A Bayesian Deep Image Prior Downscaling Approach for High-Resolution Soil Moisture Estimation

Yuan Fang ⓘ, Linlin Xu ⓘ, *Member, IEEE*, Yuhao Chen, *Member, IEEE*, Wei Zhou,
Alexander Wong, *Senior Member, IEEE*, and David A. Clausi ⓘ, *Senior Member, IEEE*

Abstract—**Soil moisture (SM) estimation is a critical part of environmental and agricultural monitoring, with satellite-based microwave remote sensing being the main SM source. However, the limited spatial resolution of most current remote sensing SM products reduces their utility for many applications, such as evapotranspiration modeling and agriculture management. In this article, to address this issue, we propose a Bayesian deep image prior (BDIP) downscaling approach to estimate the high-resolution SM from satellite products. More specifically, the high-resolution SM estimation problem is formulated as a maximum *a posteriori* problem, and solved via a neural network comprising of a deep fully convolutional neural network (FCNN) for modeling the prior spatial correlation distribution of the underlying high-resolution SM variables, and a forward model characterizing the SM map degeneration process for modeling the data likelihood. As such, the proposed BDIP approach provides a statistical framework that integrates deep learning with forward modeling in a coherent manner for combining different sources of information, i.e., the knowledge in the forward model, the spatial correlation prior in FCNN architecture, and the remote sensing data and products. Experiments on the downscaling of SM active passive SM products using the moderate resolution imaging spectroradiometer products show that SM maps estimated using the proposed method provide greater spatial detail information than other downscaling methods, with the SM estimates very close to *in situ* measurements.**

Index Terms—**Bayesian, convolutional neural network, deep image prior, MODIS, SMAP, soil moisture downscaling.**

I. INTRODUCTION

SOIL moisture (SM) highly influences hydrologic and atmospheric processes for environmental and agricultural monitoring. Microwave remote sensing (RS), with the high sensitivity to the SM variation and robustness to atmosphere conditions, is the most commonly used approach to monitor SM [1]–[3]. The SM active/passive (SMAP) mission has been providing SM at two spatial resolutions of 36 and 9 km since April 2015 [4]. However, these two spatial resolutions do not meet the requirements

Yuan Fang, Linlin Xu, Yuhao Chen, Alexander Wong, and David A. Clausi are with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: yuan.fang@uwaterloo.ca; l44xu@uwaterloo.ca; yuhao.chen1@uwaterloo.ca; alexander.wong@uwaterloo.ca; dclausi@uwaterloo.ca).

Wei Zhou is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: wei.zhou@uwaterloo.ca).

for application to evapotranspiration modeling and agriculture management [4], [5]. Therefore, improving the spatial resolution of the SMAP SM product to 1 km spatial resolution is essential. Downscaling is an inverse problem that reconstructs images at higher resolution from coarse observations. Since SM has high variation over spatial scales smaller than the SMAP resolutions, spatial heterogeneity must be properly addressed when downscaling [3].

The SM downscaling can be achieved by different strategies, e.g., data fusion or assimilation [6]–[9], geostatistical [3], [10], traditional regression [11], and machine learning (ML) [12]. Data fusion and assimilation for the downscaling is achieved by combing multisources data and extracting more accurate spatial SM information. Geostatistical methods interpolate the SM product with geographical models based on the certain spatial assumption, e.g., the geographically weighted regression [3]. The traditional regression method uses a simple regression model, e.g., a linear regression model, to analysis the correlation between the SM and other RS products [11], which could not sufficiently explore the complex relationship between them. ML methods (e.g., the decision tree regression [1]) show stronger potential in SM downscaling by building the nonlinear relationships between the SM and other indices. Recently, deep learning using multilayer perceptron (MLP) has been adopted to SM downscaling due to its capability in learning complex relationships between inputs (i.e., the coarse-resolution SM and fine-resolution ancillary products) and the target data (i.e., the fine-resolution SM), and its short inference time after training [4]. This downscaling model is trained using SM products with different resolutions, and then can be used for improving the spatial resolution of SM products by the same scale as training. However, the scale of SM products to reduce is limited by the scale difference between the two SM products used for training. In addition, MLP cannot effectively model the spatial correlation of the SM. A convolutional neural network (CNN)-based downscaling method [13] is proposed recently, which can better exploit the spatial information within adjacent pixels. However, most of the ML-, MLP-, and CNN-based methods are supervised, requiring the groundtruth data, which could be *in situ* measurements or the high-resolution SM products, and as such the performance is greatly dependent on the training dataset. [14].

Fully CNNs (FCNNs) have been widely used in various tasks, including semantic segmentation [15], [16], superresolution [17], and image denoising [18], as examples. Unlike the classic CNN, the FCNN contains no fully connected layers, and
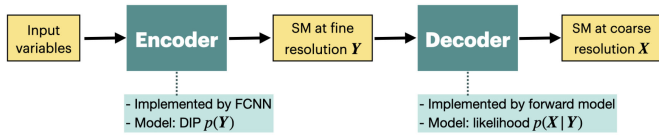
Fig. 1. BDIP model for SM downscaling. The encoder is implemented as an FCNN accounting for the spatial correlation prior of the high-resolution SM, and the decoder part works as the forward model.

it can take input of arbitrary size. Ulyanov *et al.* [19] demonstrated that the structure of an FCNN is sufficient to capture low-level image statistics [19], which is called "deep image prior (DIP)." The FCNN can capture appropriate global spatial features [20] with a wide image field-of-view compared with a CNN layer [21]. Also, compared with traditional methods [3], [11], [22], FCNN is much more computationally efficient by leveraging GPUs. In addition, many of downscaling methods assume a linear scaling relationship between optical-derived input variables and SM, which is not always satisfied [23]. Accounting for nonlinearities between SM and the input variables, FCNN, as an empirical method, has more potential compared to traditional models [23].

We integrate the DIP captured by FCNN into a Bayesian framework to address the SM downscaling inverse problem. Then, the resulting downscaling model becomes a Bayesian deep image prior (BDIP) downscaling network, where the inverse model is implemented by an FCNN accounting for DIP, the forward model is modeled by a downsampler describing the relationship between low- to high-resolution SM map.

Contributions of this article are summarized as follows.

1) We adopt a BDIP scheme to SM downscaling to account for the spatial heterogeneity in higher resolution SM maps.
2) The forward model describing the spatial resolution decreasing process from high- to low-resolution SM map is integrated into the Bayesian framework to solve the inverse problem.
3) The resulting maximum *a priori* (MAP) problem is solved by the back-propagation instead of using the typical expectation–maximization (EM) iterative method, which makes the model optimization simple and effective.
4) The proposed method reconstructs the SM in high spatial resolution only by extracting information from high-resolution RS products and the low-resolution SM using DIP, without requiring any ground-truth data for model training.

The proposed method is designed to effectively downscale SMAP SM products at 9 km spatial resolution to 1 km resolution, which can facilitate the generation of 1 km SM maps using coarse SM product and some ancillary data, and thereby can enhance the hydrological monitoring in the study area by offering more spatially detailed hydrological information of the study area. The method is evaluated qualitatively and quantitatively, and results demonstrate that the proposed approach achieves new state-of-the-art results compared with other unsupervised methods.

II. RELATED WORK

A. Downscaling Methods for RS Products

Several categories of methods have been developed to downscale RS products, i.e., image fusion algorithms [22], geostatistical methods [3], traditional regression method [11], and ML methods [12].

In recent years, neural networks have become popular in RS problems due to their capability in learning complex relationships between inputs and target data and the high training efficiency using GPUs. However, the FCNN, with better capability to capture image spatial heterogeneous features relative to a fully connected network [21], has, to the best of our knowledge, not been applied to SM downscaling. We use an FCNN to map the nonlinear relationship between the model input and high-resolution SM map to better construct the spatial information in SM map and to improve the computation efficiency.

B. Deep Image Prior (DIP)

The structure of FCNN is capable to capture image statistical information, and to impose an effective prior to restore high-quality images from low-quality images without seeing a large training dataset [21]. Recent publications show the effectiveness of DIP for image restoration [24]–[26], e.g., hyperspectral image unmixing [27], superresolution [28], image inpainting [21], and denoising [29], [30].

