Sentinel-2 60-m Band Super-Resolution Using Hybrid CNN-GPR Model

Vlad Vasilescu, Mihai Datcu, *Fellow, IEEE*, and Daniela Faur, *Member, IEEE*

***Abstract*— Sentinel-2 image super-resolution (SR) has proven advantageous in multiple data analysis pipelines, leading to a more comprehensive assessment of different environment-related metrics. This research aims to provide a method for super-resolving the 60-m bands provided by Sentinel-2 up to 10-m spatial resolution, using Gaussian process regression (GPR). While common GPR methods directly operate on raw data using carefully designed kernels, we propose a convolutional neural network (CNN)-based feature extraction kernel to directly process the input 10-m patches, applied in constructing the elements of the integrated covariance matrices. For each scene, a small number of training patches are sampled to optimize the CNN parameters and to construct the predictive mean function, the latter being further used for predicting super-resolved pixels for new input areas. We prove that our method is a reliable SR mechanism by assessing its performance both quantitatively, using metrics against other methods from literature, and qualitatively, through visual analysis of the results.**

***Index Terms*— Convolutional neural network (CNN), Gaussian process regression (GPR), Sentinel-2, super-resolution (SR).**

I. Introduction

SENTINEL-2 satellite imaging mission provides continuous high-resolution monitoring, supplying numerous applications with data presented in the form of 13 spectral bands with spatial resolutions of 10, 20, and 60 m. Such applications include monitoring crop areas [1], climate change assessment [2], vegetation health estimation [3], and management of natural disasters [4]. While sufficient for some applications, the 20- and 60-m bands can be enhanced by constructing high-resolution versions that incorporate small-scale features, providing more accurate monitoring and detection for areas of interest.

A plethora of Sentinel-2 super-resolution (SR) methods have been proposed in previous years [5], ranging from kriging methods [6], to inverse imaging problems [7] and to recent deep-learning-based methods [1], [8]. While the latter have proven to generalize well for a variety of environments, they require extensive training using synthetic data to capture the wide variety of spatial and spectral relationships. On the opposite end, previous methods relied on optimizing their internal parameters separately for each scene, fusing the information contained in the high-resolution bands with the radiometric properties of low-resolution ones, along with incorporating sensor-specific degradation-based operators to mimic the construction of low-resolution bands.

In this letter, we propose a Gaussian process regression (GPR) method for super-resolving the 60-m Sentinel-2 bands up to 10-m spatial resolution. The proposed method will incorporate a feature extractor based on a convolutional neural network (CNN) to extract high-resolution features, further used for constructing the covariance matrices through which the predictive mean can be computed for new input locations. To the best of our knowledge, this is the first SR method for Sentinel-2 bands based on GPR. The rest of this letter is organized as follows. Section II provides a theoretical description of GPR, followed by the proposed method. Section III discusses the validation process for our model, along with a comparison to other SR methods. Section IV highlights the concluding remarks and future developments.

II. Methodology

A. Gaussian Process Regression (GPR)

Let us consider an unknown function $f : \mathcal{X} \to \mathcal{Y}$ which we are trying to model using GPs. The set of input points and their observed output will be denoted as $\mathbf{X} = \{x_i\}_{1 \le i \le N}$ and $\mathbf{Y} = \{y_i\}_{1 \le i \le N}$, respectively, where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. Constructing a distribution over function values $f(x_i)$ requires defining a mean function $m(x_i)$ and a covariance function $k(x_i, x_j)$, formally written as $f(\mathbf{X}) \sim \mathcal{GP}(m(\mathbf{X}), k(\mathbf{X}, \mathbf{X}))$, where $k(\mathbf{X}, \mathbf{X}) = \{k(x_i, x_j)\}_{1 \le i,\, j \le N}$ denotes the covariance matrix between all pairs of elements from $\mathbf{X}$. Setting a prior over function f (usually Gaussian, for regression analysis) allows for sampling function values at different locations in its domain, given the mean and covariance functions [9]. For a set of new input points $\mathbf{X}_*$, the inference of its corresponding output set $\mathbf{Y}_*$ is performed by defining a joint Gaussian distribution on the values of $f(\mathbf{X})$ and $f(\mathbf{X}_*)$, followed by conditioning it on the initially known input–output pairs (x_i, y_i), as follows:

$$\begin{bmatrix} f(\mathbf{X}) \\ f(\mathbf{X}_*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{X}) \\ m(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} k(\mathbf{X}, \mathbf{X})\, k(\mathbf{X}, \mathbf{X}_*) \\ k(\mathbf{X}_*, \mathbf{X})\, k(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) \quad (1)$$

