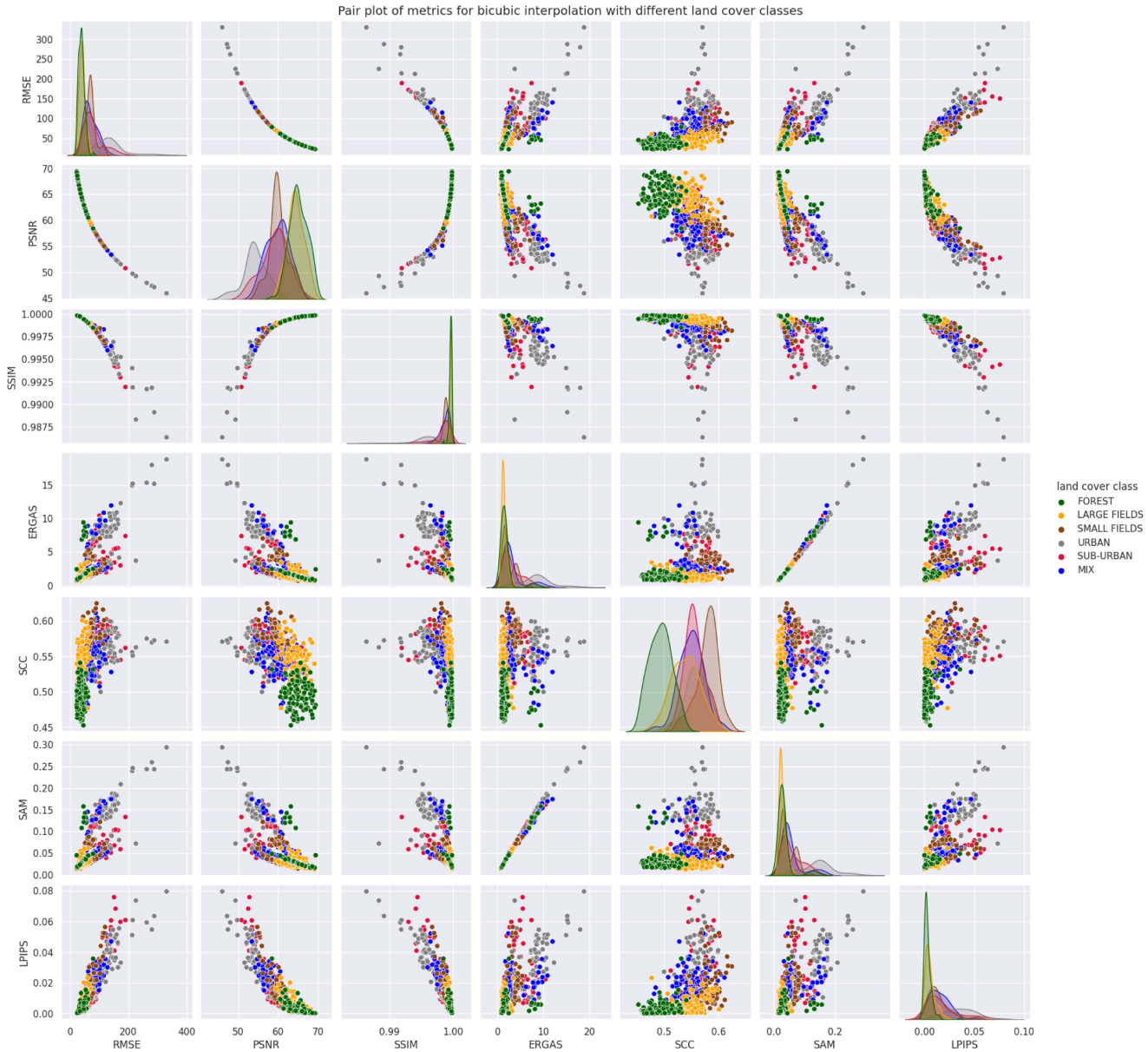

Fig. 22. Pair plot for metrics for bicubic interpolation.