C. Bayesian Data Inversion

The Bayesian method is used widely to solve the image inverse problem by integrating prior knowledge about the desired image with the posterior distributions of the observed images [31]–[33]. For example, the Markov random field has been incorporated as the prior for hyperspectral image unmixing in the Bayesian framework [32]. The rapidly developing deep learning technique solves inverse problems in many applications from the new perspective by various network architectures [34]. However, studies using deep learning in a Bayesian framework are few. It has been a trend to unify probabilistic framework integrating deep learning and Bayesian models to boost the model performance and enhance the image perception [31], e.g., variational autoencoders [35], generative adversarial networks [36], and probabilistic generalized stacked denoising autoencoders [37].

Traditionally, Bayesian inverse problems are solved by iterative algorithms [32], e.g., EM [38], which is very time-consuming and highly relies on accurate initial values of unknown parameters.

III. PROBLEM FORMULATION

We assume that the RS product with low spatial resolution is $X = \{x_i | i = 1, 2, \ldots, m \times n\}$, and the RS product with high spatial resolution is $Y = \{y_i | i = 1, 2, \ldots, \alpha^2 \times m \times n\}$, where α is the ratio between the low and high spatial resolutions. Given the forward mapping $g(\cdot)$ from Y to X, the low-resolution image X can be represented as follows:

$$X = g(Y) + N \tag{1}$$

where $N \in \mathbb{R}^{m \times n}$ is the noise matrix.

The RS product downscaling aims to infer the high-resolution image Y based on the observed low-resolution image X, which in a Bayesian framework can be achieved by maximizing the posterior distribution $p(Y|X)$, i.e.,

$$p(Y|X) \propto p(X|Y)p(Y). \tag{2}$$

Given the generative model $g(\cdot)$ of X in (1) and the posterior distribution in (2), several key factors for effective downscaling are identified as follows.

1) The effective modeling of the high-resolution image prior $p(Y)$ is critical for regulating and estimating the high-resolution image Y.
2) Meaningful modeling the data likelihood $p(X|Y)$ is essential for guiding and regulating the downscaling process.
3) An efficient optimization scheme for solving the Bayesian inverse problem is necessary.

In this article, $p(Y)$ is achieved by the DIP approach using FCNN, as detailed in Section III-A. The data likelihood $p(X|Y)$ is modeling by a distribution incorporating the forward model, as detailed in Section III-B. An efficient optimization scheme is designed and implemented in Section IV. The designed Bayesian DIP downscaling model is shown in Fig. 1.

A. Prior of the High-Resolution SM Map

There are three key requirements on the high-resolution SM Y when designing the prior $p(Y)$.

1) The large-scale heterogeneous spatial correlation effect in SM map should be fully exploited.
2) SM should be in the meaningful value range of [0,1].
3) High-resolution SM prior should allow efficient optimization.

Here, we represent the prior over the high-resolution SM Y by a distribution expressed as follows:

$$p(Y) = \frac{1}{z}\exp(-\delta(Y, E(Y))) \tag{3}$$

where $E(Y)$ is the expectation of Y, which is implemented as an FCNN, and $\delta(u, v)$ is the distance function measuring the distance between vectors u and v.

The prior spatial information of Y can be captured by an FCNN structure [19], which has a wide field of view of the input image compared with a patch-based CNN and can be optimized efficiently on GPUs. Using $f(\cdot)$ to represent the FCNN forward propagation, the expected Y is written as follows:

$$E(Y) = f(Z, \beta) \tag{4}$$

where Z is the input random noise and β is the set of model parameters, including all weights of convolution kernels and biases. We use a "hourglass" architecture with the skip connection [19] to model a mapping $f(\cdot)$ from the input variable Z to the high-resolution SM map Y due to its excellent feature extraction and noise-resistant capability.

We change the "ReLU" activation function in the original U-Net architecture to the "sigmoid" activation, because the value of SM is in the range from 0 to 1. In addition, the 1×1 convolution final layer for segmentation is changed to map the extracted

feature to the SM output with one layer. The output layer is activated by "sigmoid" which normalizes the value of the input into [0, 1]. We reduced the feature number for each layer from [64, 128, 256, 512] to [2,4,8,16]. The "hourglass" architecture used is shown in Fig. 2.

B. Data Likelihood

The data likelihood is expressed as follows:

$$p(X|Y) = \frac{1}{Z}\exp(-\delta(X, g(Y)) \tag{5}$$

where $\delta(X, g(Y))$ is the distance between the low-resolution SM map X and the reconstructed low-resolution SM map $g(Y)$. The distance function could be implemented with different specific functions based on the real data characteristic. For example, it can be implemented as an L2-norm function when the image noise satisfies the Gaussian distribution, or as an L1-norm for Laplace distribution.

IV. BDIP MODEL OPTIMIZATION

The downscaling problem in (2) can be solved by the MAP approach, where the high-resolution SM map Y is estimated by maximizing the posterior distribution of Y given the observed low-resolution SM map X, i.e.

$$\hat{Y} = \arg\max_{Y}\{p(Y|X)\}. \tag{6}$$

Maximizing $p(Y|X)$ is equivalent to minimizing its negative logarithm likelihood, i.e.,

$$\hat{Y} = \arg\min_{Y}\{-\log p(Y|X)\}. \tag{7}$$

Then, the objective function can be written as follows:

$$J_Y = \arg\min_{Y}\{-\log p(Y|X)\}$$
$$\propto \arg\min_{Y}\{-\log p(X|Y) - \log p(Y)\}. \tag{8}$$

Considering (3) and (5), the objective function can be reformulated as follows:

$$J_Y = \arg\min_{Y}\{\delta(X, g(E(Y|M)))\} \tag{9}$$

where $E(Y|M)$ is the posterior expectation of Y if given ancillary RS data with the high spatial resolution M. We use $E(Y|M)$ as the expectation of Y. To estimate parameters in $E(Y|M)$, we use M as an input to FCNN and optimize FCNN parameters. Given the estimated parameters in FCNN, we achieve $\hat{Y} = E(Y|M)$, as shown in Section III-A. When estimating parameters in FCNN for obtaining $E(Y|M)$, we use a reconstruction distance based on $\delta(X, g(E(Y|M)))$, which incorporates the forward model to constrain the meaningful Y estimation, as shown in Section V-C.

To estimate $E(Y|M)$, we first need to estimate the model parameters in FCNN, i.e., β. Here, we construct the following objective function to estimate β:

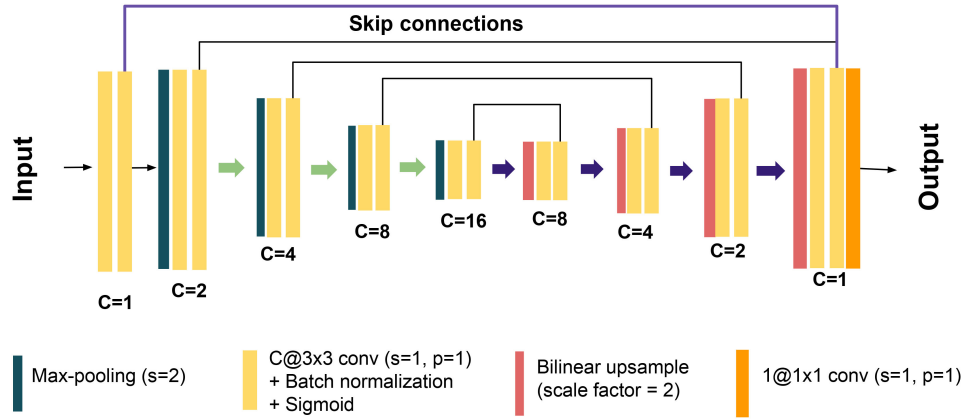$$\hat{\beta} = \arg\min_{\beta}\{\delta(X, g(f(M, \beta)))\}. \tag{10}$$

Fig. 2. "Hourglass" architecture with skip-connections of the FCNN part in Fig. 1 accounting for DIP. Blocks in the figure represent for operations rather than features. The U-Net type "hourglass" architecture [39] encodes an input image to a feature tensor with a smaller size and more channels at the bottleneck, and decodes the feature tensor to the output image of the same size as the input image. The downsampling reduces the feature size, which is essentially achieved by the max-pooling operation (green blocks). The upsampling recovers the feature size step-by-step by bilinear upsampling operations (red blocks). Double regular convolution operations (yellow blocks) are conducted after each max-pooling or TransConv operation, which does not change the feature size but increases or decreases the channel number (i.e., C) of features. The skip connection is implemented by copying and concatenating features.
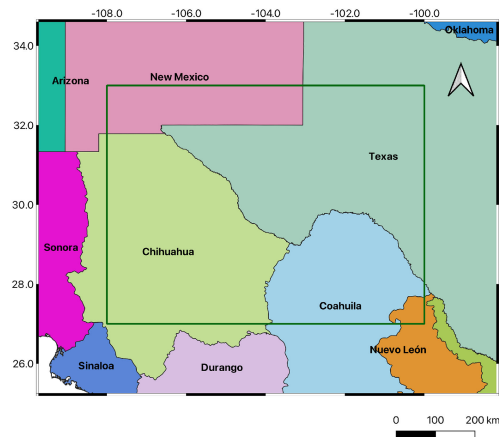

Fig. 3. Location of the study area.