Manuscript received 30 May 2023; revised 7 July 2023; accepted 11 July 2023. Date of publication 17 July 2023; date of current version 1 August 2023. This work was supported by the Romanian Ministry of Education and Research, CNCS-UEFISCDI within PNCDI III, under Project PN-III-P4-ID-PCE-2020-2120. *(Corresponding author: Vlad Vasilescu.)*

Vlad Vasilescu is with the Speech and Dialog Laboratory, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania, and also with the Center for Spatial Information, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania (e-mail: vlad.vasilescu2111@upb.ro).

Mihai Datcu is with the German Aerospace Center (DLR), Remote Sensing Technology Institute, EO Data Science, Oberpfaffenhofen, 82234 Weßling, Germany, and also with the Center for Spatial Information, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania (e-mail: mihai.datcu@dlr.de).

Daniela Faur is with the Center for Spatial Information, University POLITEHNICA of Bucharest, 060042 Bucharest, Romania (e-mail: daniela.faur@upb.ro).

Digital Object Identifier 10.1109/LGRS.2023.3296188

$$f(\mathbf{X}_*) \mid f(\mathbf{X}),\ \mathbf{X},\ \mathbf{X}_*, \sim \mathcal{N}(\bar{\mathbf{f}}_*, \mathrm{cov}(\mathbf{f}_*)) \tag{2}$$

$$\bar{\mathbf{f}}_* = m(\mathbf{X}_*) + k(\mathbf{X}_*, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}(\mathbf{Y} - m(\mathbf{X})) \tag{3}$$

$$\mathrm{cov}(\mathbf{f}_*) = k(\mathbf{X}_*, \mathbf{X}_*) - k(\mathbf{X}_*, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}k(\mathbf{X}, \mathbf{X}_*). \tag{4}$$

Equation (2) denotes the posterior distribution, while (3) and (4) give the predictive mean and covariance values for the new input points $\mathbf{X}_*$. In some scenarios, it is reasonable to assume the existence of underlying i.i.d. noise $\epsilon \sim \mathcal{N}(0, \sigma_n I)$ for observations $\mathbf{Y}$, that is, $\mathbf{Y} = f(\mathbf{X}) + \epsilon$, which leads to the adjustment $k(\mathbf{X}, \mathbf{X}) \rightarrow k(\mathbf{X}, \mathbf{X}) + \sigma_n I$ in the previous equations. A prior of $m(\mathbf{X}) = 0$ is usually assumed for computational ease.

While powerful methods, GPs suffer from a computational bottleneck associated with the inversion of an $N \times N$ matrix (3) and (4), scaling with the number of points N as $O(N^3)$. Multiple techniques have been previously developed to find a good tradeoff between the power of representation and time complexity for GPR [10], the majority of which is addressing the characteristics of $k(\mathbf{X}, \mathbf{X})$ and its inversion process. One of the most common techniques works by constructing a joint distribution over the function values for a set $\mathbf{Z} = \{z_i\}_{1 \leq i \leq M}$, $z_i \in \mathcal{X}$, of $M \ll N$ *inducing points*, further used for establishing a family of posterior distributions by conditioning the prior of f on these M function values [11], [12]. These latter methods make use of the approximation $\tilde{k}(\mathbf{X}, \mathbf{X}) \approx k(\mathbf{X}, \mathbf{Z})k(\mathbf{Z}, \mathbf{Z})^{-1}k(\mathbf{X}, \mathbf{Z})^T$ to reduce the computational expense of computing the predictive mean and covariance functions, resulting in a complexity of $O(NM^2)$.

