Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Deep feature learning and latent space encoding for crop phenology analysis

Arun Pattathal V [a,*], Arnon Karnieli [b,*]

[a] Swiss Institute for Dryland Environmental and Energy Research, Jacob Blaustein Institutes for Desert Research, Ben Gurion University of the Negev, Sede Boker Campus, 8499000, Israel

[b] The Remote Sensing Laboratory, French Associates Institute for Agriculture and Biotechnology of Dryland, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boker Campus, 8499000, Israel

ARTICLE INFO

Keywords:
Crop phenology
Smoothing
Vegetation index
VENμS
Crop fingerprint
Classification
Vegetation index curve generalization

ABSTRACT

The high spatial, spectral, and temporal resolutions of the Vegetation and Environment monitoring New Micro-Satellite (VENμS) satellite data facilitate field-level phenological analysis of crops. This study proposes deep learning (DL) based approaches to resolve the issues prevalent in crop phenology-based fingerprint estimation at field-level using VENμS satellite data. An encoder-decoder-based framework, called piece-wise kernel encoding network (PKNet), is proposed for missing data imputation of the vegetation index (VI) curves derived from time-series image data. PKNet adopts interpolation-based convolution, dynamic time wrapping (DTW) based layer formulation, and imputation-specific constraints for optimal smoothing of the irregularly sampled VI curves. Besides, PKNet learns kernel parameters dynamically. A variational encoding framework called a dynamic-projection-based generalization network (DPGNet), is proposed to generalize the pixel-level VI curves to synthesize a representative VI curve for a given field. DPGNet is more effective than the use of multiple moments as it is resilient to outliers and learns normally distributed latent space with a small number of samples. The current research also proposes a classifier, called dynamic time wrapping based capsule network (DTCapsNet), which learns a discriminative latent space and accurately models the VI curve features. The DTCapsNet considers the time-series nature of the input using DTW-based convolution layers. The feature characterization improves generalizability and gives good results, even with a limited number of training samples. Experiments using the ground truth information and satellite images, acquired over two farms in Israel, illustrate that the proposed frameworks give better results than the commonly-used existing approaches.

1. Introduction

In the agriculture domain, Earth observation (EO) data is widely explored for detailed and timely information on crop type, condition, production, and yield for several applications (Gebbers and Adamchuk 2010). Forecast of crop yield and production requires information about the surface coverage of different crops. Automatic mapping using EO imagery is the most popular and convenient approach for sampling large areas (Foerster et al. 2012; Kamir et al. 2020). EO data-based phenological analyses use vegetation index (VI) curves (e.g., normalized differential vegetation index (NDVI)), estimated from temporal images over a given season, as phenology fingerprints to model seasonal variations of different crops (Xiang et al. 2019).

Deep learning (DL) approaches, which transform inputs to intrinsic manifolds in an unsupervised manner, have reported better results than the conventional machine learning approaches for various EO data applications (Zhu et al. 2017; Ma et al. 2019; Patterson et al. 2020). DL is powerful for mining features or relationships from data as it extracts hierarchical information from vast volumes of data (Chai et al. 2020). Convolutional neural network (CNN) based deep architectures strive to learn abstract feature representations for a given set of data, and the use of shared weights resolves the issue of overfitting. The current study investigates DL approaches for addressing different issues prevalent in the phenological analysis of crops. Although the proposed approaches adopt deep convolutional architectures, the loss functions, architectures, regularizations, optimization strategies, and even the type and strategy of convolutions vary significantly in accordance with the specific issue being resolved.

* Corresponding authors.
E-mail addresses: arunvi2601@gmail.com (A. Pattathal V), karnieli@bgu.ac.il (A. Karnieli).

https://doi.org/10.1016/j.eswa.2021.115929
Received 27 December 2020; Received in revised form 25 July 2021; Accepted 16 September 2021
Available online 24 September 2021

A major issue in the EO data-based phenological analysis of crops is the effect of clouds, aerosols, and sensor malfunctions resulting in irregular VI curves (Richardson et al. 2012). Most of the existing denoising approaches adopt window-based regression techniques or static-kernel-based convolution approaches requiring the parameters to be manually fine-tuned (Cao et al. 2018; Yan and Roy 2020). Also, the used approaches are sensitive to the cross-domain and cross-modality shifts (Chen et al. 2004). This study explores DL-based unsupervised feature learning to smoothen the VI curves in a learned manifold using dynamically learned kernels.

Another issue in the phenological analysis where the application of DL can have significant improvement is the generation of field-level representation of VI curves. The field-level crop phenology monitoring requires multiple VI curves obtained at the pixel level to be transformed to field level. Usually, statistical moments such as mean, mode, median, covariance, and standard deviation are used. However, the effect of outliers and increased computational complexity in considering multiple moments affect the practical utility of the approach. Also, considering specific features or trends in the curves, rather than values alone, is significant in achieving proper generalization (Zhao et al. 2019). The current study employs DL-based approaches to project the pixel-level VI curves to a normally distributed latent space. Although different available encoding–decoding approaches are applicable for reconstructing VI curves, the specific nature and irregular sampling of the VI curves need to be considered to get an effective characteristic representation. The field-level VI curve generation proposed in this study can also be employed to generate the phenology fingerprint (of a given crop) from the representative VI curves of multiple fields.

Comparing the phenology fingerprint of a crop with a group of VI curves is crucial in identifying the crop type or anomalies (caused by different disturbances) and thereby infer certain conditions, such as pest or disease infection, water stress, desertification, etc. The substantial shift in the phenological curves of a single crop and the inter-seasonal variations affect the use of simple Euclidean distance measures in estimating the similarity between VI curves and baseline diagnostic ones. In this regard, classifiers designed to compare the phenology using temporal VI curves need to adopt relaxed one-to-many matching strategies. The dynamic time wrapping (DTW) based similarity measures optimize the alignment of two time series by establishing one-to-many associations between data points and are tolerant to both shifts and distortions. Although these approaches account for the relative importance regarding the phase difference between reference and testing points, the weights are usually pre-defined based on prior knowledge rather than learned (Baumann et al. 2017). The current study proposes a DL-based approach that can model the features of VI curves accurately and can incorporate learnable DTW-based similarity computations in layers.

In summary, the current study strives to investigate the use of advanced DL approaches for resolving the issues concerning within-season field-level monitoring of crop phenology. The main contributions of this study are (1) development of a DL-based approach, called piece-wise kernel encoding network (PKNet), for effective smoothing of VI curves; (2) development of an approach, called dynamic-projection based generalization network (DPGNet), to derive generalized field-level representation from pixel-level representations; and (3) development of a VI curve classifier, called dynamic time wrapping based capsule network (DTCapsNet), to effectively model the features of VI curves taking in to account the shifts, distortions and scale differences prevalent in them. The performance of the above-listed proposed approaches is compared with some benchmark approaches tested on the same domain.

2. Related studies

This section reviews the main commonly-used approaches related to the different issues prevalent in crop phenology-based analysis. Subsection 2.1 presents different prominent approaches that apply to the smoothing of VI curves. Different encoding-based reconstruction techniques, which can be used to generalize a given set of VI curves, are reviewed in Subsection 2.2. Finally, different similarity measures and classifiers relevant in estimating the crop labels are presented in Subsection 2.3. The specific contributions of this research and the novelty of each of the proposed frameworks concerning the corresponding existing approaches are discussed in each of these subsections.

2.1. Machine learning and deep learning approaches applicable for smoothing/denoising

Different recent machine learning approaches applicable for denoising or smoothing vegetation VI curves are reviewed in Cai et al. (2017) and Zeng et al. (2020). Most conventional smoothing approaches adopt local regression for smoothing or filling gaps of a given series data, such as VI curves. Weil et al. (2017) employed a locally weighted scatterplot smoothing function for data imputation with noisy phenological patterns. The fitting of nonlinear least squares to the asymmetric Gaussian model for NDVI smoothing is discussed by Al-Nahhal et al. (2019). Yang et al. (2019) used a weighted double-logistic function to produce gap-free NDVI time-series from original contaminated observations. Spline-based smoothing has also been employed to identify coarse resolution temporal patterns, such as the distinction between annual wet and dry season while removing the signal noise caused by individual storm events (Patterson et al. 2020). Although these regression-based smoothing approaches have been quite popular, static smoothing kernels ignore local variations and affect reconstruction accuracy. The Savitzky-Golay-filter-based approaches employ linear regression to fit adjacent data points with a low-degree polynomial to improve VI curves (Shekhar 2016; Cao et al. 2018; Kong et al. 2019). However, these approaches require polynomial order and filter length to be selected appropriately to preserve spectral characteristics of VI curves and avoid extraneous oscillations. Deep convolutional autoencoder-based denoising approaches learn nonlinear feature spaces to avoid linear events, sparsity, and low-rank assumptions of the traditional interpolation methods (Wang et al. 2019; Li et al. 2019; Lai et al. 2019; Kolbæk et al. 2019). However, generally, these approaches do not consider the irregular sampling of the data and series-specific features. DL-based approaches that attained success for processing irregularly distributed point data exploit local structures using permutation invariant feed-forward network to aggregate local features (Li et al. 2015; Qi et al. 2017; Mao et al. 2019; Guo et al. 2020; Chai et al. 2020). Although these approaches are applicable to denoising VI curves, the effective partition and selection of point clouds remain challenging. Graph convolutional network (GCN) based point cloud processing, which builds local graphs and aggregate local features through convolution or pooling, are also helpful in the reconstruction of VI curves (S. Wang et al. 2018; Wang et al. 2018; Wang et al. 2019). However, the K-nearest neighbors (KNN) based strategy, employed by most of these approaches, is sensitive to point cloud density.

Moreover, the multi-layer-perceptron-based learning of point coordinates ignores some explicitly defined geometric relations. In summary, most of the existing DL-based reconstruction techniques, applicable to the smoothing of VI curves, cannot adequately handle irregular sampling and rely on uniformly and densely sampled unaliased input data. Besides, simple convolutional encoder-decoder frameworks ignore the point nature of the data and cannot be effective in modeling local VI curve features. We hypothesize that effective smoothing or denoising requires latent space projection using interpolated convolution to consider the point nature and irregular sparse sampling of the data. Additionally, the data imputation should also consider the local features as well as the piece-wise reconstruction similarities.

2.2. Generation of field-level representation

DL-based encoding–decoding approaches, applicable for image and signal reconstruction tasks, are relevant in generating a field-level representation of a group of pixel-level VI curves. Recent advances for learning unsupervised representations are summarized in (Bengio et al. 2013; Cheriyadat 2014; Girin et al. 2020). Generative adversarial networks (GANs) and alternatives learn useful latent representations for different image processing applications such as multi-view image generation, image translation, and style transfers (Tian et al. 2018; Gui et al. 2020). Wang et al. (2018) combined generative adversarial and perceptual adversarial losses to train two feed-forward CNNs to solve image-to-image transformation tasks. Anirudh et al. (2020) used a surrogate network to approximate observation variability without the need for additional supervision or data augmentation. Emami et al. (2019) employed a spatial attention mechanism in the discriminator to help the generator focus on the discriminative regions between the source and target domains for achieving a lightweight model. However, the latent representations learned in these GAN-based approaches do not accurately model the VI curve features, and the approaches generally require a large number of training samples. Discriminative and generative variants of autoencoders are also employed for learning optimal feature representations (Tschannen et al. 2018; Girin et al. 2020). Hoshen (2018) used a stochastic function to map input samples to latent codes where multiple analogies of samples were parameterized. Although the method is much faster at inference time and leverages on large datasets, the irregular sampling and the importance of local features are not considered. Zufan Zhang et al. (2019) proposed a signal signal-specific feature fusion approach. However, the approach emphasizes feature modeling for classification rather than reconstruction. Peng et al. (2019) employed a fully connected neural network to jointly learn a collection of hierarchical representations and cluster assignments in an end-to-end manner by exploiting the prior invariant sample assignment. An unsupervised deep metric learning model designed by Kang et al. (2020) used a spatial augmentation criterion to uncover semantic relationships among land cover tiles. Kang et al. (2020) proposed a representation learning method that explicitly modeled and leveraged sample relations to encode real data manifold. However, these approaches ignore the irregular sampling and sequential nature of the VI curves. Graph-based encoding strategies have recently been proposed to encode graphs preserving their contextual and geometric information (Kipf and Welling 2016; Hasanzadeh et al. 2019). Although different generative and discriminative encoding approaches have been widely employed for various applications, we hypothesize that for proper generalization of a group of VI curves, encoding to a normally distributed space should preserve the characteristic features of the VI curves and consider irregular sampling and time-series nature. To the best of our knowledge, DL has not been used for generalizing VI curves for deriving effective field-level representations or phenology fingerprints.