Fig. 4. Distribution of the stations providing *in situ* SM measurements and Land cover map of the study area.

Backpropagation with the Adam stochastic optimizer [40] is adopted in this work to estimate β.

V. METHOD

A. Study Area and Datasets

We select a rectangular study area (i.e., the area inside the green box shown in Fig. 3) where both SMAP SM products and the moderate resolution imaging spectroradiometer (MODIS) products cover the area on all eight dates in 2020. The area is across the United States and Mexico ranging from 27°N to 33°N and 100°W to 108°W. The distribution of stations and the land cover map are shown in Fig. 4. The distribution of stations and the land cover map are shown in Fig. 4. The study area is mainly covered by different vegetation species including the shrublands, savannas, cropland, and the sparsely vegetated region. The open shrublands (in the middle in Fig. 4) are normally drier than the grassland (on the right-hand side in Fig. 4) and the woody

savannas (on the bottom-left in Fig. 4). Therefore, the SM value is lower in the middle part of the study area than side parts. So, the SM can be largely spatially variated and suitable for the SM study.

All data used in this study, including SMAP SM products at 9 km spatial resolution, the MODIS products, and the *in situ* data, are collected on a series of eight dates in 2020. The eight dates are January 25, February 25, March 13, April 14, May 16, September 5, October 7, and December 26. The objective is to downscale SMAP SM maps to the 1-km-resolution SM map. The MODIS products are utilized to provide more spatial texture information. The downscaling performance is evaluated by the *in situ* SM data.

B. SMAP Data

The SMAP mission is an L-band satellite incorporating both a radiometer and a radar dedicated to global SM measurements [1]. The daily SMAP level-3 SM passive product at 9 km (L3_SM_P_E) resolution acquired from National Snow

Fig. 5. Overall model architecture for SMAP SM downscaling. The network input M includes MODIS products NDVI, LST, and the interpolated 1 km SM from 9 km SMAP 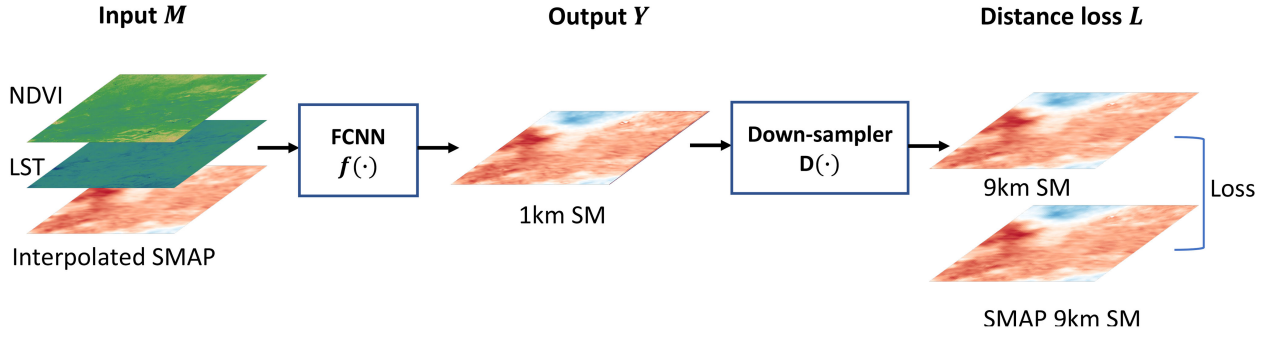SM. The output of the downscaling model is the downscaled SM Y. Unknown parameters β are network parameters of the FCNN, including weights and bias.

and Ice Data Center is downscaled to a 1-km SM map. Only the descending data acquired at 6:00 am are used.

1) MODIS Products: Normalized difference vegetation index (NDVI) and land surface temperature (LST) are physically related to SM [41] and commonly used for SM downscaling [1], [13], [14]. High-resolution auxiliary information, i.e., M in (9), MODIS products (MYD13A2 and MYD11A2) collected from the Land Processes Distributed Active Archive Center are utilized to downscale the SMAP SM products at 9 km resolution up to 1 km resolution. The MYD13A2 Version 6 product provides the NDVI and the enhanced vegetation index with a 1 km resolution. Only the NDVI layer from MYD13A2 is used in this study. The MYD11A2 Version 6 product provides an average 8-d per-pixel LST and emissivity with a 1 km spatial resolution. Only the first layer "LST_Day_1 km" from MYD11A2 is used.

2) In Situ Measurement: The international SM network hosts in-situ SM measurements collected starting 1952 to present from a total of 35 international SM networks. SM data from two networks (i.e., USCRN and SCAN) are used to evaluate the downscaling quality because the stations in these two networks are distributed more densely in the study area. There are four stations in the study area. The *in situ* SM observation measures the small point scale SM values and cannot be used directly in large-scale SM application, the shortcoming of which can be improved by RS-based SM mapping approaches. Considering that *in situ* SM measures are more accurate than RS SM products, here, we use these measures as ground truth to validate our downscaling results.

3) Data Preprocessing: For each time point, the MODIS NDVI and LST products, SMAP SM products, as well as the *in situ* measurements are prepared. The MODIS MYD13A2 and MYD11A2 products at 1 km resolution are downloaded and stitched together to achieve the global coverage for the further processing. The NDVI layer from MYD13A2, the LST layer from MYD11A2, and the SM layer from SMAP products layers are georeferenced and cropped by the longitude and latitude of the region of interest boundary. The image size of SMAP 9 km SM, NDVI, and LST, covering the study area are 74×86, 666×774, and 666×774, separately. The three-channel input of the network contains the 1 km NDVI, 1 km LST, and the 1 km interpolated SM, which is obtained from SM at 9 km using

a bilinear interpolation. Considering the fact that SMAP SM range between 0 and 1, we address the negative-valued outliers as positive values using a neighborhood refilling method, in which to remove the outliers in SMAP SM products, we refill the pixels using median values of their 3×3 neighboring pixels.

C. Model Implementation

A BDIP downscaling model is shown in Fig. 5, where the FCNN $f(\cdot)$ performs the inverse model, and the downsampler $D(\cdot)$ acts as the forward model. The inverse model will be trained, while the forward model is known and fixed. In this manner, the FCNN can achieve the downscaling purpose by learning from the forward model and inverting the downsampling operation. Then, the relationship between SM maps at 1 km (Y) and 9 km (X_9) can be expressed as follows:

$$X_9 = D_9(Y) \tag{11}$$

where $D_9(\cdot)$ is implemented using the "Lanczos" filtering with the downsampling factor 9.

The input M contains three layers, i.e., the NDVI from MYD13A2, LST from MYD11A2, and the interpolated 1 km SM map. Then, the final output of the network, which is the estimated 9 km SM, can be formulated as follows:

$$\hat{X}_9 = (D_9(f(M, \beta))). \tag{12}$$

We minimize the loss function as follows to train the FCNN:

$$L = \delta(X_9, \hat{X}_9). \tag{13}$$

Unknown parameters β are network parameters of the FCNN, including weights and bias. Once the model is trained, the intermediate output Y can be obtained as the downscaled SM.