An important step in constructing representative GP models is the choice of covariance (kernel) function $k(x_i, x_j)$, which encodes the *similarity* between input points x_i, $x_j \in \mathcal{X}$. The most commonly used covariance function is the squared exponential, defined as $k(x_i, x_j) = \sigma \cdot \exp(-((\|x_i - x_j\|_2)/2l^2))$, which is an *isotropic* function as it only depends on $|x_i - x_j|$. Most of these functions include a set of hyperparameters $\boldsymbol{\theta}_k$ (e.g., the squared exponential kernel has $\boldsymbol{\theta}_k = \{\sigma, l\}$) which needs to be tuned s.t. the posterior distribution fits the training data as good as possible. Titsias [12] provides a way for learning both the hyperparameters $\boldsymbol{\theta}_k$ and the inducing inputs $\mathbf{Z}$, which are considered to be variational parameters during the maximization of a variational lower bound to the marginal likelihood. In [11] and [13], the Sparse Variational Gaussian Process (SVGP) model provides a way to optimize the variational parameters (hyperparameters + inducing variables) by following the ascent direction given the natural gradients of the evidence lower bound (ELBO) of $p(f(\mathbf{X})|\mathbf{Y})$, written in a form that allows for mini-batch optimization. The SVGP uses variational inference s.t. $p(f(\mathbf{X})|\mathbf{Y})$ is approximated with a variational term $q(f(\mathbf{X}))$, which in conjunction with the usage of inducing variables $\mathbf{Z}$, results in the following maximization objective:

$$\mathcal{L} = \sum_{i=1}^{N} \mathbb{E}_{q(f(x_i))} \log(p(y_i|f(x_i))) - \mathbf{KL}[q(f(\mathbf{Z}))\|p(f(\mathbf{Z}))] \tag{5}$$

which allows for computing the partial derivatives with respect to the kernel hyperparameters $\boldsymbol{\theta}_k$ and inducing points $\mathbf{Z}$. Maximizing $\mathcal{L}$ implies maximizing the first term—optimizing $q(\cdot)$ s.t. it becomes a good approximation for $f(\cdot)$ on all input locations $\mathbf{X}$—and minimizing the Kullback–Leibler divergence, which serves the same goal for the inducing points $\mathbf{Z}$.

B. Proposed Method

The proposed framework is illustrated in Fig. 1. Our method is based on the previously described SVGP model, optimized for learning an input–output mapping from 10-m $k \times k$ patches, extracted from available 10-m bands B2, B3, B4, and B8, to super-resolved 10-m pixels values for 60-m bands B1 and B9. More precisely, given the 10-m patches, the model is trained to produce a *residual map*, which added over bicubically upsampled 60-m bands should yield valid 10-m predictions. The idea was to retain the radiometric distribution of the original 60-m bands (through bicubic interpolation) while increasing their high-resolution content.

Directly comparing pixel values through common kernel functions would not yield results robust with respect to translation and rotation, thus falling short of being a good proxy for the similarity between different patches. Driven by these shortcomings, we propose to include in the learning process an additional feature extractor based on a CNN, to obtain 1-D representations for 10-m input patches. The intrinsic characteristics of the convolution allow for obtaining similar embeddings for relatively close spatial regions, and also for areas exhibiting local similar patterns, driving these representations to be translation/location-invariant. This further allows for constructing covariance matrices that better encode pair-wise similarities between different patches. Let us denote our neural network function parametrized by $\boldsymbol{\theta_{net}}$ as $T_{\boldsymbol{\theta_{net}}}(\cdot) : \mathbb{R}^{k \times k \times 4} \mapsto \mathbb{R}^{d_e}$, acting on concatenated $k \times k$ patches from all 10-m bands and computing d_e-dimensional embeddings. Given two different inputs $x_i, x_j \in \mathbb{R}^{k \times k \times 4}$, their corresponding entry in the covariance matrix is now computed as $k(T(x_i), T(x_j))$, where $k(\cdot, \cdot)$ is chosen to be the commonly used squared exponential. The set of parameters to be optimized is now extended as $\boldsymbol{\theta} = \{\boldsymbol{\theta_k}, \boldsymbol{\theta_{net}}\}$, which, along with the inducing point $\mathbf{Z}$, will be modified according to the ascent direction of the loss described in (5).