2.3. *Classification and fractional area estimation*

Long short-term memory (LSTM) and variants, which are optimal for the classification of time-series data, are relevant for VI curve classification. Recently, Karim et al. (2018) proposed an attention mechanism for LSTM-based classification by augmenting convolutional blocks with squeeze-and-excitation blocks to improve accuracy. However, most LSTM-based approaches ignore the characteristic features of VI curves, and the Euclidean losses are prone to sampling biases. Among the recent GAN-based approaches applicable for classifying VI curves, Hang, Zhou, Liu, and Ghamisi (2021) proposed simultaneous optimization of reconstruction and misclassification losses to take advantage of unlabeled samples. However, the approach uses one-to-one correlation-based similarity measures and is prone to shifts and distortions prevalent in VI curves. Continual learning, CNN, and elastic weight consolidation loss were employed by Shi et al. (2019) to classify radiofrequency signals considering the phase shifts. Although Han et al. (2020) illustrated CNN to generate task-specific features, the architecture failed to effectively process irregularly sampled point data with a limited number of training samples. Mou and Zhu (2020) proposed a spectral attention module that used a gating mechanism to recalibrate spectral bands adaptively. Jiang et al. (2020) proposed an unsupervised discriminative reconstruction constrained GAN for hyperspectral anomaly detection, but the approach assumes class imbalance of lesser abnormal samples. In order to address the issue of limited availability of training samples, Jia and Liu (2020) proposed eigenmap-based feature extraction and dimensionality compression to significantly reduce the number of parameters in the DL model. However, most of these approaches yield limited results for VI curves, where shape similarity is a major consideration for accurate recognition. Moreover, most of the existing classifiers consider VI curves as vectors ignoring the characteristic features that are physically significant in estimating the phenological similarity of crops. The capsule-based approach, proposed in Arun and Karnieli (2021), addressed the modeling of crop-specific phenological events. However, the approach is computationally complex and requires a large number of training samples. DTW-based approach, proposed in Grabocka and Schmidt–Ieme (2019), used a wrapping function as an upper-level neural network to model the alignment of time-series indices in deep representation space. Another similar approach leveraged DTW to better feature extraction by employing a stochastic backpropagation scheme (Cai et al. 2019). The approach by Iwana et al. (2020) employed a DTW-based neural unit but did not explore it for convolutional frameworks and also had the requirement of fine-tuning the slope constraint. Although DTW-based approaches have yielded acceptable results, the parameter tuning requirements affect their effectiveness and generalizability (Iwana et al. 2020). We hypothesize that DTW-based convolutional layers and capsule-based feature learning can model sequential nature and characteristic features of the VI curves effectively. The convolutional implementation of time wrapping is hypothesized to dynamically learn kernels for estimating feature-specific shape similarity correspondences of VI curves.

3. Materials and methods

This section discusses the datasets used, proposed approaches, and implementations. A brief discussion of the study area and datasets used are presented in Subsection 3.1. The proposed approaches for resolving each of the issues are discussed in Subsection 3.2. The implementation details of the proposed approaches regarding the data considered in this study are presented in Subsection 3.3.

3.1. *Datasets*

The current study employs the Vegetation and Environment monitoring New Micro-Satellite (VENµS) data collected over two agricultural farms in Israel for field-level phenology assessment and crops monitoring (https://karnieli-rsl.com/ven%C2%B5s). The VENµS sensor is characterized by a high spatial resolution of 5 m, a high spectral resolution of 12 narrow bands in the visible to near-infrared regions of the spectrum, and a high revisit time of 2 days at the same viewing and azimuth angles. The NDVI values of barley, wheat, and potato crop fields computed over the crop years 2017–2018 and 2018–2019 are used for various analyses. The study area consists of different fields, among them 23 fields of barley, 90 fields of wheat, and 20 fields of potato. Analysis of the proposed frameworks is conducted for both years. The image metadata provided the cloud cover, and cloud masks were applied in the VENµS level-2 image data.. The actual cloud affected VI curves are used to evaluate the proposed smoothing approach. The shapefiles of crop fields, along with the cropping and harvesting information obtained from framers, serve as ancillary data for labeling the VI curves.

3.2. *Proposed approaches*

The overall workflow involved in the phenology fingerprint generation and analysis is summarized in Fig. 1. The shaded blocks indicate the use of the proposed approaches in resolving the issues prevalent in related analyses. The proposed smoothing framework, PKNet, adopts a dynamic kernel-based strategy to learn filter parameters from the data. The generalization framework, DPGNet, proposed in this research, is employed for obtaining representative characteristic curves for each field. DPGNet is also employed for generating phenology fingerprints of each crop from the representative curves of multiple fields. The proposed DTCapsNet classifier uses DTW layers and features specific transformations to assign crop labels to a given VI curve or detect the anomaly.

3.2.1. DL-based smoothing of VI curves

In the proposed PKNet framework, presented in Fig. 2, the vectorized input VI curve is encoded to multiple scales and abstraction levels, resulting in multi-level, multi-resolution feature representations. The decoder successively synthesizes the complete data, based on the prior, starting at low-resolution feature maps (denoting large-scale structures) up to high-resolution feature maps (representing fine-scale structures). PKNet uses interpolation-based convolutional units instead of regular ones, thereby addressing the problem of the irregular nature of VI curves and modeling of local features. The interpolation-based convolution of vectorized VI curve $\mathbf{\upsilon}$ with a kernel κ(.), centered at a location x, is implemented as:

$$\mathbf{v}^{*}\kappa(\mathrm{x}) = \sum_{\mathrm{x'}}\frac{1}{\mathrm{N}_{\mathrm{x'}}}\sum_{\mathrm{k}_{\alpha}}\varphi(\kappa_{\alpha},\mathrm{x'})\mathbf{v}(\mathrm{x}+\mathrm{x}_{\alpha}).\kappa(\mathrm{x'}) \tag{1}$$

where φ(.,.) is an interpolation function that computes the weights based on a filter weight vector k_α and a given input point x', and $N_{x'}$ is the density normalization term to make the convolutions sparsity invariant. Apart from conventional hyper-parameters, interpolation-based convolution uses kernel length defined as the distance between two adjacent weight vectors to control the receptive field.

The convolution units in the encoder employ padded convolutions followed by leaky-Rectified Linear Units (leaky-ReLU) activations and pooling operations. Convolutional units following the down-sampling steps increase the number of feature maps where the stride selection and sampling factor depends on the application and the data. The decoder consists of up-sampling of the feature maps followed by a concatenation of the same with the copied encoder feature maps at the corresponding resolution. The use of skip connections resolves the issue of vanishing gradients, and the concatenated feature maps are reprojected using convolution layers. The final layer employs 1 × 1 convolutions to map multiple feature maps to the desired output. For training the network, in addition to the L2 regularization loss, to ensure piecewise similarity, a multiscale version of the structural dissimilarity loss is also employed as:

$$\mathrm{L_{SD}} = \sum_{\mathrm{p}\in\mathbf{P}} 1 - \Omega(\mathrm{p})$$

$$\Omega(\mathrm{p}) = \frac{2\mu_{\mathrm{p}}\mu_{\mathrm{p}}^{'} + \mathrm{C}_1}{\mu_{\mathrm{p}}^{2} + \mu_{\mathrm{p}}^{'\,2} + \mathrm{C}_1}.\frac{2\sigma_{\mathrm{p}}\sigma_{\mathrm{p}}^{'} + \mathrm{C}_2}{\sigma_{\mathrm{p}}^{2} + \sigma_{\mathrm{p}}^{'\,2} + \mathrm{C}_2} \tag{2}$$

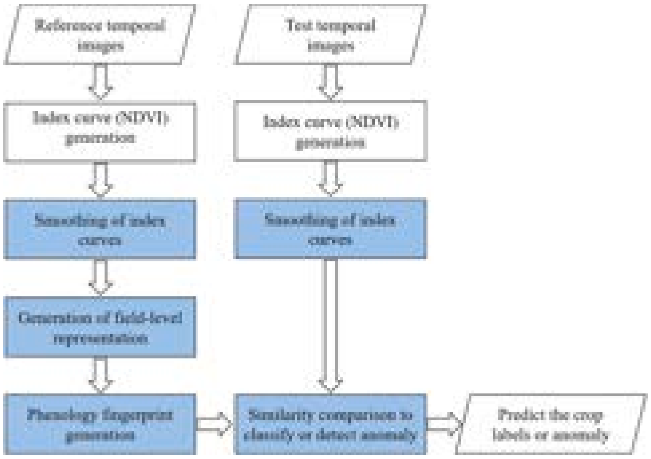

Reference temporal images
Test temporal images
Index curve (NDVI) generation
Index curve (NDVI) generation
Smoothing of index curves
Smoothing of index curves
Generation of field-level representation
Phenology fingerprint generation
Similarity comparison to classify or detect anomaly
Predict the crop labels or anomaly

Fig. 1. Workflow of phenology-based crop analysis (Shaded boxes denote the applicability of the proposed approaches).

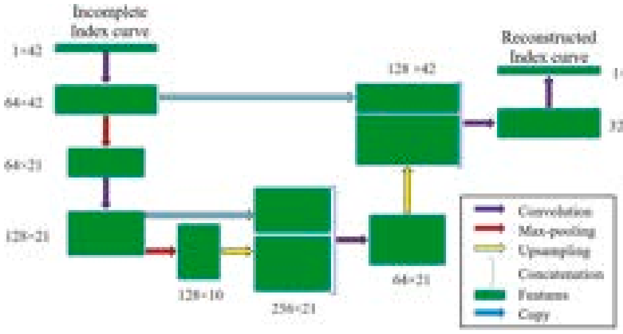


Fig. 2. Architecture of the proposed PKNet for index curve smoothening and reconstruction (Size and shape of feature blocks represent relative differences in feature dimensions).

where P ⊆ R denotes the set of all relative locations on the VI curve, C_1 and C_2 are constants, μ_p and $\mu_{p'}$ respectively represents the means of the patches of the reconstructed and ground-truth VI curves while σ_p and $\sigma_p^{'}$, respectively, denote the corresponding standard deviations. The means and standard deviations are computed in neighborhoods (context) of varied extents to implement multiscale measurements of structural dissimilarity.

A back-projection-based refinement is proposed to further refine the results for missing or erroneous value imputation (not applicable for improving sparsely sampled VI curves). The errors in the reconstruction of known values of the VI curve are estimated. The weighted average of the reconstruction error is computed for the known values in the context of each missing value. The weights are determined based on the spatial distance and spectral similarity of the given value with respect to the missing one. The weighted error estimated for the missing value at the i^{th} position of the VI curve is computed as:

$$\delta\mathrm{E_i} = \sum_{\mathrm{j}\in\mathrm{C_i}}\left(\mathrm{e}^{-\left((\mathrm{p_i}-\mathrm{p_j})+\lambda(\mathrm{v_i}-\mathrm{v_j})\right)}\right)\delta\mathrm{E_j} \tag{3}$$

where δE_j is the reconstruction error of the known value at the j^{th} location of the VI curve, λ is the normalizing factor, C_i denotes the spatial context of the i^{th} position on the VI curve, p_i and p_j are the i^{th} and j^{th} positions, and v_i and v_j are the values on the reconstructed curve at the i^{th} and j^{th} locations.