1) Downsampler Selection: The downsampler works as the forward model mapping the LR SM to the HR SM. To find an appropriate forward model with the best capability of preserving the spatial information and the highest downscaling accuracy, we try the average pooling, max-pooling, and the downsampling with Lanczos kernel. Although it is claimed that no consistent differences are found among these downsampling methods for RGB images superresolution [19], it is critical to find out their performance of downscaling on the RS imagery products.
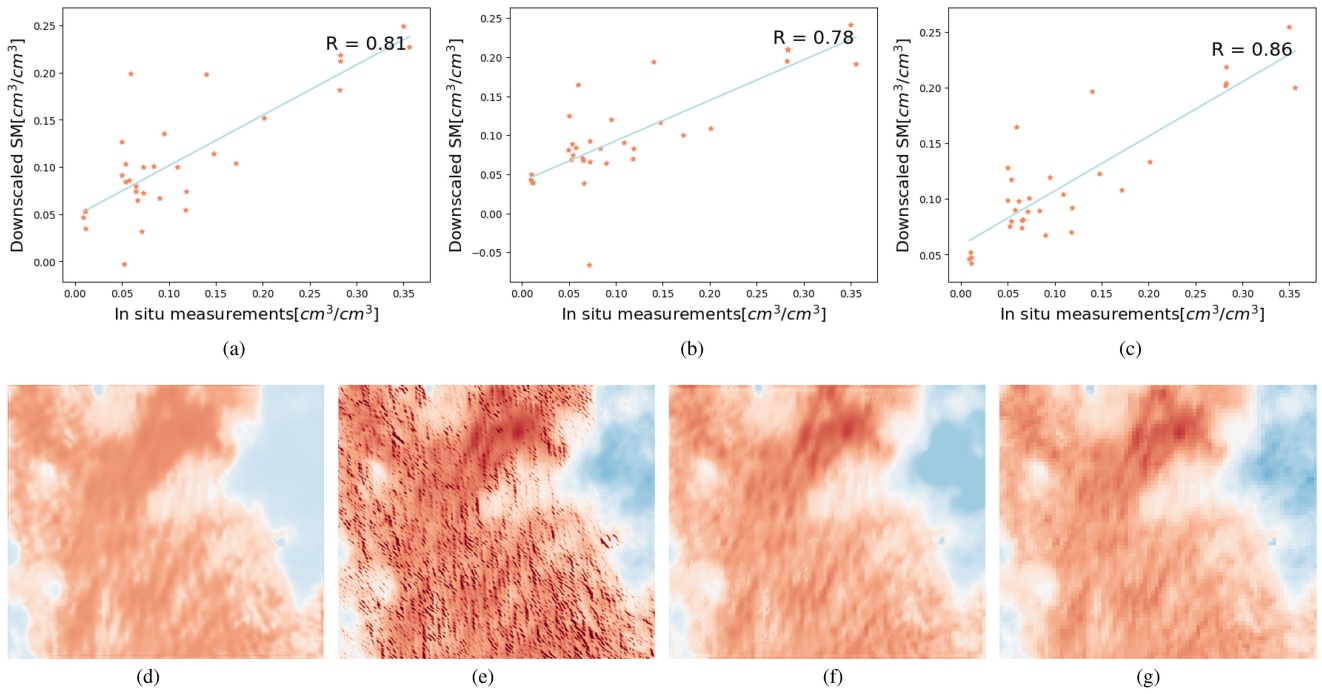
Fig. 6.   Scatters of downscaled 1 km SM against *in situ* SM measurements over eight dates and the downscaled 1 km SM maps on Jan 25th by models with different downsamplers. Using the downsampler with Lanczos kernel gives the sharpest SM map with the most spatial texture preserved. Average pooling smooths linear spatial feature and max pooling introduce fake spatial features. (a) Average pooling. (b) Max pooling. (c) Lanczos. (d) Average pooling. (e) Max pooling. (f) Lanczos. (g) SMAP 9 km.
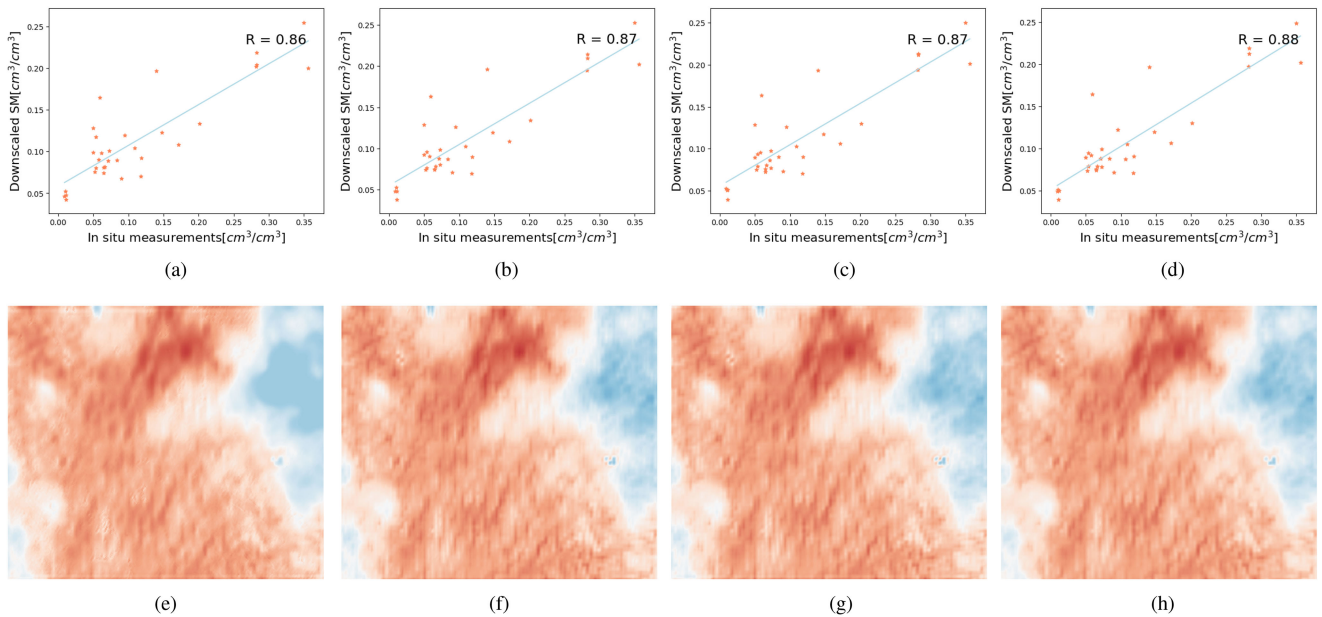
Fig. 7.   Downscaled 1 km SM maps by networks with different loss implementations. L2 loss tends to smooth the image. L1 loss can better accommodate high-frequency information. Although (f) is sharper than (e), the R value does not increase. Then, the SSIM loss using for preserving image structural feature and the perceptual loss using for extracting spatial information from the feature domains are added one-by-one. As a result, (h) shows the richest spatial information and its corresponding R value achieved 0.88. (a) L2. (b) L1. (c) L1+SSIM. (d) L1+SSIM+perceptual. (e) L2. (f) L1. (g) L1+SSIM. (h) L1+SSIM+perceptual.
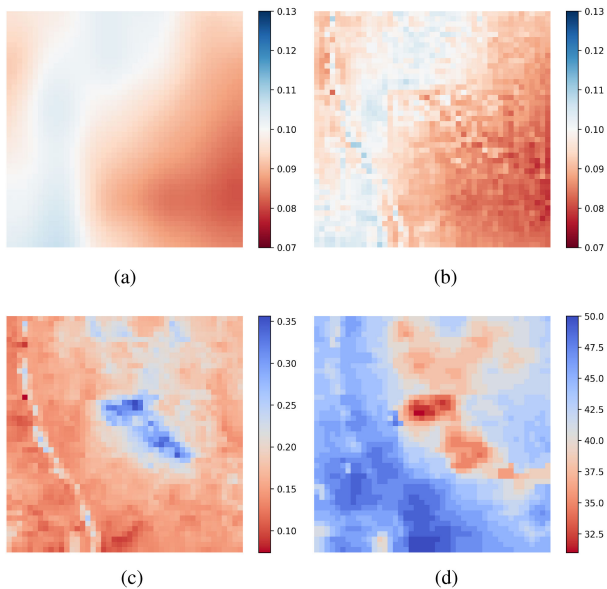
Fig. 8. Downscaled 1 km SM maps (unit: cm^3/cm^3) by networks (a) with and (b) without the additional skip connection (indicated by the purple line in Fig. 2). (b) Fuses spatial information from the (c) NDVI and (d) LST better than (a). (a) without skip. (b) with skip. (c) NDVI. (d) LST.