To optimize the previously discussed set of parameters, the model was trained in a reduced-resolution context by generating synthetic input–output pairs, with spatial resolutions of 60, and 10 m, respectively. Given a set of 10-m input bands $\mathbf{X}_{10} \in \mathbb{R}^{H \times W \times 4}$, one can simulate the degradation process by a factor of 6 using the sensors' modulation transfer function (MTF), which encompasses a depth-wise convolution operation between $\mathbf{X}_{10}$ and Gaussian kernel g_σ with variance σ, followed by a depth-wise convolution with a 6×6 averaging filter d, applied with stride 6

$$\mathbf{X}_{10} \downarrow_6 = \mathrm{MTF}(\mathbf{X}_{10}) = (\mathbf{X}_{10} * g_\sigma) * d \in \mathbb{R}^{\frac{H}{6} \times \frac{W}{6} \times 4} \tag{6}$$

where $\mathbf{X}_{10} \downarrow_6$ denotes the simulated 60-m bands, and the variance for the Gaussian kernel is chosen as $\sigma = 3$, following the protocol described in [14]. For a $k \times k \times 4$ patch extracted from $\mathbf{X}_{10} \downarrow_6$, the prediction target was chosen to be the corresponding center pixel from 60-m bands

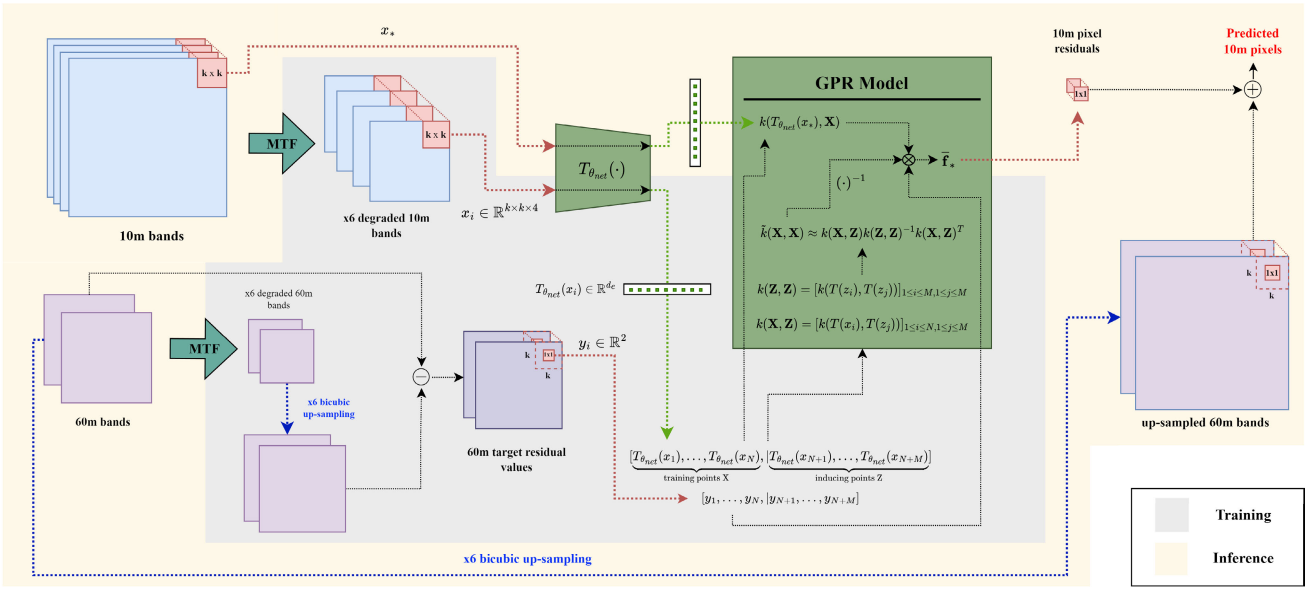

Fig. 1. Proposed SR method for Sentinel-2 60-m bands. Training points x_i were generated using a degraded version of the 10-m bands, while the *residual values* y_i extracted from 60-m bands were taken as output. The GPR model (including $T_{\theta_{\text{net}}}$) is trained to approximate the mapping between 60-m pixels from degraded 10-m bands to 60-m pixels in the original 60-m bands. The inference phase utilizes this learned mapping to obtain 10-m predictions for the super-resolved 60-m bands, taking as input data 10-m patches x_* and applying the previously optimized algorithm to compute the residual values $\bar{\mathbf{f}}_*$.