3.2.1.1. Generation of field-level representation. This study also investigates the application of DL in learning a generalized or abstract representation of a group of VI curves in the context of crop classification. In this regard, DPGNet employs a variational autoencoder (VAE) based network to estimate the generative and encoding models G_Θ and D_ϕ, respectively, for a group of VI curves. The nonlinear encoder function D_ϕ maps the group of VI curves to a low-dimensional manifold representation while G_Θ decodes the latent vectors back to the corresponding VI curves. It may be noted that the approach employs interpolation-based convolution (Subsection 3.2.1) instead of normal convolution. The architecture of DPGNet adopted in this study is presented in Fig. 3. The skip connections are employed to resolve the vanishing gradient problem. As shown in the figure, mean (μ) and standard deviation (σ) vectors are used to sample the latent vectors using the reparameterization trick, ensuring that the latent vectors

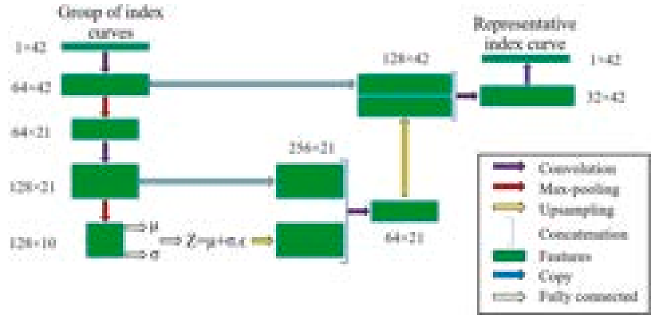


Fig. 3. Architecture of the proposed DPGNet for representative index curve generation (Size and shape of feature blocks represent relative differences in feature dimensions).

follow a unit Gaussian distribution.

The learned models (G_Θ and D_ϕ) will be able to describe the manifold of the VI curve variability accurately. Any arbitrary observation v_i of an VI curve belongs to the set $v_i \in \{G_\Theta(z) : z \in R^k\}$ and thus can be equivalently represented by a corresponding low-dimensional vector z $\in R^K$ in the latent space of the generative model G_Θ. As the output of VAEs varies smoothly with changes of the latent variables (Kingma and Welling 2014), the latent representations generalize the group of VI curves effectively and are robust to outliers. Unlike the existing VAEs, the loss function of DPGNet is composed of a fitting data term (maximizing the data likelihood) and a latent compression term (enforcing that q(z|v) stay close to the latent prior $p_\theta(z)$) along with label information embedding factor as:

$$L_{VAE} = E_v\left[E_{q_\tau(z|v)}\left[-\log(p_{\tilde{\imath}}(v|z))\right]\right] - E_v\left[D_{KL}(q_\tau(z|v)||N(z|\mu_y, I))\right] - \beta D_{KL}\left(q_\tau(z)||\prod_{j=1}^{d} q_\tau(z_j)\right) \tag{4}$$

where v is the input VI curve, μ_y is the posterior mean of class y, z is the latent representation, $q_\tau(z|v)$ is the encoding distribution, $p_{\tilde{\imath}}(v|z)$ is the decoding distribution, d is the dimension of the latent space modeled by the encoder, β is the entanglement penalty factor, and $D_{KL}(.||.)$ is the Kullback–Leibler divergence. Although penalizing $D_{KL}\left(q_\tau(z|v)||N(z\middle|\mu_y, I)\right)$ facilitates disentanglement amounts to the loss of the information about v stored in z resulting in a poor reconstruction. The reformulation adopted in equation (4) resolves the effect of the stochastic sampling in the latent code, where the last term measures the dependence for multiple variables. In addition, $p_{\tilde{\imath}}(z)$ is considered as a mixture of unit Gaussians and is conditioned on the label information to refine the latent space based on the class label information. The embedding of label information obviates the need to marginalize over all classes to compute the K-L divergence.

The DPGNet constraints latent representations of the given group of pixel-level VI curves to be in unit Gaussian, preserving the crop-specific information. The mean of the distribution is sampled and decoded to yield the representative VI curve for the given field. A similar approach is used to obtain the phenology fingerprint of a given crop from a group of characteristic curves. Instead of sampling the mean value, a bundle of curves in the range of one standard deviation around the mean is also averaged to get an alternate characteristic representation.

3.2.1.2. Feature-based similarity estimation. This subsection proposes a VI curve classifier that estimates shape-based similarity and effectively considers the VI curve features. The estimation of fractional area coverage is also investigated based on the anomaly of a given VI curve from its representative curve. The proposed DTCapsNet classifier uses DTW-based convolutions and capsule units instead of normal neural units. Capsules are a group of neurons that model characteristics of features as the orientation of their output vectors. The architecture of DTCapsNet, adopted in this study, is presented in Fig. 4.

As shown in Fig. 4, the input VI curves are fed to successive convolutional units, where max-pooling layers follow the second and third convolutional units. Unlike regular convolutional units, DTW-based nonlinear units (Iwana et al. 2020) are implemented in a convolutional framework for feature extraction. The activation of a given DTW node n is computed as:

$$a_n = \phi\left(\sum_{(i,j)\in M} ||w_{n,i} - x_j||\right) \tag{5}$$

where × is a vector of the input values $x_{1,\ldots,}x_m$ and w_n is a vector of the respective weights $w_{n,1},\ldots,\ w_{n,m}$ to the node n. The function ϕ(·) is a nonlinear activation function applied to the result, and M is a set of all the matched pairs between × and w_n computed using the dynamic wrapping approach discussed in (Iwana et al. 2020). In addition to modifying the neural units, the convolution operation is also modified to consider the specific characteristics and noisy nature of the VI curves. An interpolation-based convolution (Subsection 3.2.1) is employed instead of the normal convolution.

The feature tensors from convolutional layers form the feature capsules (FCs) (as shown in Fig. 4) and are connected to each of the class capsules. The length of FC's output vectors denotes the likelihood of finding the corresponding feature/pattern, while the orientation denotes the instantiation parameters. Unlike the regular Capsulenet (Sabour et al. 2017), DTCapsNet appends the features derived at different hierarchies (for modeling the local and global features) to obtain the feature capsules. The capsule layers employ a DTW-based dynamic routing to refine the transformations in conjunction with the back propagations. Unlike in the regular capsule network implementations, the dynamic routing is modified to consider the shape and characteristic features of the VI curves. In this regard, the log priors that measure the similarity between the output of the j^{th} spectral class capsule (e_j) and the prediction vectors $\hat{u}_{j/i}$ is computed as:

$${b_{ij}}^k = {b_{ij}}^{k-1} + \sum_{m,n\in M} ||{e_j}^{k-1}{}_m - {\hat{u}_{j/i}}^{k-1}{}_n||^2 \tag{6}$$

where ${b_{ij}}^k$ denotes the logits at each iteration k, and M is the set containing the indices of elements along the warping path between the ${e_j}^{k-1}$ and ${\hat{u}_{j/i}}^{k-1}$. The wrapping path between two vectors is computed according to the discussions in (Cuturi and Blondel 2017; Iwana et al. 2020).

The 1D capsules, shown in Fig. 4, learn features of VI curves and their characteristics in conjunction with the objective of crop classification. The DTW-based shared convolution weights and GAN-based data augmentation, coupled with a reduction in the number of hyperparameters are the characteristics of the proposed approach. The

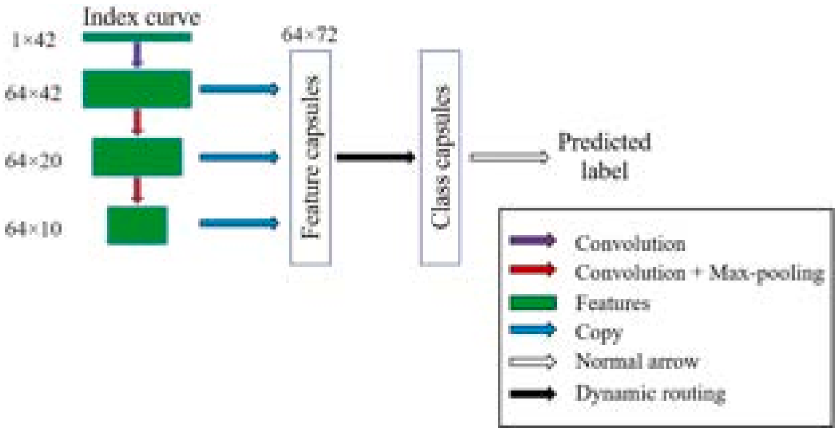


Fig. 4. Architecture of the proposed DTCapsNet classifier (Size and shape of feature blocks represent relative differences in feature dimensions).

cross-entropy loss is minimized to train the network weights. To model DTCapsNet for estimating fractional abundances from field-level VI curves, the normalized length of the output vector of each class is used as an estimate of the fractional abundance.

3.3. *Implementation details of the frameworks considered in the current study*

This subsection discusses the design and optimal hyper-parameter settings of each of the architectures and approaches proposed and compared in the current study. It is worth noting that the architectures and hyper-parameters need to be fine-tuned concerning the input data. The configurations adopted are discussed in the following subsections.

3.3.0.1. *Smoothing of NDVI curves*

The encoder-decoder-style PKNet, employed for solving missing data imputation as shown in Fig. 1, uses multi-size kernels of sizes 1 × 2, 1 × 3, 1 × 5, 1 × 7, and 1 × 9 in the encoder and decoder streams. The padded convolutions are followed by ReLU activations in all layers except the last two decoding layers that use tanh activation. The stride of pooling and unpooling layers is kept to two for respectively halving and doubling the resolution of the resulting feature maps. Convolutions following down-sampling steps double the number of feature maps while is halved by the convolutions following upsampling. The convolution kernels use Gaussian interpolation (as the interpolation function) and the Gaussian bandwidth (3σ) is fixed to 0.1. The number of filters in the first encoding unit is empirically set to 32.

The input data is jittered by a zero-mean Gaussian noise with 0.02 standard deviation at selected random locations. The network is trained for 480 epochs with an initial learning rate of 0.001 and a decay rate 0.7 every 80 epochs with batch size 30. Hyper-parameter optimization, proposed in (Bochinski et al. 2018), is employed to optimize the parameters of the proposed network. The mean squared error (MSE)-based loss and cosine dissimilarity loss (Arun et al. 2019), along with the proposed piece-wise dissimilarity loss, are employed to learn the network weights. For implementing multiscale structural dissimilarity measurements, the context extents are varied from 1, 3, 5, 7, and 9.

The implementations of benchmark denoising approaches are directly adopted from the corresponding literature. In the least-square-based fitting approaches (Yang et al. 2019; Al-Nahhal et al. 2019), the key temporal points and parameter sets are decided based on the phenological growth transitions, planting, heading, and harvesting information obtained from the farmers. For the spline-based (Patterson et al. 2020) and Savitzky-Golay-filter-based (Cai et al. 2017) approaches, the smoothing parameter, upper-envelope iterations, upper-envelope strength, and stiffness adaptation strength are respectively set to 2.34, 3, 7, and 0.49 using 10-fold cross-validation. The hyperparameters for the approach by Lai et al. (2019), such as learning rate, momentum rate, and training epochs, are respectively set to 0.3, 0.1, and 800. A total of 6 hidden layers are employed with respective dimensions as 36, 18, 12, 12, 18, and 36. The approach by Chai et al. (2020) is implemented as a fifty layer framework including 19 convolutional (with ReLU activation), max-pooling, up-sampling, and concatenate layers. In all the DL-based approaches, adopted as benchmarks in the current study, max-pooling and upsampling layers respectively downscale and upscale the resolution of the features by a factor of 2. Also, a hyperparameter optimization technique, discussed in (Gulcu and Kus 2020), is employed to optimize the network parameters.