TABLE I
PARAMETERS CONFIGURATION FOR DIFFERENT MODELS

Downsampler	Loss	Learning rate	Epoch
Average	L2	0.003	3000
Max	L2	0.003	3000
Lanczos	L2	0.003	3000
Lanczos	L1	0.003	3000
Lanczos	0.1×L1+1×SSIM	0.001	3000
Lanczos	0.1×L1+1×SSIM+0.5×perceptual	0.001	3000

*2) Loss Function Design:* The reconstruction loss is initialized as an L2 loss, which is commonly used in image reconstruction tasks [19]. However, given its performance of blurring some detailed spatial information, L1 loss, L1 loss combined SSIM loss [42], as well as the combination of L1 loss, SSIM loss, and perceptual loss [43] are tested to better reconstruct the structural spatial texture in HR SM maps.

3) Skip Connections: Skip connections in FCNNs solves the degradation problem and ensures the feature reusability by copying and concatenating features from shallower layers to deeper layers. To better preserve the spatial feature in the input data, besides the skip connections existing in a classic U-Net architecture, we add a skip connection (indicated by the purple line in Fig. 2) by concatenating the output feature of the input layer to the output feature of the last second layer (i.e, the last 3×3 convolution layer indicated by the yellow block in Fig. 2).

4) Parameters Configuration: The learning rate and training epochs for different models tested in Sections V-C1 and V-C2 are listed in Table I.

TABLE II
METHODS ASSESSMENT

Method	MODIS used?	Performance		
		Variation consistent?	Spatial texture increased?	SM value range correct?
BDIP	Yes	Yes	Yes	Yes
Bicubic	No	Yes	Yes	No
GFPCA [44]	Yes	Yes	No	No
PCA [45]	Yes	No	Yes	No
CNMF [46]	Yes	No	Yes	No

D. Methods Comparison

The compared methods include Bicubic, GFPCA [44], PCA [45], and CNMF [46]. Bicubic is a standard interpolation approach based on the cubic interpolation. GFPCA is designed for the fusion of hyperspectral and RGB image based on PCA [44]. PCA, as a standard data transformation method, has been used for RS data pansharpening [45], [47]. CNMF is developed based on nonnegative matrix factorization unmixing and applied to hyperspectral and multispectral data fusion and downscaling [46], [48], [49].

The compared methods are conducted using the downscaling toolbox from Github. The source code is available at https://github.com/codegaj/py_pansharpening. These methods all require two sets of inputs, which are the high-resolution channels and the low-resolution channels. The summation of the NDVI and LST is used for the high-resolution input. The low-resolution channel is the 9 km SMAP SM.

E. Evaluation Strategy

Following the commonly used evaluation scheme for downscaling algorithms, the downscaled SM map is evaluated from three aspects, i.e.,
1) the consistency of the spatial variation pattern with the SMAP SM maps [1], [14],
2) the numerical accuracy of the SM values to in-situ SM measurements [1], [13], [14], and
3) the amount of the spatial textural information compared with the SMAP SM map [1], [13].

Downscaling results are evaluated in both visual and numerical ways. For the visual evaluation, the downscaled SM maps will be presented together with the SMAP SM maps at 9 km spatial resolution, as well as the estimation by the other four downscaling methods. For the numerical evaluation, the *in situ* groundtruth measurements on eight dates are used as the reference. The classical statistical metrics are calculated to represent the error scores, including correlation coefficient (R), mean square error (mse), the difference of the mean values (BIAS), root-mean-square error (RMSE), normalized root-mean-square error (nrRMSE), and unbiased root-mean-square error (ubRMSE).

VI. RESULT AND DISCUSSION

1) Downsampler Selection: The downscaling results obtained by different downsampler are shown in Fig. 6. Using the downsampler with Lanczos kernel gives the sharpest SM
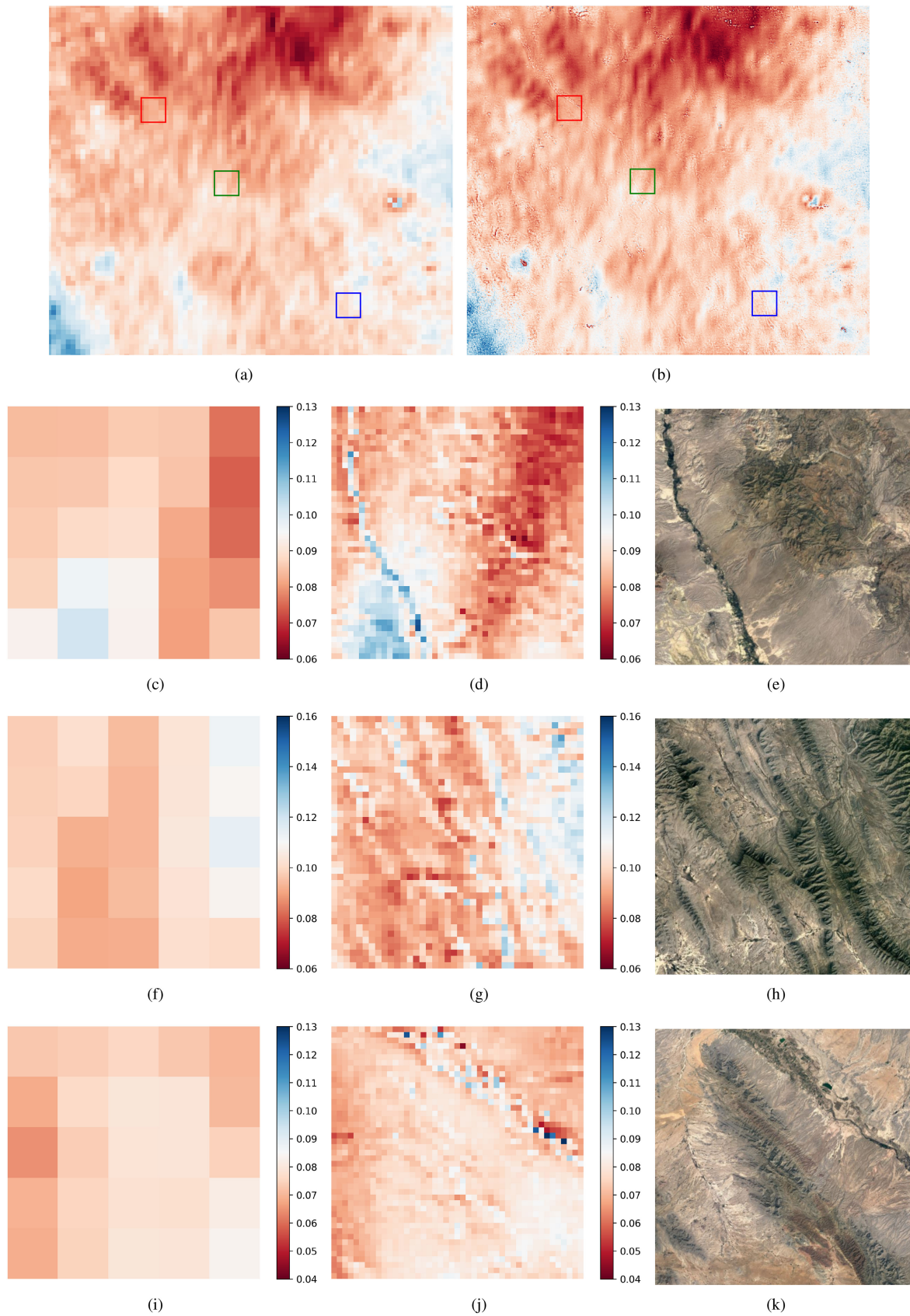
Fig. 9. Comparison between the 9 km SMAP SM maps and downscaled 1 km SM maps (unit: cm³/cm³) on Dec 26th. (a) and (b) are separately the 9 km SMAP and the 1 km downscaled SM. (c) and (d) are separately the zoomed region indicated with the green box in (a) and (b). (e) is the corresponding area clipped from the Google Earth. (f), (g) correspond to the blue box in (a) and (b). (i), (k) correspond to the red box in (a) and (b). The downscaled SM map shows not only the consistent variation pattern with the 9 km SMAP SM, but also much more spatial detail information. For example, the green linear region in (e) is the cropland with higher water content, which is indicated by the blue linear feature in (d).