(see Fig. 1 for illustration of the process). Since the model predicts residual values to be added over bicubically magnified low-resolution bands, the target values were constructed by first degrading the 60-m bands (using (6)), followed by applying $\times 6$ bicubic upsampling and subtracting the result from the original 60-m bands.

For each scene, a number of patches have to be sampled for training the SR model, balancing the representational power and the computational complexity of $k(\mathbf{X}, \mathbf{Z})$. Given that these N training points should be as representative as possible for the current scene, we propose a sampling process based on per-patch variance, to avoid near-flat areas: 1) compute the variance of all $p \times p$ nonoverlapping 60-m patches; 2) normalize each to their sum and construct a discrete probability distribution over their center locations by assigning the normalized variances to their probabilities; and 3) sample N points from the constructed distribution to serve as training data.

III. Experiments

A. Sentinel-2 Data and Evaluation Metrics

To validate the performance of our model, we used Level-1C products provided by Sentinel-2, each spanning an approximate area of 100×100 km^2. Three areas were selected for performance assessment, further referenced by their geographic location: Bucharest, Romania[1]; coastline of the Tyrrhenian sea, Italy[2]; Dilo, Ethiopia.[3] As for the numerical evaluation framework, we adopted *Wald's* protocol [15] for assessing the performance in reduced-resolution conditions. This coincides with how we create our synthetic training data, taking the 60-m bands as targets and the degraded 10-m bands as inputs. Since the target consists of 60-m spatial resolution data, we further denote this process as 360–60 m SR. Our target represented a residual value, which, added over the bicubically upsampled 60-m bands, yielded the 10-m super-resolved response. Before any processing, all bands were divided by their corresponding maximum value. To measure the error between our prediction and the original 60-m bands, we used as evaluation metrics root-mean-square error (RMSE), signal-to-reconstruction error (SRE), and spectral angle mapper (SAM).

B. Performance Comparison and Analysis

For the SR framework presented in Fig. 1, we choose the following architecture for the CNN implementing feature extraction $T_{\theta_{\text{net}}}(\cdot)$: three 2-D convolutional layers, with 16–32–64 filters of spatial dimension 3×3, followed by two fully connected layers with 256–128 units ($d_e = 128$). *ReLU* activation is used after each layer, except for the last one. The simplicity of the chosen architecture is motivated by two factors: the use of relatively small patches for training—to avoid taking into account redundant spatial details for prediction—and to bypass the possibility of overfitting. Squared exponential is used for measuring the distance between two 128-D embedding vectors. For each of the three Sentinel-2 scenes, we trained a different SR model by sampling $N = 8000$ input–output pairs, from which we selected $M = 200$ inducing points by applying *k-means* on all N sampled points, selecting the $k = M$ centroids as our initial $\mathbf{Z}$ values. The models were optimized using $10k$ iterations of the *L-BFGS-B* algorithm [16] to adapt all parameters $\boldsymbol{\theta}$. We used patches of size 13×13, as this resulted in the best performance. Several methods were used to compare

[1]S2A_MSIL1C_20210823T090601_N0301_R050_T35TMK_20210823T112038.

[2]S2A_MSIL1C_20220323T095031_N0400_R079_T33TWE_20220323T104033.

[3]S2A_MSIL1C_20161230T074322_N0204_R092_T37NCE_20161230T075722.