3.3.0.2. *Generation of field-level representation*

The generative and encoder models of the proposed DPGNet are learned from the given group of VI curves. The latent space dimension is empirically set to be 10 (K = 10) as it adequately captures the variability of the VI curves of different crops (potato, barley, and wheat) considered in this study. The encoder network constitutes of two convolutional units where downscaling halves the resolution of the feature maps and convolution doubles the number of features. The number of filters of the first convolution unit is set to 32. All the layers employed multi-sized kernels of sizes 1 × 2, 1 × 3, 1 × 4, 1 × 5, 1 × 7, and 1 × 9 pixels. A fully connected two-layer network is found to be optimal for sampling (through reparameterization) the 10-dimensional latent space. The decoder network employs three convolutional units, of which two are followed by upsampling layers (scale factor = 2). The decoding convolutional layers, except the last layer, halve the number of features. The last decoding convolution layer combines all the features to reconstruct the VI curves and uses tanh activation function.

Hyper-parameter optimization, proposed in Bochinski et al. (2018), is employed to optimize the parameters of DPGNet. The MSE-based loss, cosine dissimilarity loss, and the proposed piece-wise dissimilarity loss are minimized to learn the network weights. The context extents are varied from 1, 3, 5, 7, and 9 for implementing the multiscale measurements of structural dissimilarity. The network is trained for 250 epochs using batch optimization with a mini-batch size equal to one-third of the total amount of training data for each field.

Different benchmark reconstruction approaches discussed in this study are implemented based on the descriptions in the corresponding research papers. The optimization approach, proposed in Gulcu and Kus (2020), is adopted to fine-tune the hyper-parameters of the DL-based reconstruction models. For implementing the approach by Hoshen (2018), the settings of unconditional generative model, VAE encoder and mapping module are adopted from the literature with the latent code dimension set to 36. Besides, a learning rate of 0.3 is used across all parameters that are updated with every training batch.

The method by Wang et al. (2018) is implemented by alternatively optimizing the discriminator and transformation networks using Adam solver with a learning rate of 0.001 and a momentum of 0.5. The hyperparameters for balancing the influence of generative and perceptual adversarial losses are empirically set to 1. The results reported for Anirudh et al. (2020) are obtained using a 12-layered network that is optimized using RMSProp with learning rates set as 0.01 and 0.03, respectively, for the first and second inner loop iterations. The encoder-decoder approach, proposed in Kang et al. (2020), is implemented as a 10-layer framework with a fully connected mapping layer separating the encoder and decoder. The filter sizes are empirically set to be 1 × 3 and a number of filters are set to be 256 in all layers except the last layer.

3.3.0.3. Classification and fractional area estimation

In the proposed implementation of DTCapsNet, the cross-entropy loss is minimized for training the network. The features used as ‘feature capsule-1′ are not translation invariant as they are derived before pooling. The max-pooled features used for ‘feature capsule-2′ and ‘feature capsule-3′ capture the feature hierarchies for effectively modeling the VI curves. For crops having similar VI curves, consideration of pooled features improves the separability and classification accuracy. Instead of using max pooling for capsule-2 and capsule-3, the dilated convolution has also been experimented with.

DTCapsNet has 32 1D-filters in the first, second, and third convolution units in the current implementation. The sizes of filters in each layer are 1 × 3, 1 × 5, 1 × 7, and 1 × 9, and for each size, there are 16 filters in each unit aggregating to 64. For the NDVI curves, having a length of 42, the dimensions of the output of the first, second, and third convolution units are 64 × 42, 64 × 20, and 64 × 10, respectively. The final feature vector, having a length of 72 (42 + 20 + 10), comprises features (of the NDVI curves) at different hierarchies. In other words, the feature capsules have nine channels of convolutional 8D-capsules. The mini-batch size for training is set to 80; the momentum for backpropagation is set to 0.3 (obtained through cross-validation); the learning rate is initially set to 0.4 and is depreciated by a factor of 2 after every 50 epochs. Hyper-parameter optimization, proposed in Bochinski et al. (2018), is employed to optimize the parameters of DTCapsNet. As the length of the class capsule’s output vector is used as an estimate of the fractional abundance, no further network modification is required to estimate the fractional area covered by each crop in a given field (represented by the field-level VI curve).

The implementation of baseline classifiers applicable to the classification of VI curves is adopted from the corresponding literature. The approach, proposed by Rubwurm and Korner (2017), is implemented as an 8-layer cascaded LSTM classifier trained using Adam optimizer. The framework, discussed in Jia and Liu (2020), is adopted as constituting 3 dual-scale convolution and bi-channel fusion units. The RMSprop algorithm is employed, with a learning rate of 0.006 and weight decay of 0.005, to optimize the framework. As for all the DL-based benchmark approaches, the hyperparameter optimization in Gulcu and Kus (2020) is employed to fine-tune the hyperparameters such as the number of filters, network depth, and size of filters. The approach proposed in Mou and Zhu (2020) is implemented as an end-to-end trainable 10-layer CNN framework consisting of a spectral attention module. The parameters and optimization settings as recommended in the corresponding literature along with a with learning rate of 0.006 are adopted to train the network. The GAN-based approach proposed by Hang et al. (2021) is implemented as composed of a 4-layered encoder module, 4-layered decoder module, and 4-layered classifier module. The hyper-parameters for balancing the influence of generative adversarial loss and perceptual adversarial loss are empirically set to 0.2 and 0.06, respectively, while the fusion weights are set to be 0.26, 0.34, and 0.40. A similar GAN-based approach, discussed in Jiang et al. (2020), is implemented as an 8-layer autoencoder framework. The framework is trained using adversarial loss constrained by enhanced representation and shrink constraints, each having a scale factor of 1.

Among the baseline spectral unmixing approaches considered in this study, Dou et al. (2020) is implemented as an autoencoder framework where the encoder and decoder are trained alternatively. An Adam optimizer is employed with a learning rate of 0.008 and momentum of 0.7. Besides, the hyperparameters such as neighborhood are set to 1, and the scale factors of loss functions are respectively set to 0.5, 1, 1, and 10. For the approach by Qian et al.(2020), a 6-layered feed-forward network is used to implement linear mixing for blind unmixing. The RMSprop algorithm is employed, with a learning rate of 0.07 and weight decay of 0.002, to optimize the framework. For implementing the unmixing architecture discussed in Borsoi et al. (2019), a 6-layer network with ReLU activations is trained using Adam optimizer with a mini-batch size equal to 100. The scaling factors to constrain the latent representations are empirically selected to be 0.3 and 0.01. Su et al. (2019) proposed the benchmark unmixing approach, which is implemented with the hyperparameters for the regularizations on the mixing matrix and abundance fractions both set to 0.1.

4. Results

To verify the effectiveness of the proposed methods, extensive experiments have been conducted using VI curves derived from VENμS satellite images. The ablation analyses of the proposed frameworks are discussed in Subsections 4.1.1, 4.2.1, and 4.3.1, and Subsections 4.1.2, 4.2.2, and 4.3.2 present comparative analysis of the proposed approaches with the benchmark approaches in the corresponding domain. Peak signal to noise ratio (PSNR) is used to evaluate the effectiveness of the reconstruction methods (denoising and generalization approaches). Confusion-matrix-based Kappa statistics is applied to evaluate the classification results. High values of PSNR and Kappa statistics indicate high accuracy. The Z-score-based test statistics (discussed in Herrmann et al. (2013)) is employed to analyze the significance of the results presented in this study.

4.1. Smoothening of field-level NDVI curves

The pixel-level NDVI curves are used to train PKNet. Different downscaling strategies such as bilinear, bi-cubic, and nearest neighbor interpolations are employed to generate training and testing samples. Multiple downscaling strategies are adopted to avoid the bias of the trained network towards a particular approach. A GAN-based augmentation, similar to the one adopted in Zhiwen Zhang et al. (2019), is also used to increase the number of training samples. Besides, a random approach is employed to remove values or add Gaussian noise at irregular intervals. The effectiveness of denoising is also tested on samples affected by the cloud. PKNet is extensively analyzed over the data of three crops over two consecutive crop years. The reconstruction PSNR, measured between the original ground truths and the corresponding smoothened (reconstructed) VI curves, is used to evaluate the missing data imputation/denoising approaches.

4.1.1. Ablation analysis of PKNet

Analyses of the effect of the different modules on PKNet are presented in Table 1. As is evident, the interpolated convolution (Subsection 3.2.1), instead of the normal convolution, significantly improves the PSNR. The improvement in accuracy can be attributed to the

Table 1
Comparative analysis of the effect of architectural variations for PKNet for 60% of training samples.

Architectural variations	PSNR
PKNet without using interpolated convolution	29.63
PKNet without using DTW-based convolution	30.76
Stacked denoising autoencoder	23.12
Stacked denoising variational autoencoder	25.47
PKNet without back-projection-based refinement	32.53
Proposed PKNet implementation	**35.81**

Table 2
Analysis of the effect of proposed losses/constraints on PKNet.

Losses/Constraints	PSNR
PKNet without piece-wise loss	32.19
PKNet without cosine dissimilarity loss	33.53
Proposed PKNet implementation	**35.81**

learnable interpolation implemented through convolution. In addition, the proposed strategy facilitates an adaptive receptive field (Subsection 3.2.1), unlike in conventional CNNs, which further improves the handling of missing data. The better performance of the PKNet, compared with the stacked denoising autoencoder (Ca et al. 2010) and stacked denoising VAE (Im et al. 2015), is due to the use of upsampling layers in conjunction with the skipped connections. The VAE requires more training samples and also suffers from reconstruction blur as it attempts to learn the true posterior distribution through optimization of the variational lower bound. The proposed back-projection-based refinement (Subsection 3.2.1) exploits the prior information about the input to give a significant improvement in reconstruction PSNR.

An analysis of the effect of proposed constraints on reconstruction accuracy is summarized in Table 2. As is evident, the use of piece-wise loss (Subsection 3.2.1) significantly improves the reconstruction PSNR. It is also observed that the proposed constraints (Subsection 3.2.1) significantly reduce the training sample requirement as they facilitate learning the intrinsic manifold of the data. The performance of PKNet can be attributed to the proposed constraints, losses, and interpolation-based convolutions.

Analysis of the sensitivity of PKNet shows that an increase in network depth, number of filters, and filter size improves the accuracy to a limit beyond which it deteriorates or saturates. The use of multi-sized kernels and skip connections significantly improves the results without mainly affecting the execution time. An illustration of the sensitivity analysis of the network layers toward network parameters is presented in Fig. 5. The higher the PSNR, the better the reconstruction. Experiments indicate that the proposed back-projection makes PKNet less sensitive to the change in network depth. The increase in kernel size and the number of kernels improve the accuracy to a limit, but the trend saturates or deteriorates gradually. The variational encoding results in better accuracy as compared to the normal autoencoders. Also, PKNet is less sensitive to slight changes in the depth of the reparameterization stream.

Table 3
Comparison of PKNet with benchmark smoothing approaches for 60% of training samples*.

Benchmark smoothing approaches	PSNR
Least-squares fitting to double logistic functions(Yang et al. 2019)	27.34
Spline smoothing (Patterson et al. 2020)	28.68
Autoencoder-based (Lai et al. 2019)	30.45
Savitzky-Golay filtering-based(Cai et al. 2017)	29.42
Deep learning-based (Chai et al. 2020)	31.36
Proposed PKNet implementation	**35.81**

*Benchmark methods are implemented based on the available GitHub implementations and are fine-tuned based on the related publications.

4.1.1.1. Comparison of PKNet with the commonly-used smoothening/denoising approaches. The commonly-used denoising approaches, applicable to VI curve smoothing, are compared with the proposed PKNet. The selected benchmark approaches are improved versions of the ones that reported the state-of-the-art results in Cai et al. (2017) and Julien and Sobrino (2019). Additionally, some of the DL-based approaches are also considered. Some of the benchmark approaches are modified for the one-dimensional VI curves. The result of the comparative analysis is summarized in Table 3 and Fig. 6. Table 4 confirms the significance of the comparison at a confidence level of 95%. As is evident from the results, PKNet gives higher PSNR (better reconstruction accuracy) than the existing approaches. An illustration of the denoising results of PKNet is presented in Fig. 7. Among the conventional smoothing approaches, the local filtering methods (Savitzky-Golay fitting and locally weighted scatterplot smoothing) have given better performance with optimized parameter settings. However, fitting methods (asymmetric Gaussian function fitting and double logistic function fitting) are less sensitive to the parameters. DL-based approaches give better results than the conventional approaches except Savitzky-Golay fitting, where the latter give comparable results when the training samples are scarce.