Fig. 10. Scatters of the 1 km SM estimated by the different method and the *in situ* groundtruth over eight dates with R values. Two networks are separately calculated. The first row is USCRN network and the second row is the SCAN network. The last row is the result obtained by all stations from two networks. (a) BDIP. (b) Bicubic. (c) PCA. (d) GFPCA. (e) CNMF.

Fig. 11. SMAP SM map at 9 km resolution in column (a), the downscaled 1 km SM maps by the different methods from column (b) to (f), and the input NDVI map in column 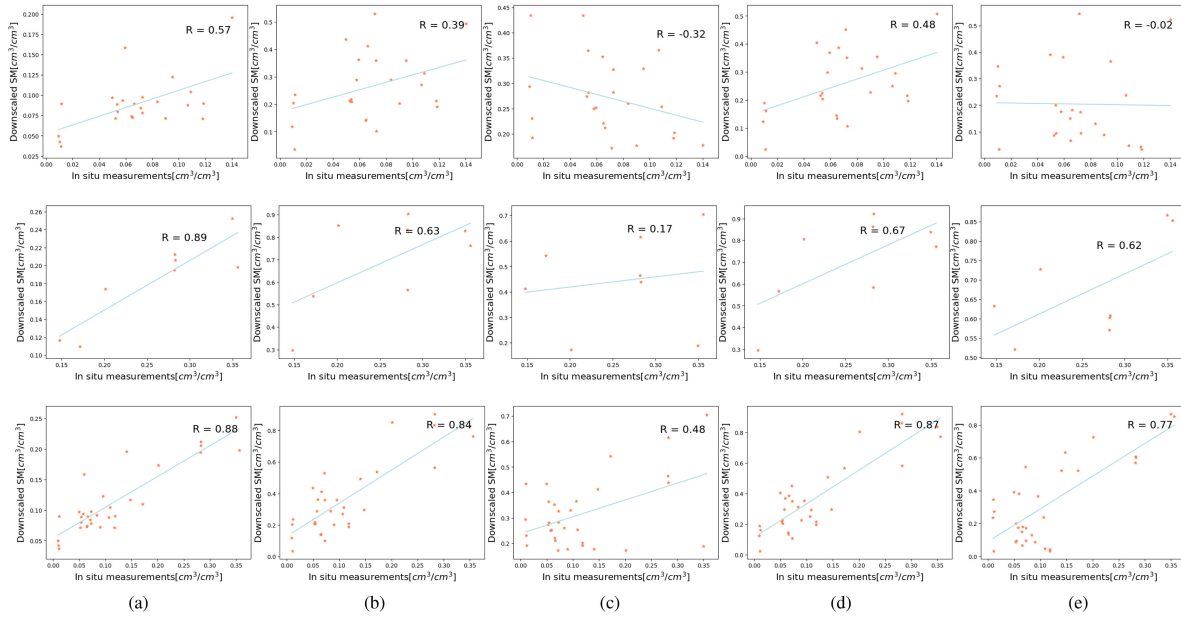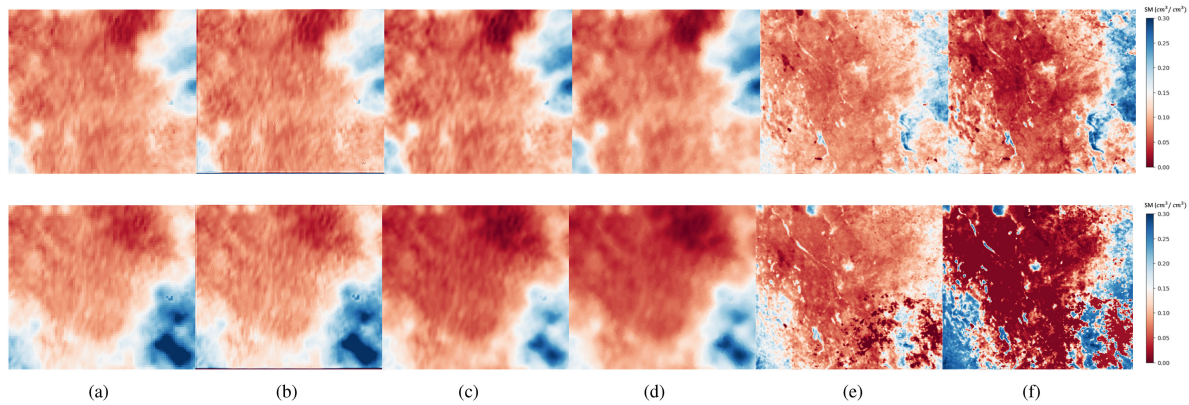(g) at April 14 (first row) and September 5 (second row). The proposed method estimates the high-resolution SM map with sharp and clear boundaries. SM maps generated by the Bicubic and GFPCA share consistent variation patterns with SMAP SM maps, but with large bias. GFPCA SM maps get more blurred than 9 km SMAP SM. SM maps achieved by PCA and CNMF preserve much information in the NDVI and LST than the SMAP SM map, which is fake SM texture. They are not able to properly extract and balance the spatial feature information from the SMAP SM and MODIS products. (a) SMAP. (b) BDIP. (c) Bicubic. (d) GFPCA. (e) PCA. (f) CNMF.

map with the most spatial texture preserved. R value over the time series corresponding to Lanczos is also the highest. The average pooling smooths some linear spatial features and the max pooling brings fake spatial features. The results indicate that using different downsamplers significantly affects the downscaling performance.

2) Performance of Loss Functions: Once the downsampler with Lancnos kernel is selected, we fix the downsampler and change the loss function. The 1 km downscaled SM map obtained with L2 loss is blurred, especially on the right-hand side of Fig. 7(a), although the R value is high. Given that L2 loss is sensitive to high-frequency signals and tends to smooth the

image, we tried L1 loss instead, which can better accommodate high-frequency information. As a result, Fig. 7(f) is sharper than Fig. 7(e) and shows more spatial texture, with R value increasing. Then, the SSIM loss using for preserving image structural feature and the perceptual loss using for extracting spatial information from feature domains are added one-by-one. As a result, Fig. 7(h) shows the richest spatial information, and its corresponding R value achieved 0.88. The results indicate the importance of designing loss functions for downscaling visual performance. Although the R value does not highly increased, the spatial information shown in the downscaled SM map gets sharper and richer.

TABLE III
RMSE(CM^6/CM^6), BIAS(CM^3/CM^3), RMSE(CM^3/CM^3), NRRMSE(CM^3/CM^3), AND UBRMSE(CM^3/CM^3) OF THE VALIDATION FOR THE 1 KM DOWNSCALED SM WITH THE MEASUREMENT OF *IN SITU* STATIONS FROM TWO NETWORKS

		R	MSE	BIAS	RMSE	nrRMSE	ubRMSE
PCA	All(4 stats)	0.4752	0.0541	-0.1988	0.2327	0.3341	0.1210
	SCAN (1 stat)	0.1653	0.0659	-0.1837	0.2568	0.4600	0.1795
	USCRN (3 stats)	-0.3154	0.0502	-0.2038	0.2241	0.3638	0.5267
GFPCA	All	0.8713	0.0909	-0.2529	0.3013	0.3302	0.1637
	SCAN	0.6685	0.2243	-0.4472	0.4737	0.6119	0.1562
	USCRN	0.4750	0.0462	-0.1882	0.2151	0.4315	0.1042
CNMF	All	0.7659	0.0765	-0.2054	0.2765	0.3225	0.1851
	SCAN	0.6238	0.1803	-0.4140	0.4246	0.5904	0.0943
	USCRN	-0.0186	0.0419	-0.1359	0.2046	0.3817	0.1530
Bicubic	All	0.8427	0.0928	-0.2563	0.3047	0.3406	0.1649
	SCAN	0.6318	0.2182	-0.4386	0.4671	0.6176	0.1607
	USCRN	0.3871	0.0511	-0.1955	0.2260	0.4345	0.1134
Ours	All	**0.8823**	**0.0028**	**0.0034**	**0.0537**	**0.1551**	**0.0536**
	SCAN	**0.8917**	**0.0072**	**0.0761**	**0.0853**	**0.3464**	**0.0384**
	USCRN	**0.5683**	**0.0014**	**-0.0208**	**0.0378**	**0.2023**	**0.0315**

3) Performance of Skip Connection: Downscaling performances are compared between the U-Net with and without the additional skip connection. The result is shown in Fig. 8. The downscaling result with the skip connection [see Fig. 8(b)] shows much richer spatial information than that without the skip connection [see Fig. 8(a)]. So, the skip connection in the U-Net can better preserve the low-level feature in NDVI and LST.