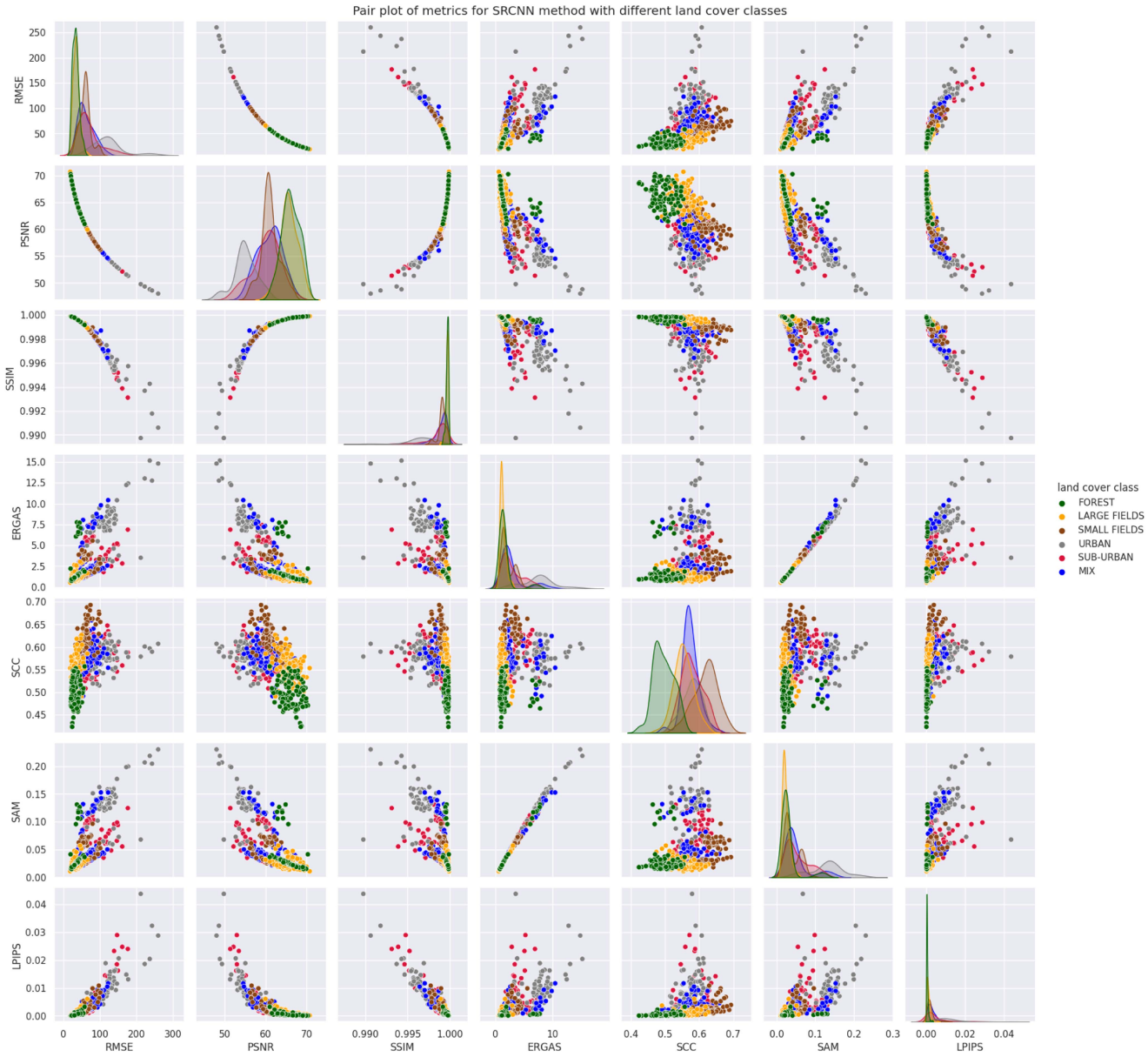

Fig. 23. Pair plot for metrics for SRCNN.