The better results of PKNet as compared to the conventional and other prominent approaches (including the existing DL-based ones) can

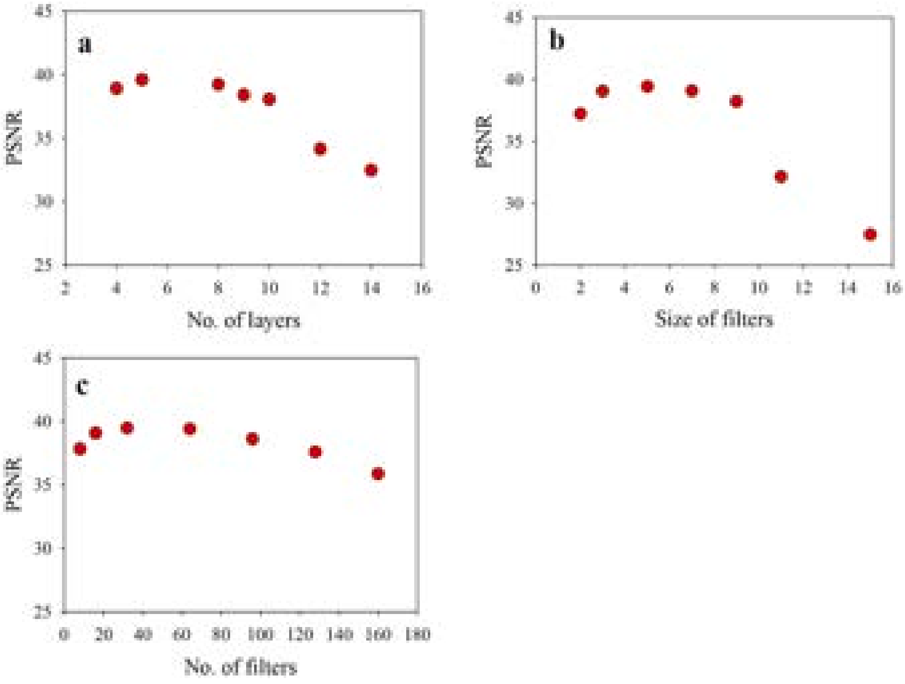


Fig. 5. Analysis of the sensitivity of PKNet towards (a) depth of the network layers; (b) size of filters; and (c) number of filters for 80% of the training samples.

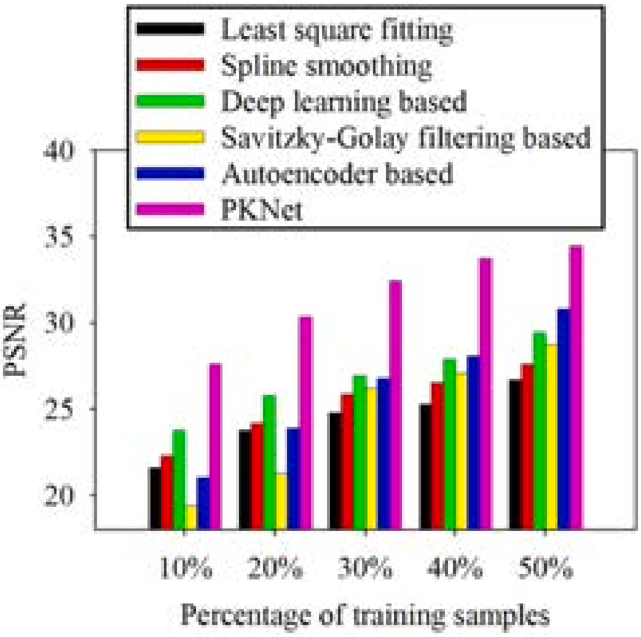


Fig. 6. Accuracy analysis of PKNet and benchmark smoothing approaches with respect to the change in the percentage of training samples.

Table 4
Z-score-based significance analysis of the results of PKNet *.

Benchmark smoothing approaches	Z-score with respect to PKNet
Least-squares fitting to double logistic functions(Yang et al. 2019)	2.32
Spline smoothing (Patterson et al. 2020)	2.71
Deep Learning -based (Lai et al. 2019)	2.18
Savitzky-Golay filtering-based(Cai et al. 2017)	2.23
Deep learning-based (Chai et al. 2020)	1.99

* Z-score > 1.96 shows a significant (>95 %) improvement by PKNet.

be attributed to its ability to address the irregular data sampling and the use of learned feature space for data imputation. Besides, PKNet also addresses the issue of fine-tuning the context and weight parameters as the same is learned dynamically. The spectral dissimilarity loss and the proposed piece-wise back-projection, significantly improve the generalizability and give better results even with limited training samples. The feature-based latent space resolves the issues of domain bias and inter-field variability of VI curves.

Although least squares-based approaches, such as Yang et al. (2019) and Al-Nahhal et al. (2019) are computationally efficient, these approaches do not consider the crop-specific features. Similar is the case with the spline-based (Patterson et al. 2020) and weighted regression-based (Weil et al. 2017) approaches. The Savitzky-Golay filtering-based (Kong et al. 2019) smoothing approaches give good results but require finetuning based on the prior information regarding the crop phenological events. In contrast to the DL-based approaches such as Chai et al. (2020) and Lai et al. (2019), the proposed use of DTW-based neural units along with the interpolated convolution significantly improves the denoising results of PKNet. Besides, the proposed constraints and losses along with the skip connections can also be attributed to the improved performance, especially when the availability of training samples is limited.

4.2. Generation of field-level representation

The DPGNet, which attempts to find a feature-based latent space for generalizing a group of VI curves, is trained to reproduce the inputs. To analyze different approaches regarding the generation of representative VI curves, reconstruction accuracy (in terms of PSNR) is estimated. A GAN-based augmentation (Sandfort, Yan, Pickhardt J, & Summers M, 2019) is used to generate enough training samples for the fields having a limited number of samples.

4.2.0.1. Ablation analysis of DPGNet

An analysis of the adopted architectural choices for DPGNet is summarized in Table 5. It is observed that the projection of VI curves to a normally distributed latent space, guided by adversarial loss, improves the generalizability of DPGNet. The use of skip connections and upsampling layers improves the results. DPGNet outperforms the use of stacked denoising autoencoders and stacked denoising VAEs (in terms of the reconstruction PSNR). The proposed interpolation-based convolution (Subsection 3.2.1) also resolves the issues of irregular sampling and data biases.

The effect of proposed loss functions on the reconstruction accuracy of the DPGNet is presented in Table 6. The use of piece-wise constraints (Subsection 3.2.2) and interpolation-based convolutions (Subsections 3.2.2 and 3.2.1) significantly improve reconstruction accuracy. The use of adversarial loss along with the other proposed constraints in conjunction with the skip connections resolves the issue of reconstruction blur prevalent in VAEs.

Analysis of the sensitivity of DPGNet shows that an increase in network depth improves the accuracy to a limit beyond which it deteriorates. The use of multi-sized kernels and skip connections significantly improves the results without significantly affecting the execution time. The increase in kernel size improves accuracy to a limit beyond which the trend saturates. If the number or size or kernels are further increased without increasing the training samples, the approach gives a poor reconstruction PSNR. An illustration of the sensitivity analysis of the network layers toward network parameters is presented in Fig. 8.

4.2.0.2. Comparison of DPGNet with the commonly-used reconstruction approaches

The proposed DPGNet is compared with the main signal/image reconstruction approaches. The results are summarized in Table 7 and Fig. 9. The approval of the significance of the proposed results at a confidence level of 95% by baseline approaches is illustrated in Table 8. An illustration of the field-level fingerprints generated by DPGNet for randomly selected potato, wheat, and barley fields are presented in Fig. 10. Most of the reconstruction approaches selected for comparison with DPGNet, are DL-based methods and give state-of-the-art accuracies. As is evident, DPGNet outperforms other approaches in terms of PSNR. The approved results of DPGNet can be attributed to its ability to project the VI curves to well-distributed latent representations that follow unit Gaussian. The proposed use of variational encoding and the adversarial loss significantly improves the results and yields better generalization. Similar VI curves are observed to be distributed very near to the mean in the latent space. Hence, the proposed approach of using the mean from the latent space results in better generalization than using moments derived from the VI curves. The use of multiple moments (such as mean, mode, median, and standard deviation), derived from pixel-level VI curves, increases the number of parameters and thereby computational complexity of further analyses. The spectral dissimilarity and piece-wise dissimilarity losses improve the generalizability and give better results even with limited training samples. The crop-specific feature-based latent space, learned from the data, resolves the issues of domain bias and inter-field variability of VI curves.

Although the use of variational encoder, as discussed in Hoshen (2018), facilitates the generalization of a group of VI curves, the DPGNet is preferred due to skip connections and DTW-based units in addition to the interpolated convolution. It is also observed that the use of the piece-wise similarity consideration preserves the features more accurately than the use of MSE-based losses. Besides, the embedding of label information before the latent space transformation and the use of

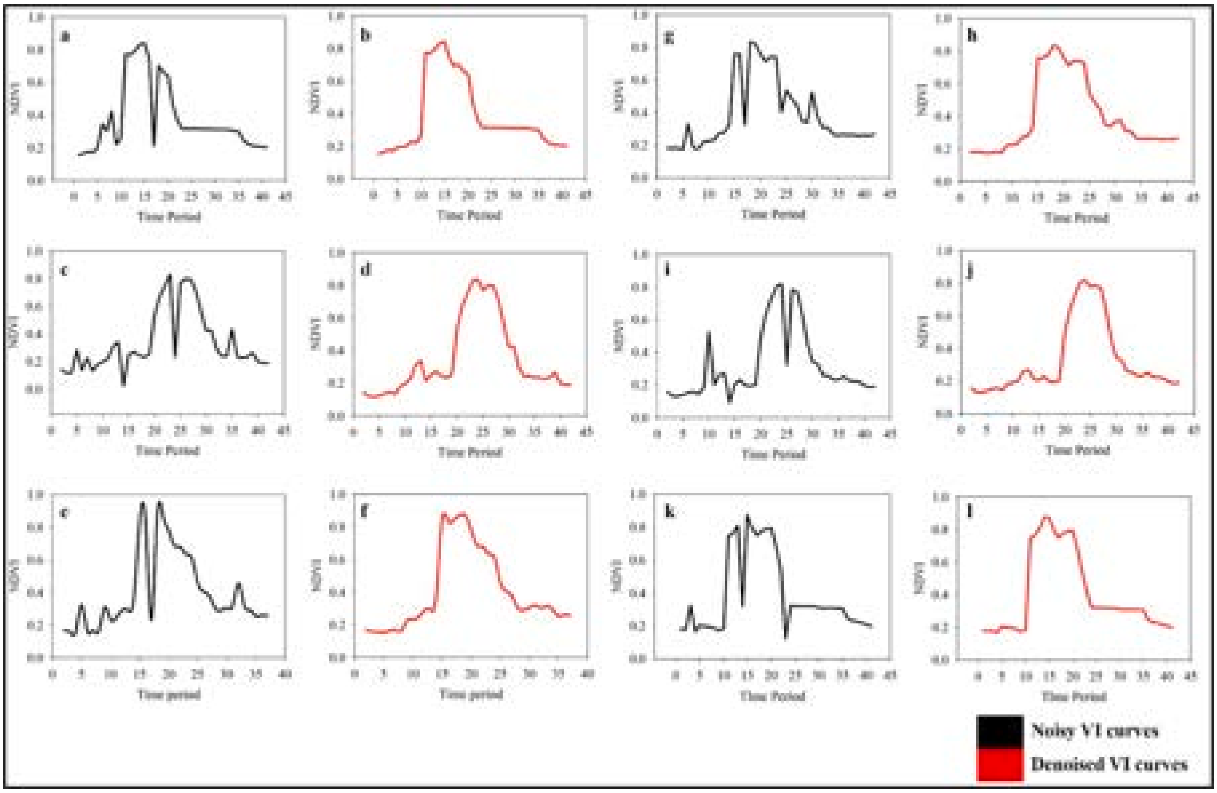


Fig. 7. Illustration of the denoising results for PKNet.