TABLE I
RESULTS ON REDUCED-RESOLUTION EVALUATION FOR 360 → 60 m SR

		Italy			Romania			Ethiopia		
		RMSE↓	SRE (dB)↑	SAM (deg)↓	RMSE↓	SRE (dB)↑	SAM (deg)↓	RMSE↓	SRE (dB)↑	SAM (deg)↓
B1	Bicubic	96.42	27.55	2.507	45.06	30.3	1.753	111.56	22.02	4.593
	SSSS	40.57	34.97	1.022	43.48	30.6	1.692	67.56	26.38	2.772
	VDSen2	46.51	33.81	1.181	20.45	37.16	0.794	57.05	27.85	2.328
	DSen2	39.88	35.21	1.009	17.91	38.31	0.696	41.70	30.57	1.699
	ATPRK	23.26	39.89	0.588	**10.83**	**42.68**	**0.421**	47.50	29.44	1.949
	GPR	**19.94**	**41.23**	**0.492**	14.49	40.15	0.559	**32.4**	**32.77**	**1.327**
B9	Bicubic	158.7	22.27	4.447	52.88	20.22	5.63	64.29	18.33	6.961
	SSSS	89.18	27.27	2.508	22.01	27.83	2.33	50.03	20.51	5.339
	VDSen2	76.18	28.64	2.119	28.76	25.51	3.052	35.43	23.51	3.802
	DSen2	66.55	29.81	1.858	28.16	25.69	2.988	31.01	24.67	3.332
	ATPRK	54.55	31.54	1.503	13.88	31.83	1.468	38.91	22.7	4.201
	GPR	**34.62**	**35.49**	**0.964**	**10.64**	**34.14**	**1.122**	**21.08**	**28.02**	**2.291**