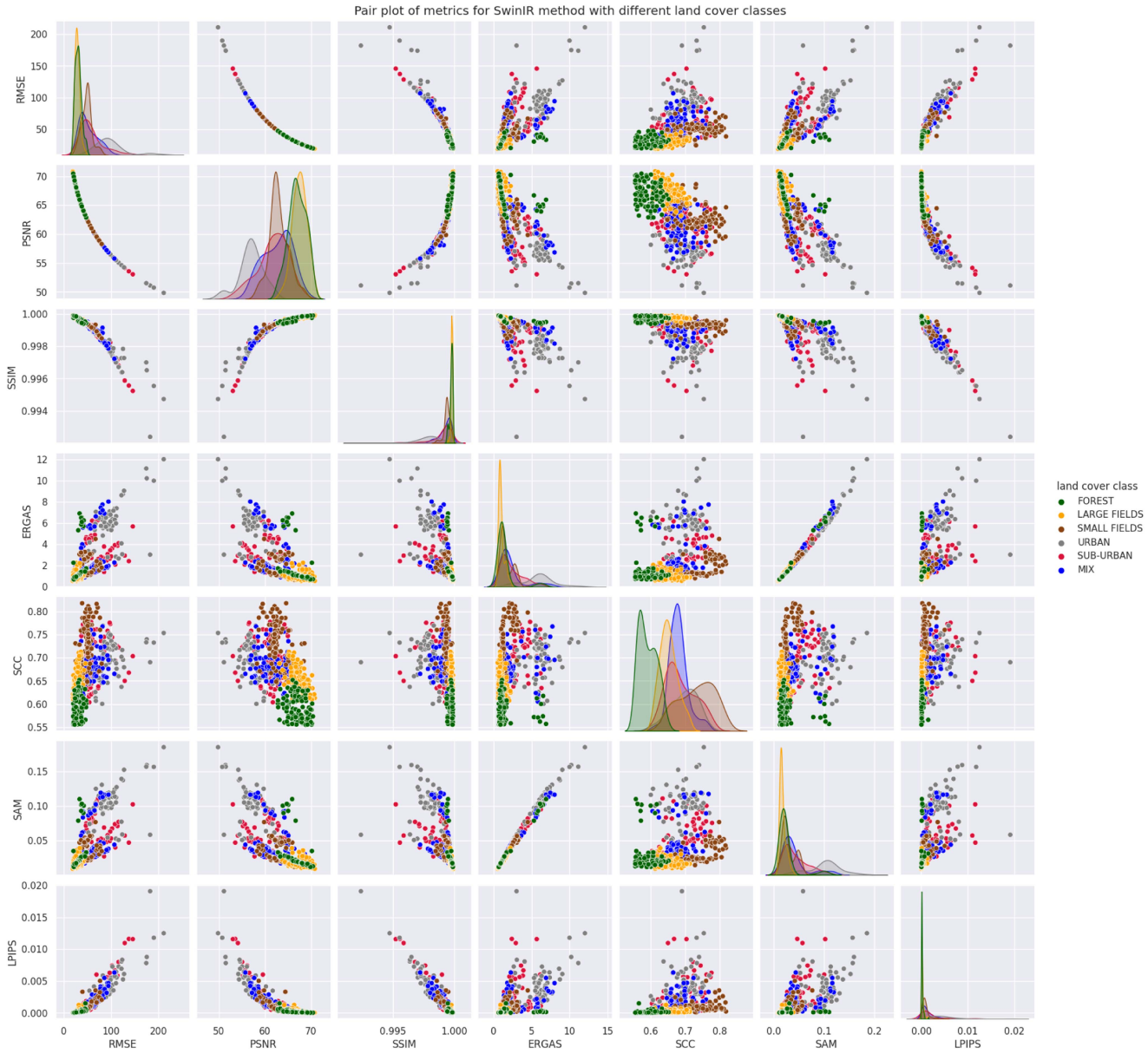

Fig. 24. Pair plot for metrics for SwinIR.

ACKNOWLEDGMENT

This work was carried out as a part of the statutory tasks of Department of Electronics of AGH University of Krakow (Poland), within the subsidy of Polish Ministry of Science and Higher Education. This work was supported by the AGH University of Krakow through "Excellence Initiative—Research University" program.

REFERENCES

[1] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3365–3387, Oct. 2021.

[2] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

[3] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2021, pp. 1833–1844.

[4] L. Liebel and M. Körner, "Single-image super resolution for multispectral remote sensing data using convolutional neural networks," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLI-B3, pp. 883–890, 2016.