Table 5
Analysis of the effect of the losses/constraints on DPGNet for generating field-level representation.

Architectural variations	PSNR
DPGNet without interpolated convolution	37.85
DPGNet without DTW-based convolutional units	36.93
Stacked denoising AE	33.98
Stacked denoising VAE	34.43
Proposed DPGNet implementation	**38.94**

Table 6
Analysis of the effect of the losses/constraints on DPGNet for generating field-level representation.

Losses/Constraints	PSNR
DPGNet without piece-wise constraint	36.29
DPGNet without cosine dissimilarity constraint	36.88
DPGNet without adversarial loss	36.42
Proposed DPGNet implementation	**38.94**

adversarial loss for improving the latent distribution with respect to the input further improves the performance. These proposed strategies make the DPGNet more preferable as compared to the GAN-based approaches (Wang et al. (2018) and Anirudh et al. (2020)) that do not account for the crop-label information and require more training samples. The representation learning approach, proposed by (Z. Kang et al. 2020), did not yield as good results as DPGNet due to the constraints for leveraging the sample relations that ignore the characteristic features of the VI curves.

4.3. Classification and fractional area estimation

The field-level NDVI curves are compared with phenology fingerprints of different crops to classify them or detect anomalies. The proposed DTCapsNet, presented in Fig. 4, is analyzed concerning the classification of VI curves. The ground truth ancillary data is used to generate labels for VI curves and train the network. For the crops with a limited number of training samples, a GAN-based augmentation (Cubuk et al. 2019) generates more training samples. The confusion-matrix-based Kappa statistics is used to evaluate the effectiveness of different classifiers, while RMSE is used for comparing fractional area estimates by different unmixing approaches.

4.3.0.1. Ablation analysis of DTCapsNet

An illustration of the analysis of the architectural variations for DTCapsNet is presented in 9. The use of DTW-based network layers (Subsection 3.2.3) significantly improves the results (in terms of the Kappa statistics) as it effectively estimates the similarity between two given VI curves (Table 9). In addition, the consideration of interpolation-based convolution (Subsection 3.2.3) has improved the classification due to its capability in handling the irregularly sampled data.

Analysis of the sensitivity of DTCapsNet with respect to the network parameters is summarized in Fig. 11. As is evident from the results, an increase in network depth improves the accuracy to a limit beyond which it deteriorates. The reduction in accuracy at increased network depth can be attributed to over-/under-fitting. Hence, the network's depth needs to be tuned in conjunction with the input's spectral dimension. Generally, deeper networks increase computation time. Empirically, for input curves having a length of 24–48, a 2–6 layered network yields the best results.

The increase in the number of kernels improves the accuracy to a limit, but the trend saturates gradually. The size of filters is a critical factor and needs to be tuned following the data. As the length of spectral features can vary from even one to a few pixels, too big-sized kernels may sometimes ignore essential features. Also, very small-sized kernels may capture the noise instead of actual spectral features. Besides, the increase in size and number of filters exponentially increases the

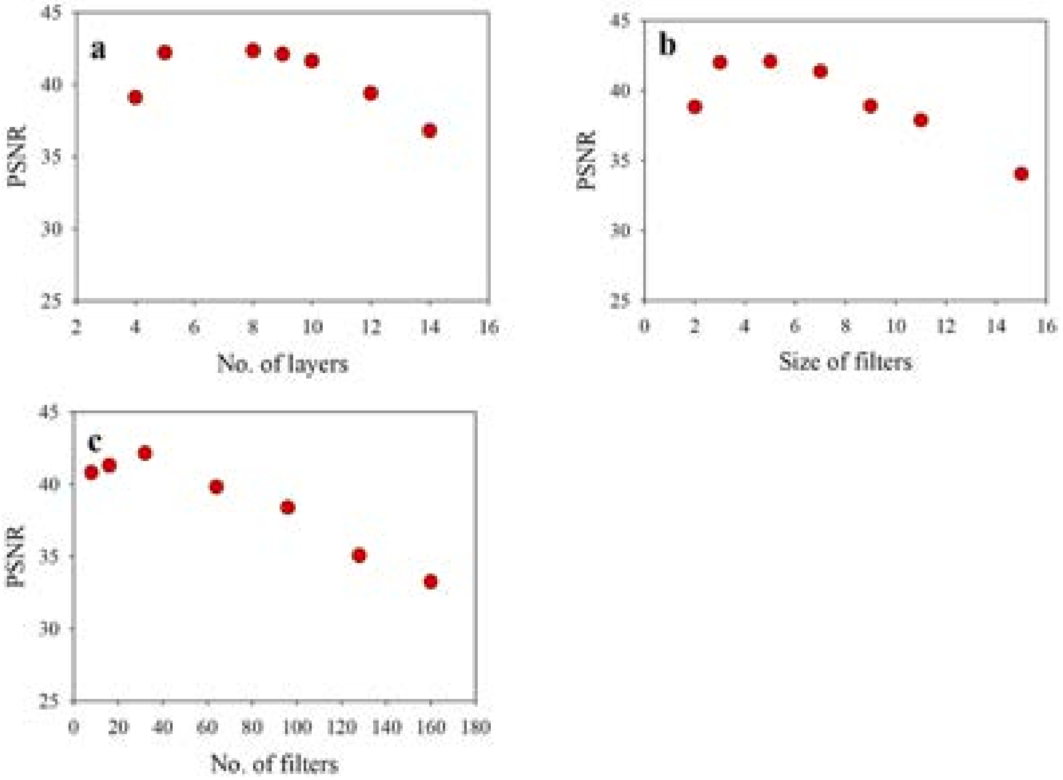


Fig. 8. Analysis of the sensitivity of DPGNet towards (a) depth of the network layers; (b) size of filters; and (c) number of filters for 80% of the samples.

Table 7
Comparison of DPGNet with the benchmark reconstruction approaches for 60% of the training samples*.

Benchmark reconstruction approaches	PSNR
Variational autoencoder (Hoshen 2018)	33.72
Perceptual adversarial network (C. Wang et al. 2018)	30.08
Relation-guided representation (Z. Kang et al. 2020)	32.57
GAN based (Anirudh et al. 2020)	34.63
Proposed DPGNet	**38.94**

computational complexity of the network. Hence, a better trade-off needs to be adopted. The use of multi-sized kernels is a viable alternative as it significantly improves the results without much affecting the execution time.

4.3.0.2. *Comparison of DTCapsNet with the commonly-used classifiers*

A comparative analysis of DTCapsNet with recently published approaches relevant to the classification of VI curves, is summarized in Table 10. The significance of the results of DTCapsNet (at a confidence level of 95%) in comparison with the benchmark methods is illustrated in Table 11. Based on the discussions in (Fawaz et al. 2018; S. Li et al. 2019; Imani and Ghassemian 2020), some main existing classifiers are selected as the benchmark methods for comparison. An analysis of the variation in the accuracy of different approaches based on the variation in the percentage of training samples is presented in Fig. 12. Besides, an illustration of the confusion matrices of PKNet and two benchmark approaches is presented in Figs. 13 and 14. As is evident from the results, DTCapsNet better models the VI curves as compared to other prominent approaches. The proper modeling of features significantly improves the generalization capability of the network and results in better classification accuracies, even with a small number of training samples. The DTW and interpolation-based convolutions facilitate the effective transformation of vectorized VI curves to a latent space that is more discriminative than the original space.

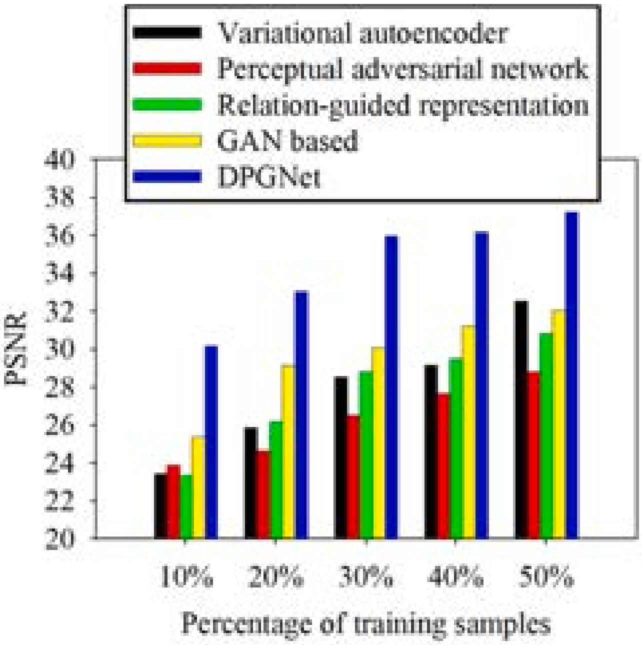


Fig. 9. Accuracy analysis of DPGNet and the benchmark reconstruction approaches with respect to the change in percentage of training samples.

The better performance of DTCapsNet in comparison with the convolutional network-based approaches, such as Jia and Liu (2020) and Hang et al. (2021), can be attributed to the effective modeling of the characteristic features of the VI curve using capsules. In addition, the DTW-based routing and convolution along with the proposed loss functions significantly improve the classification results. Although the use of spectral attention modules as in Mou and Zhu (2020) improves the

Table 8
Z-score-based significance analysis of the results of DPGNet*.

Benchmark reconstruction approaches	Z-score with respect to DPGNet
Variational autoencoder (Hoshen 2018)	2.53
Perceptual adversarial network (C. Wang et al. 2018)	2.14
Relation-guided representation (Z. Kang et al. 2020)	2.39
GAN based (Anirudh et al. 2020)	2.60

results for spectral curves, the approach did not yield good results for VI curves when the training samples are limited. The GAN-based approaches, such as Hang et al. (2021) and Jiang et al. (2020), do not yield acceptable results for a small percentage of training samples due to the inability to model the crop-specific features of the index curves. Although the LSTM-based classifier, proposed by Rubwurm and Korner (2017), attempts to model the sequential nature of the VI curves, the results indicate that the approach fails to effectively model the characteristic features.

It is worth pointing that, in contrast to the existing capsule-based approaches, DTCapsNet employs a shallow-capsule-layer strategy where the routing is modified to consider the specific characteristics of the VI curves. The DTCapsNet requires only a smaller number of training samples and is computationally more efficient as compared to the existing approaches. The absence of variational encoding and the use of adversarial loss reduces the network depth and facilitates faster convergence. Additionally, the DTW-based convolution and routing along with the VI curve-specific constraints further improve the performance.

The DTCapsNet-based approach is found to be effective in estimating the fractional area covered by different crops from the representative VI curve of a given field (Table 12). The significance of the proposed results is validated in Table 13. Instead of using an additional SoftMax layer, the normalized length of the output vector of each crop class is used as an estimate of the fractional abundance. Recent prominent unmixing approaches such as those based on GAN (Borsoi et al. 2019), convolutional autoencoder (Palsson et al. 2019), deep autoencoder (Su et al. 2019), sparse autoencoder (Dou et al. 2020), iterative shrinkage thresholding (Qian et al. 2020) and VAE (Xie et al. 2020) are chosen as the benchmark methods for comparison. The results presented in Table 12 indicate that the proposed approach outperforms the existing approaches. These results can be attributed to the effective modeling of characteristic VI curve features, resulting in an optimal discriminative latent space.

5. Discussion

Experiments over different datasets illustrate the better performance of the proposed approaches as compared to prominent existing approaches. A detailed analysis of the results of each of the proposed approaches is presented in the following subsections.

Table 9
Analysis of the alternative architectural choices of DTCapsNet for classification of index curves.