4) Spatial Detail Restoration: To check the downscaling effectiveness of the proposed method, the downscaled 1 km SM maps and the 9 km SMAP SM maps are zoomed in different scales, as shown in Fig. 9. The downscaled SM map shows not only the consistent variation pattern with the 9 km SMAP SM, but also much more spatial detail information, which is consistent with satellite RGB images. For example, the green linear region in Fig. 9(e) is the cropland with higher water content, which is indicated by the blue linear feature in Fig. 9(d).

5) Methods Comparison: Table III lists R values, BIAS, and RMSE values between the 1 km downscaled SM map and the *in situ* groundtruth over eight dates. In total, 32 points (i.e., 4 stations × 8 d) in total are used to calculate metrics.

The validation scatterplots over eight dates shown in Fig. 10 show consistent correlation degrees with Table III. Scatters of the proposed BDIP method show an obvious linear relationship between the downscaled SM and groundtruth data. By observing the scatters, we found that the measurements within the USCRN network are generally smaller than the SCAN network because the USCRN stations distributed in the shrublands, and the SCAN station is in the grassland, where the soil normally contains more water. It is found that the score of SCAN network is generally better than USCRN network. The possible reasons are listed as follows.

1) The better statistical score could be caused by fewer station points.
2) The SM is overall higher at SCAN station than that at USCRN stations.
3) Sensors of these two networks could be different.

Fig. 11 displays the downscaled SM maps by different methods on two dates. SM maps generated by the Bicubic and GFPCA share consistent variation patterns with SMAP SM maps, but with large bias. GFPCA SM maps get more blurred than 9 km SMAP SM. The proposed method, on the contrary, estimates the high-resolution SM map with sharp and clear boundaries. SM maps achieved by PCA and CNMF preserve much information in the NDVI and LST than the SMAP SM map, which is the fake SM texture. They fail to properly extract and balance the spatial feature information from the SMAP SM and MODIS products. The abovementioned results' description is listed in Table II.

The PCA and CNMF methods were designed for multispectral, hyperspectral images pansharpening, where the HR images and the LR images share the similar spatial texture. PCA and CNMF are also used to enhance the contrast of the original image. However, for the SM downscaling guided by NDVI and LST, the HR NDVI, LST, and the LR SMAP SM have different spatial texture. So, simply extracting the spatial textural information from all bands leads to the failure of data fusion. GFPCA performs better than PCA and CNMF because a transformation from NDVI and LST to the SMAP SM was conducted instead of extracting information from all of MODIS and SMAP products. However, the downscaled SM still gets blurred, which could be caused by the transformation or upsampling procedure. Bicubic interpolates the SMAP SM directly without using MODIS products, leading to insufficient spatial details.

To sum up, the downscaled 1 km SM by our proposed method not only has the consistent variation pattern with the SMAP maps, but also restores more spatial details than other methods with higher accuracy.

VII. CONCLUSION

In this article, we proposed a Bayesian DIP downscaling model for SMAP SM products by integrating the FCNN into a Bayesian framework. MODIS products was used as the model input to guide the downscaling procedure. An hourglass FCNN was adopted to map the nonlinear relationship between MODIS products and high-resolution SM map and to better construct the spatial heterogeneous information in SM map. The MAP inverse problem was solved by back propagation instead of EM iterations, which makes the model optimization simpler and faster. Experiments on the time series data showed that SM maps estimated by the proposed method provided more spatial texture details than other existing unsupervised downscaling methods, and the estimated SM was very close to *in situ* measurements

with a high overall R value 0.88. The proposed Bayesian downscaling model is very effective for SMAP SM downscaling.

Despite of the successful of the BDIP downscaling approach based on the *in situ* and visual validation, this article has a shortcoming of insufficient result analysis from the geographical perspective, such as how the downscaled SM map correlates in finescale with the land cover types, precipitation, and the elevation. Another shortcoming of this article is the insufficient comparison with more advanced downscaling methods considering that the unsupervised downscaling approach is limited. However, this unsupervised approach has larger potential than supervised ones to be widely used without high-resolution maps required. Moreover, since the proposed model is flexible to fuses multisource RS products and its downsampler part can be adjust according to the resolution of existing SM products, it has the big potential to be applied to more SM products with different spatial resolutions and to fuse more RS products, such as precipitation and terrain products.

REFERENCES

[1] Z. Wei, Y. Meng, W. Zhang, J. Peng, and L. Meng, "Downscaling SMAP soil moisture estimation with gradient boosting decision tree regression over the tibetan plateau," *Remote Sens. Environ.*, vol. 225, pp. 30–44, 2019.

[2] J. Peng and A. Loew, "Recent advances in soil moisture estimation from remote sensing," *Water*, vol. 9, no. 7, 2017, Art. no. 530.

[3] Y. Jin, Y. Ge, J. Wang, Y. Chen, G. B. M. Heuvelink, and P. M. Atkinson, "Downscaling AMSR-2 soil moisture data with geographically weighted area-to-area regression Kriging," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2362–2376, Apr. 2018.

[4] S. H. Alemohammad, J. Kolassa, C. Prigent, F. Aires, and P. Gentine, "Global downscaling of remotely sensed soil moisture using neural networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 10, pp. 5341–5356, 2018.

[5] S. Sabaghy *et al.*, "Comprehensive analysis of alternative downscaled soil moisture products," *Remote Sens. Environ.*, vol. 239, 2020, Art. no. 111586.

[6] E. G. Njoku *et al.*, "Observations of soil moisture using a passive and active low-frequency microwave airborne sensor during SGP99," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 12, pp. 2659–2673, Dec. 2002.

[7] Y. H. Kaheil, M. K. Gill, M. McKee, L. A. Bastidas, and E. Rosero, "Downscaling and assimilation of surface soil moisture using ground truth measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1375–1384, May 2008.

[8] U. Narayan and V. Lakshmi, "Characterizing subpixel variability of low resolution radiometer derived soil moisture using high resolution radar data," *Water Resour. Res.*, vol. 44, no. 6, 2008, Art. no. W06425.

[9] J. Pellenq *et al.*, "A disaggregation scheme for soil moisture based on topography and soil depth," *J. Hydrol.*, vol. 276, no. 1–4, pp. 112–127, 2003.

[10] G. Kim and A. P. Barros, "Downscaling of remotely sensed soil moisture with a modified fractal interpolation method using contraction mapping and ancillary data," *Remote Sens. Environ.*, vol. 83, no. 3, pp. 400–413, 2002.

[11] J. Rhee, J. Im, and G. J. Carbone, "Monitoring agricultural drought for arid and humid regions using multi-sensor remote sensing data," *Remote Sens. Environ.*, vol. 114, no.12, pp. 2875–2887, 2010.

[12] P. K. Srivastava, D. Han, M. R. Ramirez, and T. Islam, "Machine learning techniques for downscaling SMOS satellite soil moisture using MODIS land surface temperature for hydrological application," *Water Resour. Manage.*, vol. 27, no. 8, pp. 3127–3144, 2013.

[13] W. Xu, Z. Zhang, Z. Long, and Q. Qin, "Downscaling SMAP soil moisture products with convolutional neural network," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4051–4062, 2021, doi: 10.1109/JS-TARS.2021.3069774.

[14] F. Wen, W. Zhao, Q. Wang, and N. Sánchez, "A value-consistent method for downscaling SMAP passive soil moisture with MODIS products using self-adaptive window," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 913–924, Feb. 2020.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[16] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, 2018.

[17] Y. Li, J. Hu, X. Zhao, W. Xie, and J. J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, 2017.

[18] J. Zhang, J. Pan, W. S. Lai, R. Lau, and M. H. Yang., "Learning fully convolutional networks for iterative non-blind deconvolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6969–6977.

[19] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 9446–9454.

[20] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.

[21] S. Ozkan and G. B. Akar, "Deep spectral convolution network for hyperspectral unmixing," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 3313–3317.