[5] A. Malczewska, "SRCNN and SWINIR modification for satellite images e.g. Senitnel-2, clustering of images," 2023. [Online]. Available: https://github.com/anmaran/Super-resolution-Sentinel-2,

[6] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[7] Q. Yuan et al., "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716.

[8] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Front.*, vol. 7, no. 1, pp. 3–10, 2016.

[9] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[10] Y. Bai, X. Sun, Y. Ji, J. Huang, W. Fu, and H. Shi, "Bibliometric and visualized analysis of deep learning in remote sensing," *Int. J. Remote Sens.*, vol. 43, no. 15/16, pp. 5534–5571, 2022.

[11] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018.

[12] I. Kotaridis and M. Lazaridou, "Remote sensing image segmentation advances: A meta-analysis," *ISPRS J. Photogrammetry Remote Sens.*, vol. 173, pp. 309–322, Mar. 2021.

[13] M. Schmitt, J. Prexl, P. Ebel, L. Liebel, and X. X. Zhu, "Weakly supervised semantic segmentation of satellite images for land cover mapping–challenges and opportunities," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. V–3–2020, no. 3, pp. 795–802, Aug. 2020.

[14] S. Talukdar et al., "Land-use land-cover classification by machine learning classifiers for satellite observations–a review," *Remote Sens.*, vol. 12, no. 7, Apr. 2020, Art. no. 1135.

[15] D. Hong et al., "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, Aug. 2021.

[16] B. Yu, L. Yang, and F. Chen, "Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3252–3261, Sep. 2018.

[17] R. Youssef, M. Aniss, and C. Jamal, "Machine learning and deep learning in remote sensing and urban application: A systematic review and meta-analysis," in *Proc. 4th Ed. Int. Conf. Geo-IT Water Resour., Geo-IT Water Resour.*, 2020, pp. 1–5.

[18] C. Qiu, M. Schmitt, C. Geiß, T. H. K. Chen, and X. X. Zhu, "A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 163, pp. 152–170, May 2020.

[19] L. Graf, H. Bach, and D. Tiede, "Semantic segmentation of Sentinel-2 imagery for mapping irrigation center pivots," *Remote Sens.*, vol. 12, no. 23, pp. 1–19, Dec. 2020.

[20] J. Pena, Y. Tan, W. Boonpook, J. Pena, Y. Tan, and W. Boonpook, "Semantic segmentation based remote sensing data fusion on crops detection," *J. Comput. Commun.*, vol. 7, no. 7, pp. 53–64, Jul. 2019.

[21] B. Hejmanowska, P. Kramarczyk, E. Głowienka, and S. Mikrut, "Reliable crops classification using limited number of Sentinel-2 and Sentinel-1 images," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3176.

[22] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[23] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "An investigation on self-normalized deep neural networks for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 3607–3610.

[24] Y. Chen, X. Gui, J. Zeng, X.-L. Zhao, and W. He, "Combining low-rank and deep plug-and-play priors for snapshot compressive imaging," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2023.3294262.

[25] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 148–164, 2018.

[26] Y. Zhao, J. Zhao, C. Zhao, W. Xiong, Q. Li, and J. Yang, "Robust real-time object detection based on deep learning for very high resolution remote sensing images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1314–1317.

[27] W. Diao, X. Sun, X. Zheng, F. Dou, H. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 137–141, Feb. 2016.

[28] L. Khelifi and M. Mignotte, "Deep learning for change detection in remote sensing images: Comprehensive review and meta-analysis," *IEEE Access*, vol. 8, pp. 126385–126400, 2020.

[29] W. Zhang and H. Fan, "Application of isolated forest algorithm in deep learning change detection of high resolution remote sensing image," in *Proc. IEEE Int. Conf. Artif. Intell. Comput. Appl.*, 2020, pp. 753–756.

[30] A. Shafique, G. Cao, Z. Khan, M. Asad, and M. Aslam, "Deep learning-based change detection in remote sensing images: A review," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 871.

[31] X. Wang et al., "A review of image super-resolution approaches based on deep learning and applications in remote sensing," *Remote Sens.*, vol. 14, no. 21, Oct. 2022, Art. no. 5423.