Architectural variations	Kappa statistics	Overall Accuracy
DTCapsNet without DTW-based convolutional units	0.91	95.09
DTCapsNet without DTW-based routing	0.91	93.86
DTCapsNet without interpolation-based convolution	0.92	94.62
Proposed DTCapsNet implementation	**0.93**	**97.40**

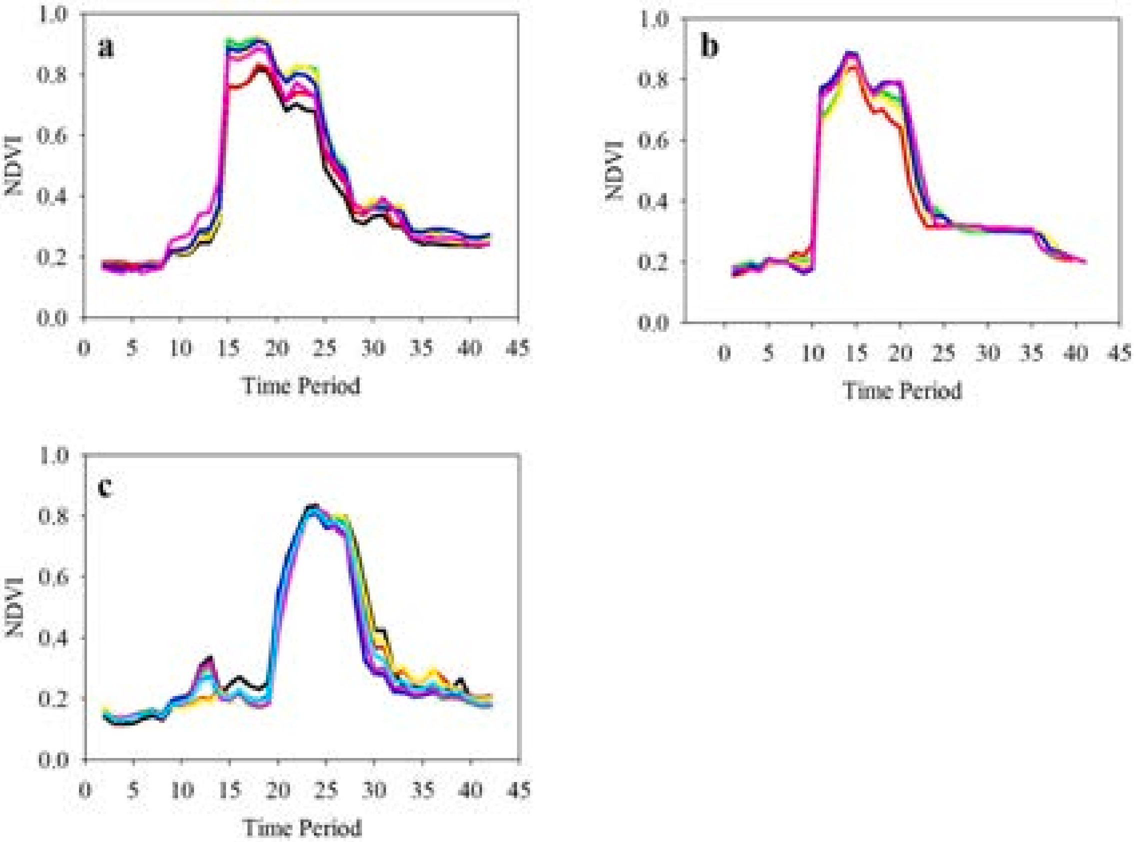


Fig. 10. Examples of field-level phenological fingerprints learned by DPGNet for randomly selected fields of (a) Barley; (b) Wheat; and (c) Potato.

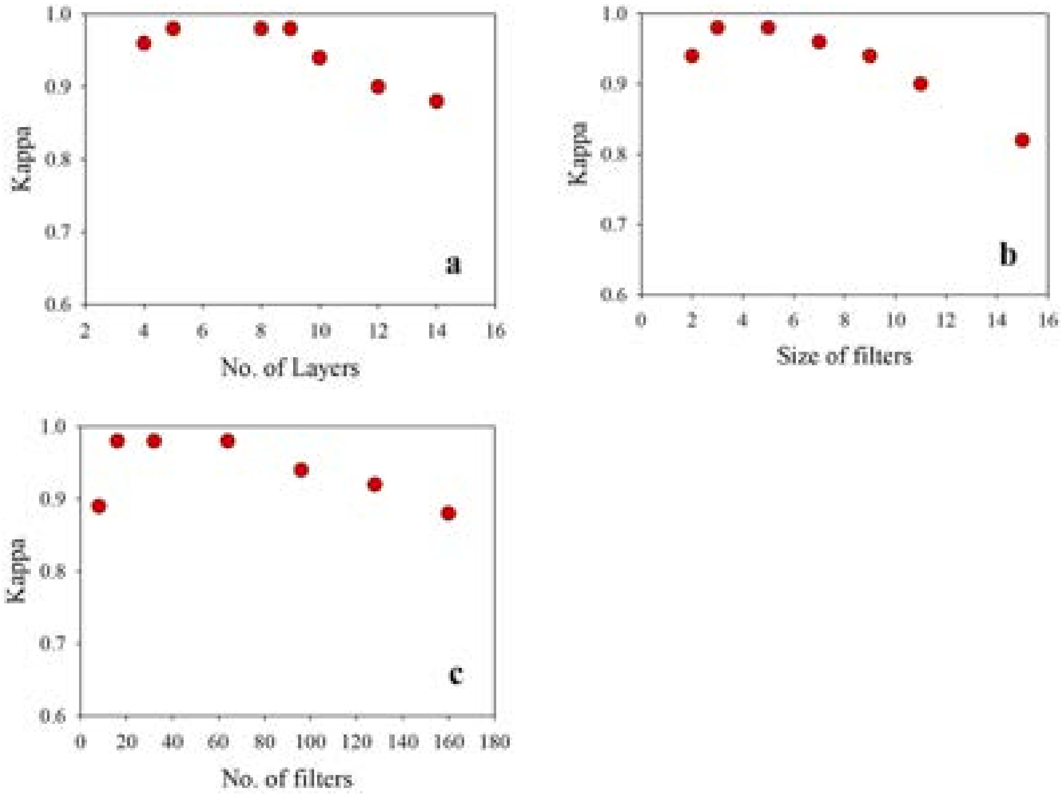


Fig. 11. Analysis of the sensitivity of DTCapsNet towards (a) depth of the network layers; (b) size of filters; and (c) number of filters for 80% of the training samples.

Table 10
Comparison of DTCapsNet with the benchmark classifiers for 60% of the training samples*.

Benchmark classifiers	Kappa statistics	Overall accuracy
LSTM based (Rubwurm and Korner 2017)	0.90	94.08
Multi-task GAN based (Hang et al., 2021)	0.89	92.37
GAN based (Jiang et al. 2020)	0.85	88.76
Spectral attention based (Mou and Zhu, 2020)	0.87	90.53
CNN based (Jia and Liu, 2020)	0.90	94.19
Proposed DTCapsNet	**0.93**	**97.40**

Table 11
Z-score based significance analysis of the results of DTCapsNet*.

Benchmark classifiers	Z-score with respect to DTCapsNet
LSTM based (Rubwurm and Korner 2017)	2.32
Multi-task GAN based (Hang et al., 2021)	2.46
GAN based (Jiang et al. 2020)	1.99
Spectral attention based (Mou and Zhu 2020)	2.18
CNN based (Jia & Liu, 2020)	2.29

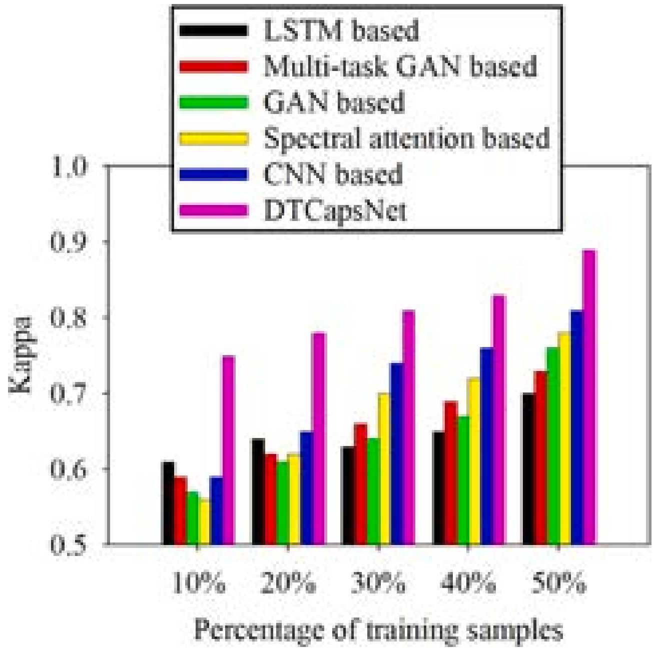
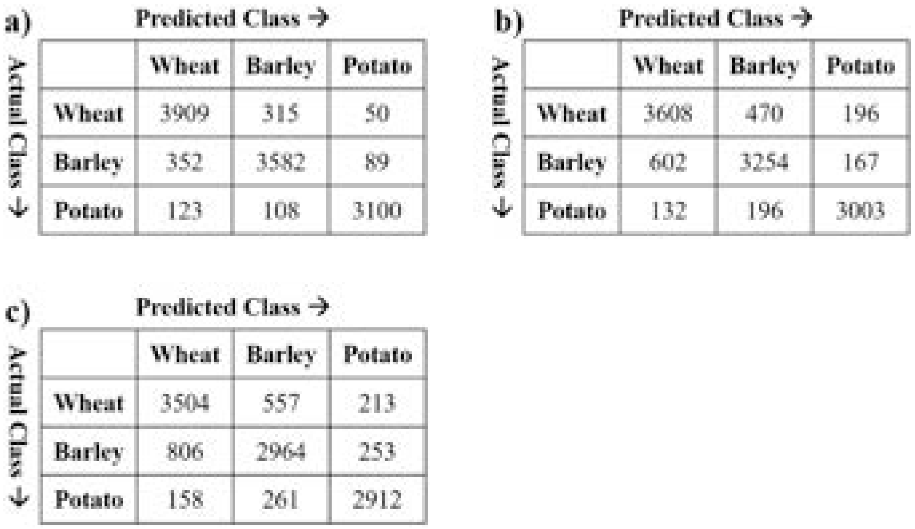


Fig. 12. Accuracy analysis of DTCapsNet and the benchmark classifiers with respect to the change in the percentage of training samples.

5.1. Smoothening of field-level NDVI curves

The interpolation in learned latent feature space, as adopted in PKNet, models the local features effectively. The multiple kernels and multi-layer abstractions facilitate the modeling of different characteristic features at different resolutions significant for proper interpolation. The network depth determines the level of abstraction, whereas the number of kernels determines the latent space dimension. Different experiments in this study illustrate that using multi-size kernels facilitates modeling features more effectively than using single-size kernels. The use of static kernels for interpolating the point data affects the results, especially when the data is irregularly sampled, as in the case of noisy NDVI curves. The interpolated convolution considers the irregularity in the data placement and does not assume a grid structure. Also, the deep convolutional architecture facilitates the learning of kernels from the data rather than the use of static kernels.