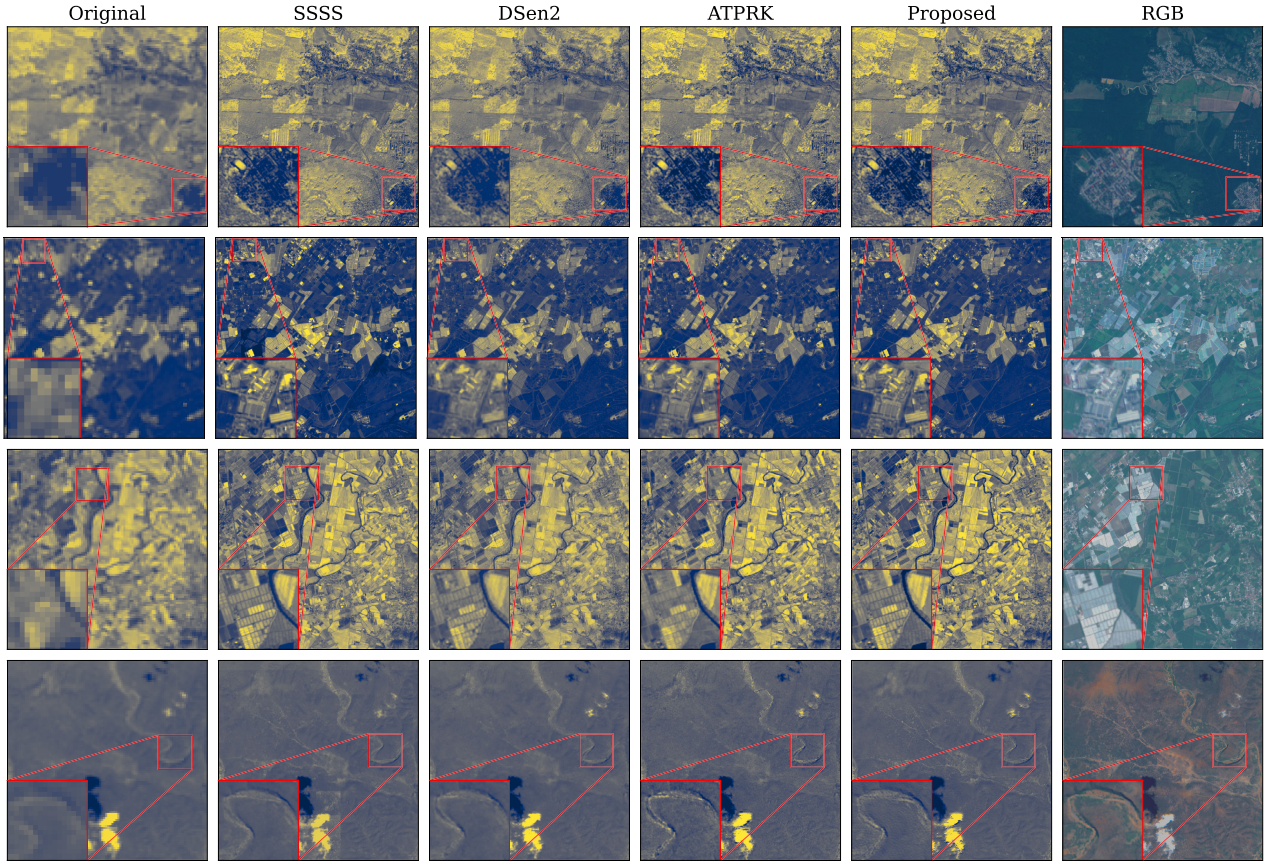

Fig. 2. Results on full-resolution evaluation for 60 → 10 m SR. The illustrated super-resolved images cover an approximate area of 6 × 6 km^2, with the zoomed-in regions covering 1 × 1 km^2. The first, third, and fourth rows represent results for band B9, while the second is for band B1. The first area is extracted from the Romania tile, the next two from Italy, and the last one from Ethiopia.

our model's performance to: bicubic upsampling, SSSS [7], DSen2, and its deeper version VDSen2 [8] and ATPRK [6]. It is important to note that all numerical evaluations were performed by excluding the N sampled training points, ensuring a fair comparison between methods.

Numerical outcomes for reduced-resolution evaluation (360 → 60 m SR) are presented in Table I, for all methods and separately for each Sentinel-2 area. All results are reported in the original range of values (nonnormalized), for each band. Our method resulted in the best performance in all three areas for band B9, followed by ATPRK with the second-best results. On band B1, the proposed method gave the best performance on super-resolving Italy and Ethiopia tiles, while achieving the second-best performance for the Romania area, being surpassed by ATPRK. To investigate this, we measured the Pearson correlation between our prediction and ATPRK's on band B1, and the available 10-m bands. This resulted in a mean coefficient of 0.625 for ATPRK and 0.547 for ours,

TABLE II
QNR [%] ↑ FOR FULL-SCALE SR 60 → 10 m

	Italy	Romania	Ethiopia	Average
Bicubic	70.45	77.76	62.59	70.26
SSSS	79.05	88.20	69.38	78.87
DSen2	79.72	86.00	68.27	77.99
ATPRK	80.70	88.54	70.00	79.74
GPR	81.05	89.38	69.52	**79.98**

the main difference coming from band B4 where ATPRK resulted in a coefficient of 0.977, while ours in 0.728. Since band B4 is representative of urban areas, along with vegetation crops, and since Romania tile is highly reflected in these characteristics, we concluded that an increased performance for ATPRK was achieved through the ease of controlling its correlation with other 10-m bands. Therefore, ATPRK may be a better fit in the case of areas highly represented by the specifics of an available 10-m band, while falling short in environmentally heterogeneous areas, where linear combinations of these 10-m bands are not sufficient. Note that the results obtained by SSSS for this band are fairly similar to the ones obtained through bicubic interpolation, indicating a deviation from the reflectance distribution of the original band. DSen2 takes third place in almost all evaluations, surpassing its deeper version VDSen2 by a large margin, showcasing better generalization.