[32] P. Wang, B. Bayram, and E. Sertel, "A comprehensive review on deep learning based remote sensing image super-resolution methods," *Earth-Sci. Rev.*, vol. 232, Sep. 2022, Art. no. 104110.

[33] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.

[34] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[35] S. Vitale and G. Scarpa, "A detail-preserving cross-scale learning strategy for CNN-based pansharpening," *Remote Sens.*, vol. 12, no. 3, Feb. 2020.

[36] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "Pannet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1753–1761.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Dec. 2015, pp. 770–778.

[38] M. Gargiulo, A. Mazza, R. Gaetano, G. Ruello, and G. Scarpa, "A CNN-based fusion method for super-resolution of Sentinel-2 data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4713–4716.

[39] M. Gargiulo, "Fast super-resolution of 20 m Sentinel-2 bands using convolutional neural networks," *Remote Sens.*, vol. 11, no. 22, Nov. 2019, Art. no. 2635.

[40] M. Gargiulo, "Advances on cnn-based super-resolution of Sentinel-2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3165–3168.

[41] C. Lanaras, J. Bioucas-Dias, S. Galliani, E. Baltsavias, and K. Schindler, "Super-resolution of Sentinel-2 images: Learning a globally applicable deep neural network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 305–319, 2018.

[42] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1132–1140.

[43] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.

[44] L. Wagner, L. Liebel, and M. Körner, "Deep residual learning for single-image super-resolution of multi-spectral satellite imagery," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. IV-2-W7, no. 2/W7, pp. 189–196, Sep. 2019.
[45] M. Galar, R. Sesma, C. Ayala, and C. Aranda, "Super-resolution for Sentinel-2 images," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W16, pp. 95–102, 2019.
[46] M. Galar, R. Sesma, C. Ayala, L. Albizua, and C. Aranda, "Learning super-resolution for Sentinel-2 images with real ground truth data from a reference satellite," *ISPRS Ann. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. V-1-2020, pp. 9–16, 2020.
[47] N. Latte and P. Lejeune, "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 Sens.*, vol. 12, no. 15, Jul. 2020, Art. no. 2366.
[48] M. Zabalza and A. Bernardini, "Super-resolution of Sentinel-2 images using a spectral attention mechanism," *Remote Sens.*, vol. 14, no. 12, Jun. 2022, Art. no. 2890.
[49] W. Xu, G. Xu, Y. Wang, X. Sun, D. Lin, and Y. Wu, "High quality remote sensing image super-resolution using deep memory connected network," *Int. Geosci. Remote Sens. Symp.*, vol. 2018-Jul., pp. 8889–8892, Oct. 2020.
[50] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, "Learning texture transformer network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5790–5799.
[51] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, and T. Zeng, "Transformer for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2022, pp. 456–465.
[52] W. G. C. Bandara and V. M. Patel, "Hypertransformer: A textural and spectral feature fusion transformer for pansharpening," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 1757–1767.
[53] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Hyperspectral image super-resolution with optimized rgb guidance," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11653–11662.
[54] Z. Wang, B. Chen, R. Lu, H. Zhang, H. Liu, and P. K. Varshney, "Fusionnet: An unsupervised convolutional variational network for hyperspectral and multispectral image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 7565–7577, Jun. 2020, doi: 10.1109/TIP.2020.3004261.
[55] Y. Chen, J. Zeng, W. He, X.-L. Zhao, and T.-Z. Huang, "Hyperspectral and multispectral image fusion using factor smoothed tensor ring decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5515417.
[56] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[57] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[58] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[59] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?," in *Proc. 3rd Conf. Fusion Earth Data: Merging Point Meas., Raster Maps Remotely Sensed Images*, 2000, pp. 99–103.
[60] R. H. Yuhas, A. F. H. Goetz, J. W. Boardman, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," *JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop, Volume 1: AVIRIS Workshop*, Jun. 1992.
[61] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat tm and spot panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.
[62] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.
[63] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.
[64] C. Lin, C. C. Wu, K. Tsogt, Y. C. Ouyang, and C. I. Chang, "Effects of atmospheric correction and pansharpening on LULC classification accuracy using WorldView-2 imagery," *Inf. Process. Agriculture*, vol. 2, no. 1, pp. 25–36, May 2015.
[65] J. K. Gilbertson, J. Kemp, and A. van Niekerk, "Effect of pan-sharpening multi-temporal landsat 8 imagery for crop type differentiation using different classification techniques," *Comput. Electron. Agriculture*, vol. 134, pp. 151–159, Mar. 2017.
[66] P. Bose, D. Halder, O. Rhaman, and T.H. Pial, "Effectivity of super resolution convolutional neural network for the enhancement of land cover classification from medium resolution satellite images," in *Proc. 38th Asian Conf. Remote Sens.*, 2017.
[67] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens., Vol. 63 , No. 6 , Jun. 1997, pp. 691–699. Amer. Soc. Photogrammetry Remote Sens.*, vol. 63, pp. 691–699, no. 6, Jun. 1997.
[68] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
[69] I. S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-means, spectral clustering and normalized cuts," *J. Mach. Learn. Res.*, vol. 5, pp. 1435–1467, 2004.
[70] H. Zhou et al., "A novel spectral clustering algorithm based on neighbor relation and Gaussian kernel function with only one parameter," *Soft Comput.*, 2023, doi: 10.1007/s00500-023-09309-z.
[71] B. Scholkopf and A. J. Smola, "Nonlinear component analysis as a kernel eigenvalue problem," in *Advances in Kernel Methods*. Cambridge, MA, USA: MIT Press, 1998, pp. 327–352.
[72] A. Y. Ng, M. I. Jordan, and Y. Weiss, "Spectral clustering and transductive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, pp. 849–856.
[73] B. Schölkopf, A. Smola, and K. R. Müller, "Kernel principal component analysis," in *Artificial Neural Networks Lecture Notes in Computer Science*, vol. 1327, W. Gerstner, A. Germond, M. Hasler, J. D. Nicoud, Eds. Berlin, Heidelberg, Germany: Springer, 1997, doi: 10.1007/BFb0020217.
[74] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