[22] K. Vittala Shettigara, "A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set," *Photogramm. Eng. Remote Sens.*, vol. 58, no. 5, pp. 561–567, 1992.

[23] B. Fang, V. Lakshmi, R. Bindlish, T. J. Jackson, and P.-W. Liu, "Evaluation and validation of a high spatial resolution satellite soil moisture product over the continental United States," *J. Hydrol.*, vol. 588, 2020, Art. no 125043.

[24] K. Gong, C. Catana, J. Qi, and Q. Li, "Pet image reconstruction using deep image prior," *IEEE Trans. Med. Imag.*, vol. 38, no. 7, pp. 1655–1665, Jul. 2019.

[25] E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, "Deep phase decoder: Self-calibrating phase microscopy with an untrained deep neural network," *Optica*, vol. 7, no. 6, pp. 559–562, 2020.

[26] K. C. Zhou and R. Horstmeyer, "Diffraction tomography with a deep image prior," *Opt. Exp.*, vol. 28, no. 9, pp. 12872–12896, 2020.

[27] Y. Fang, Y. Wang, L. Xu, R. Zhuo, A. Wong, and D. A Clausi, "BCUN: Bayesian fully convolutional neural network for hyperspectral spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5523714.

[28] X. Ma, Y. Hong, and Y. Song, "Super resolution land cover mapping of hyperspectral images using the deep image prior-based approach," *Int. J. Remote Sens.*, vol. 41, no. 7, pp. 2818–2834, 2020.

[29] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 2124–2132.

[30] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 524–533.

[31] H. Wang and D.-Y. Yeung, "A. survey on Bayesian deep learning," *ACM Comput. Surv.*, vol. 53, no. 5, pp. 1–37, 2020.

[32] Y. Fang, L. Xu, J. Peng, H. Yang, A. Wong, and D. A Clausi, "Unsupervised Bayesian classification of a hyperspectral image based on the spectral mixture model and Markov random field," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3325–3337, Sep. 2018.

[33] T. Wang, F. Fang, F. Li, and G. Zhang, "High-quality Bayesian pansharpening," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 227–239, Jan. 2019.

[34] Y. Bai, W. Chen, J. Chen, and W. Guo, "Deep learning methods for solving linear inverse problems: Research directions and paradigms," *Signal Process.*, vol. 177, no. 2, 2020, Art. no. 107729.

[35] P. Diederik Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[36] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[37] H. Wang, X. Shi, and D.-Y. Yeung, "Relational stacked denoising autoencoder for tag recommendation," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, vol. 29, pp. 3052–3058.

[38] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382. Hoboken, NJ, USA: Wiley, 2007.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.- Assist. Interv.*, 2015, pp. 234–241.

[40] P. Diederik Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[41] A. Abdalla Hassaballa, A. Nasir Matori, and H. Z. M. Shafri, "Surface moisture content retrieval from visible/thermal infrared images and field measurements," *Caspian J. Appl. Sci. Res.*, vol. 2(AICCE'12 & GIZ' 12), pp. 182–189, 2013.

[42] 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.

[43] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[44] W. Liao, "Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter," in *Proc. 7th Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2015, pp. 1–4.

[45] P. Kwarteng and A. Chavez, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 1, pp. 339–348, 1989.

[46] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.

[47] G. Vivone *et al.*, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[48] N. Yokoya, T. Yairi, and A. Iwasaki, "Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization," in *Proc. 3rd Workshop Hyperspectral image Signal Process.: Evol. Remote Sens.*, 2011, pp. 1–4.

[49] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of EO-1/Hyperion and Terra/ASTER," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 419–426, Apr. 2013.

Yuan Fang received the B.Eng. and M.Sc. degrees in geomatics engineering from the China University of Geosciences, Beijing, China, in 2016 and 2018, respectively. She is currently working toward the Ph.D. degree in remote sensing image processing with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada.

Her research interests include remote sensing image processing, hyperspectral imaging, machine learning, and environmental monitoring.

Linlin Xu (Member, IEEE) received the B.Eng. and M.Sc. degrees in geomatics engineering from the China University of Geosciences, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree in geography from the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON, Canada, in 2014.

He is currently a Research Assistant Professor with the Department of Systems Design Engineering, University of Waterloo. He has authored or coauthored various papers on high-impact remote sensing journals and conferences. His research interests include hyperspectral and synthetic aperture radar data processing and their applications in various environmental applications.

Yuhao Chen (Member, IEEE) received the B.S. and Ph.D. degrees in electrical and computer engineering from Purdue University, West Lafayette, IN, USA in 2015, 2019, respectively.

He was a Member of the Video and Image Processing Laboratory, Purdue University. In 2020, he joined as a Postdoctoral Fellow of systems design engineering with Vision and Image Processing Lab, University of Waterloo, Waterloo, ON, Canada. His current research interests include computer vision, machine learning, deep learning, data representation, and image processing.

Wei Zhou received the Ph.D. degree in computer vision from the University of Science and Technology of China, Hefei, China, in 2021 (joint with the University of Waterloo, Waterloo, ON, Canada from 2019 to 2021)

He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo. He was a Visiting Scholar with the National Institute of Informatics, Tokyo, Japan in 2017, a Research Assistant with Intel, Santa Clara, CA, USA from 2016 to 2018, and a Research Intern with Microsoft Research, Redmond, WA, USA and Alibaba Cloud, Hangzhou, China, in 2018 and 2019, respectively.

Alexander Wong (Senior Member, IEEE) received the B.A.Sc. degree in computer engineering, the M.A.Sc. degree in electrical and computer engineering, and the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 2005, 2007, and 2010, respectively.

He is currently the Canada Research Chair of artificial intelligence and medical imaging, the Co-Director of the Vision and Image Processing Research Group, a Professor with the Department of Systems Design Engineering, University of Waterloo, and a Member of the College of the Royal Society of Canada, Ottawa, ON, Canada. He has authored more than 560 refereed journal and conference papers and patents, in various fields, such as computational imaging, artificial intelligence, computer vision, graphics, image processing, and multimedia systems. His research interests include integrative biomedical imaging systems design, operational artificial intelligence, and scalable and explainable deep learning.

Dr. Wong was the recipient of a number of awards, including two Outstanding Performance Awards, the Distinguished Performance Award, Engineering Research Excellence Award, Sandford Fleming Teaching Excellence Award, Early Researcher Award from the Ministry of Economic Development and Innovation, Outstanding Paper Award at the CVPR Workshop on Adversarial Machine Learning in Real-World Computer Vision Systems and Online Challenges, in 2021, Best Paper Award at the NIPS Workshop on Transparent and Interpretable Machine Learning, in 2017, Best Paper Award at the NIPS Workshop on Efficient Methods for Deep Neural Networks, in 2016, two Best Paper Awards by the Canadian Image Processing and Pattern Recognition Society, in 2009 and 2014, respectively, Distinguished Paper Award by the Society of Information Display, in 2015, two Best Paper Awards for the Conference of Computer Vision and Imaging Systems, in 2015 and 2017, respectively, Synaptive Best Medical Imaging Paper Award, in 2016, two Magna Cum Laude Awards and one *Cum Laude* Award from the Annual Meeting of the Imaging Network of Ontario, AquaHacking Challenge First Prize, in 2017, Best Student Paper at Ottawa Hockey Analytics Conference, in 2017, and Alumni Gold Medal.

David A. Clausi (Senior Member, IEEE) received the Ph.D. degree in systems design engineering from the University of Waterloo, Waterloo, ON, Canada, in 1996.

He was with Mitra Imaging Inc., Waterloo, ON, USA, in medical imaging. He was an Assistant Professor of geomatics engineering with the University of Calgary, Calgary, AB, Canada, in 1997. In 1999, he returned to the University of Waterloo, where he is currently a Professor who specializes in the field of intelligent systems and the Associate Dean Research and External Partnerships with the Faculty of Engineering. He has numerous contributions, publishing refereed journal articles and conference papers in the areas of his research interests which include remote sensing, computer vision, image processing, and algorithm design.

Dr. Clausi was the recipient of many scholarships, paper awards, research awards, and teaching excellence awards, and his efforts have led to successful commercial implementations, culminating in the creation and sale of a high-tech company. From 2004 to 2006, he was the Co-Chair of IAPR Technical Committee 7-Remote Sensing.