In the case of full-scale SR, that is, applying the previous algorithms on the original 10-m bands, the evaluation relies solely on visual inspection for the predicted bands. While any numerical evaluation for SR with no reference can lead to divergent conclusions, several efforts have been conducted to develop such evaluations. One of them is quality no-reference (QNR) [17] which combines spatial and spectral distortion in a single metric, extended in [14] for S2 SR. Table II presents the QNR for full-scale SR, given as percentages. Our method and ATPRK result in the best performances with, however, a very small difference between them. Some visual results sampled from the three scenes are presented in Fig. 2, along with the 10-m RGB representation for detailed comparison. DSen2 results in somewhat blurry details, especially for the last, partially clouded area from the Ethiopian desert. SSSS caused fairly different radiometric distributions for band B1, relative to the original bands, supporting the high RMSE magnitudes for reduced-resolution evaluation from Table I. Our method and ATRPK obtained the best overall visual results, aligning with the observed high-frequency details in the 10-m RGB images and with the results from Table II, thus validating the quality of induced high-resolution information. Finally, we would like to enhance that, even if the QNR can sometimes provide a proxy for the quality of SR results, not relying on a reference image may lead to inconsistent comparisons during the evaluation phase. Thus, greater emphasis should be placed on evaluation protocols that incorporate reference data.

IV. CONCLUSION

In this research, we proposed a Sentinel-2 60-m band SR method using GPR based on a CNN feature extractor. Our method was optimized on synthetically degraded image patches and was tested in degraded and full-resolution contexts, obtaining top performance in both quantitative and qualitative evaluations. The main drawback of our system is given by the necessity of being optimized separately for each scene, which could otherwise contribute to significant growth of the covariance matrix, leading to impractical solutions. Future advancements cover the development of techniques aimed at increasing the generalization capabilities, for the same model to be used on multiple areas and the use of predicted covariance values in a postprocessing step targeting high-covariance regions.

REFERENCES

[1] L. Tulczyjew, M. Kawulok, N. Longépé, B. Le Saux, and J. Nalepa, "Graph neural networks extract high-resolution cultivated land maps from Sentinel-2 image series," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[2] R. Barella et al., "Combined use of Sentinel-1 and Sentinel-2 for glacier mapping: An application over central east Alps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4824–4834, 2022.

[3] Z. Chen, T. Shi, X. Zhang, K. Jia, H. Jiang, and B. Yuan, "A hybrid leaf area index estimation method of dioscorea polystachya turczaninow using Sentinel-2 vegetation indices," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4415713.

[4] G. I. Drakonakis, G. Tsagkatakis, K. Fotiadou, and P. Tsakalides, "OmbriaNet—Supervised flood mapping via convolutional neural networks using multitemporal Sentinel-1 and Sentinel-2 data fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2341–2356, 2022.

[5] S. E. Armannsson, M. O. Ulfarsson, J. Sigurdsson, H. V. Nguyen, and J. R. Sveinsson, "A comparison of optimized Sentinel-2 super-resolution methods using Wald's protocol and Bayesian optimization," *Remote Sens.*, vol. 13, no. 11, p. 2192, Jun. 2021.

[6] Q. Wang, W. Shi, Z. Li, and P. M. Atkinson, "Fusion of Sentinel-2 images," *Remote Sens. Environ.*, vol. 187, pp. 241–252, Dec. 2016.

[7] C.-H. Lin and J. M. Bioucas-Dias, "An explicit and scene-adapted definition of convex self-similarity prior with application to unsupervised Sentinel-2 super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3352–3365, May 2020.

[8] 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. Photogramm. Remote Sens.*, vol. 146, pp. 305–319, Dec. 2018.

[9] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *J. Math. Psychol.*, vol. 85, pp. 1–16, Aug. 2018.

[10] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian process meets big data: A review of scalable GPs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4405–4423, Nov. 2020.

[11] J. Hensman, N. Fusi, and N. D. Lawrence, "Gaussian processes for big data," in *Proc. 29th Conf. Uncertainty Artif. Intell.*, 2013, pp. 282–290.

[12] M. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *Proc. Artif. Intell. Statist.*, 2009, pp. 567–574.

[13] J. Hensman, A. Matthews, and Z. Ghahramani, "Scalable variational Gaussian process classification," in *Proc. Artif. Intell. Statist.*, 2015, pp. 351–360.

[14] V. Vasilescu, M. Datcu, and D. Faur, "A CNN-based Sentinel-2 image super-resolution method using multiobjective training," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 4700314.

[15] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[16] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, pp. 550–560, 1997.

[17] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogramm. Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, Feb. 2008.