Anna Malczewska received the B.Sc degree in close-range photogrammetry for small historical objects in 2014 and M.Sc. degree in mobile lidar measurement systems in 2019 from AGH University of Krakow, Poland, Faculty of Geo-Data Science, Geodesy and Environmental Engineering. She is currently working toward the Ph.D. degree in remote sensing super-resolution with the Department of Photogrammetry, Remote Sensing of Environment and Spatial Engineering, AGH Krakow, Poland.

In 2022, she joined to the same Department as a Teaching/Research Assistant. Her research interests include remote sensing, measurement techniques, computer vision, especially super-resolution.

Maciej Wielgosz received the M.Sc. degree in wavelet image compression algorithms for low latency computing from the Faculty of Electrical Engineering, Automatics, Computer Science, and Electronics, AGH University of Science and Technology, Krakow, Poland, in 2005 and the Ph.D. (hons.) degree in high performance reconfigurable computing from the Department of Electronics, AGH University of Science and Technology, Krakow, Poland in 2010.

His research interests include artificial intelligence, image processing, anomaly detection, reinforcement learning, and embedded machine learning. He was awarded the Habilitation Degree in 2021 in the discipline of information and communication technology for his thematically related cycle of scientific articles titled "Design of low-latency architectures for machine learning algorithms." Over the years, he has held positions in various research institutions. He has been associated with the NIBIO Norwegian Institute of Bioeconomy Research since September 2022 as a Machine Learning Researcher, where he focuses on ML solutions for Point Cloud, 3D data, and images in Forestry. Prior to that, he worked at the Centre de Visió per Computador (CVC) at Universitat Autònoma de Barcelona, focusing on "Adversarial Cases for Autonomous Vehicles (ARCANE)." Furthermore, since 2010, he has been an Assistant Professor with the AGH University of Science and Technology. His expertise has also been sought by international organizations such as the European Organization for Nuclear Research (CERN), where he worked on detecting anomalous behavior of LHC superconducting magnets. He has also collaborated with Cadence Design Systems, NTNU University of Science and Technology, and the Department of Telematics in Trondheim, Norway.