The encoding–decoding approach is more effective for NDVI reconstruction than the GAN-based approaches among the different DL-based



a) Predicted Class → Actual Class ↓

	Wheat	Barley	Potato
Wheat	3909	315	50
Barley	352	3582	89
Potato	123	108	3100

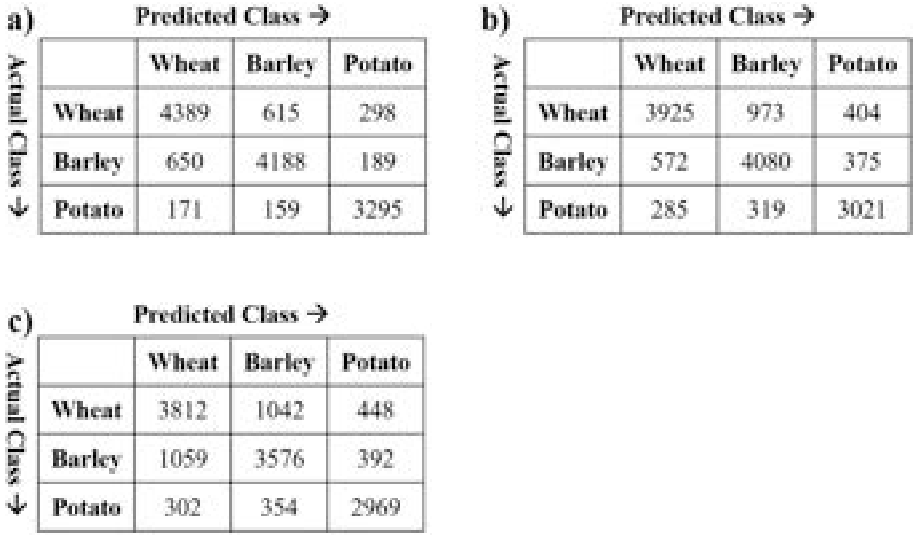
b) Predicted Class → Actual Class ↓

	Wheat	Barley	Potato
Wheat	3608	470	196
Barley	602	3254	167
Potato	132	196	3003

c) Predicted Class → Actual Class ↓

	Wheat	Barley	Potato
Wheat	3504	557	213
Barley	806	2964	253
Potato	158	261	2912

Fig. 13. Confusion matrices of (a) PKNet (b) Jia and Liu (2020) and (c) Rubwurm and Korner (2017) for 50% of the total samples as test samples.



a) Predicted Class → Actual Class ↓

	Wheat	Barley	Potato
Wheat	4389	615	298
Barley	650	4188	189
Potato	171	159	3295

b) Predicted Class → Actual Class ↓

	Wheat	Barley	Potato
Wheat	3925	973	404
Barley	572	4080	375
Potato	285	319	3021

c) Predicted Class → Actual Class ↓

	Wheat	Barley	Potato
Wheat	3812	1042	448
Barley	1059	3576	392
Potato	302	354	2969

Fig. 14. Confusion matrices of (a) PKNet (b) Jia and Liu (2020) and (c) Rubwurm and Korner (2017) for 60% of the total samples as test samples.

latent space projection strategies. Deep autoencoders have multiple convolutional and deconvolutional layers and are prone to degradation due to information loss. Skip connections in autoencoders are found to resolve the issue and facilitate effective reconstruction with even limited training samples.

Lack of generalizability usually makes the DL networks trained on one crop less effective for another crop or a different area. The data and domain bias, as well as the limited availability of training samples, also affect the effectiveness of DL approaches. The back-projection-based refinement proposed in this study effectively resolves this issue by matching the non-noisy data points in the original signal with the corresponding ones in the reconstructed version to have an accurate refinement. The conventional Euclidean losses, which consider point-wise matching, are not effective for NDVI reconstruction. The use of piece-wise dissimilarity losses, in addition to the mean-squared-error-based losses, improves reconstruction accuracy.

The use of the VAE model, instead of the model adopted in PKNet, can constraint the latent features to a normally distributed space. However, additional sampling layers in VAE are found to adversely affect the smoothing of VI curves, especially when training samples are limited.

Table 12
Comparison of DTCapsNet-based fractional area estimation with prominent unmixing approaches.

Commonly-used unmixing approaches	RMSE
(Su et al. 2019)	0.51
(Borsoi et al. 2019)	0.49
(Qian et al. 2020)	0.56
(Dou et al. 2020)	0.34
DTCapsNet-based approach	**0.12**

Table 13
Z-score-based significance analysis of the DTCapsNet-based fractional area estimation.

Commonly-used unmixing approaches	Z-score with respect to the DTCapsNet based unmixing approach
(Su et al. 2019)	2.45
(Borsoi et al. 2019)	2.32
(Qian et al. 2020)	2.13
(Dou et al. 2020)	1.98

The experiments conducted in this study use simulated samples from cloud-free VI curves, as discussed in Subsection 4.1. These simulations (adding noises and reconstructing the original) are essential for quantifying the performance of different smoothing approaches. Smoothing of real cloud-affected VI curves using PKNet and visual inspection confirms better accuracy compared to the existing approaches. The network training on a specific type of crop is more effective than training on multiple crops and can be preferred when sufficient training samples are available. Also, when multi-year data is used, and the data acquisition dates are not exact, DTW-based convolutional layers should be used instead of the normal convolution.

5.2. Generation of field-level representation

The normally distributed latent representations obtained from DPGNet provide a more effective field-level representation rather than multiple moments. The use of interpolated convolutions, DTW-based layers, skip connections, adversarial loss, and piece-wise dissimilarity loss improve the reconstruction accuracy compared to the conventional VAEs. In addition, the approach of embedding the crop-label information and input prior in the latent space also give a significant improvement in reconstruction PSNR. The mean computed in the learned latent space better represents the field-level VI curves than the same computed in the VI curve space. The generalization capability of DPGNet can be attributed to the effectiveness of the learned latent space build on the characteristic features of the VI curves. Training on a specific type of crop is better than training on multiple crops to generate field-level representations. The approach is resilient to the issues when multi-year data (with slight date shifts) is used due to the use of DTW-based convolutional layers.

5.3. Classification and fractional area estimation

The VI curve features and their characteristics, such as the depth, width, and position, are crucial in the phenology-based classification of crops. Although encoding networks designed for smoothing (Subsection 3.2.1) or VI curve generalization (Subsection 3.2.2) coupled with a fully connected network can classify VI curves, considering spectral characteristics needs a capsule-based approach. DTCapsNet uses capsules (a group of neurons) instead of neurons, and the approach is effective in learning features and their characteristics. The use of multiple kernels and multi-layer abstractions facilitates modeling features at different resolutions, while multi-sized kernels enable modeling of varied length features.

The conventional convolutional units, aggregated as capsules, do not consider the time-series nature of the input NDVI data while computing the similarity measures. The DWT-based convolutional units, which employ time wrapping similarity measures to compute weight vectors instead of scalars, resolve the issue of distortions and shifts prevalent in index-curve-based classification.

The use of capsule network has significantly improved the generalizability and the approach gives good results even with a limited number of training samples. The performance of DTCapsNet can be attributed to the effective modeling of features and their characteristics. Analysis of the activation maps indicates that the capsules learn physically significant features compared to those learned using normal convolutional units. The use of reconstruction loss, along with cross-entropy loss, is found to affect the classification accuracy. Hence, it is recommended that for the classification of time-series data such as VI curves, misclassification loss alone is effective. The approach provides good results even when trained with multi-year data (having slight shifts in features) and can be attributed to the feature-specific learning and DTW layers.

The normalized length of the output vector of class capsules gives an accurate estimate of the fractional abundances of the corresponding classes. The use of shallow capsule layers for spectral unmixing overcomes the issue of vanishing gradient, and the approach gives better results even with a limited number of training samples.

6. Conclusion

This research proposes DL-based approaches for within-season field-level monitoring of crops using VENμS satellite data. The time-series VI curves (such as NDVI), which form the basis of crop phenology analyses, are usually affected by atmospheric and sensor effects resulting in missing or erroneous data. The proposed PKNet considers relevant features of the VI curves to implement missing data imputation. The constraints and encoding scheme and the DTW-based similarity measures, achieve effective denoising with minimal training samples. Unlike the conventional convolution, a point-based convolution is proposed to process the irregularly sampled VI curves. The better results of PKNet than the prominent existing approaches can also be attributed to the improved generalizability achieved through skip connections and piece-wise dissimilarity losses. It may be noted that PKNet learns smoothing parameters and kernels dynamically from the data.

The VAE-based transformations and the adversarial constraints, adopted in DPGNet, transform the VI curves to a normally distributed latent space that is a close representation of the intrinsic manifold. The mean sampled from the unit Gaussian space learned from a group of VI curves, is reconstructed to get a characteristic representation of the group. The generalization using DPGNet has given better results than the use of statistical moments and other reconstruction techniques.

The use of capsules in the proposed DTCapsNet classifier facilitates effective modeling of spectral characteristics of the VI curves and thereby improves the generalizability of the approach. The interpolated convolution, along with DTW-based weight computation, considers irregular sampling and the series nature of the VI curves. DTCapsNet also accurately estimates the fractional area covered by different crops from a given field-level VI curve. The better unmixing results can be attributed to the feature-based modeling. Experiments on simulated and real datasets indicated that the proposed smoothing, generalization, and classification frameworks give better results than the corresponding baseline methods considered in this study.

CRediT authorship contribution statement

Arun Pattathal V: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Visualization, Software, Writing – original draft, Writing – review & editing. **Arnon Karnieli:** Conceptualization, Funding acquisition, Supervision, Data curation, Project administration, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The project was partially funded by Israeli French High Council for Scientific & Technological Cooperation, research program “Maïmonide-Israel”, contract no. 3-15832.

References

Al-Nahhal, I., Dobre, O. A., Basar, E., Moloney, C., & Ikki, S. (2019, November 1). A Fast, Accurate, and Separable Method for Fitting a Gaussian Function [Tips & Tricks]. IEEE Signal Processing Magazine. Institute of Electrical and Electronics Engineers Inc. 10.1109/MSP.2019.2927685.

Anirudh, R., Thiagarajan, J. J., Kailkhura, B., & Bremer, P. T. (2020). MimicGAN: Robust Projection onto Image Manifolds with Corruption Mimicking. *Int. J. Comput. Vision, 128*(10–11), 2459–2477. https://doi.org/10.1007/s11263-020-01310-5

Arun, P. V., & Karnieli, A. (2021). Deep Learning-Based Phenological Event Modeling for Classification of Crops. Remote Sensing 2021, Vol. 13, Page 2477, 13(13), 2477. 10.3390/RS13132477.

Baumann, M., Ozdogan, M., Richardson, A. D., & Radeloff, V. C. (2017). Phenology from Landsat when data is scarce: Using MODIS and Dynamic Time-Warping to combine multi-year Landsat imagery to derive annual phenology curves. *Int. J. Appl. Earth Observ. Geoinform., 54*, 72–83. https://doi.org/10.1016/j.jag.2016.09.005

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell., 35*(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Bochinski, E., Senst, T., & Sikora, T. (2018). Hyper-parameter optimization for convolutional neural network committees based on evolutionary algorithms. In Proceedings - International Conference on Image Processing, ICIP (Vol. 2017-September, pp. 3924–3928). IEEE Computer Society. 10.1109/ICIP.2017.8297018.

Borsoi, R. A., Imbiriba, T., & Bermudez, J. C. M. (2019). Deep Generative Endmember Modeling: An Application to Unsupervised Spectral Unmixing. *IEEE Trans. Comput. Imaging, 6*, 374–384. https://doi.org/10.1109/TCI.2019.2948726

Ca, P. V., Edu, L. T., Lajoie, I., Ca, Y. B., & Ca, P.-A.-M. (2010). Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pascal Vincent Hugo Larochelle Yoshua Bengio Pierre-Antoine Manzagol. *J. Mach. Learn. Res., 11*.

Cai, X., Xu, T., Yi, J., Huang, J., & Rajasekaran, S. (2019). DTWNet: A dynamic time warping network. In Advances in Neural Information Processing Systems (Vol. 32).

Cai, Z., Jönsson, P., Jin, H., & Eklundh, L. (2017). Performance of Smoothing Methods for Reconstructing NDVI Time-Series and Estimating Vegetation Phenology from MODIS Data. *Remote Sens., 9*(12), 1271. https://doi.org/10.3390/rs9121271

Cao, R., Chen, Y., Shen, M., Chen, J., Zhou, J.i., Wang, C., et al. (2018). A simple method to improve the quality of NDVI time-series data by integrating spatiotemporal information with the Savitzky-Golay filter. *Remote Sens. Environ., 217*, 244–257. https://doi.org/10.1016/j.rse.2018.08.022

Chai, X., Gu, H., Li, F., Duan, H., Hu, X., & Lin, K. (2020). Deep learning for irregularly and regularly missing data reconstruction. *Sci. Rep., 10*(1), 1–18. https://doi.org/10.1038/s41598-020-59801-x

Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., & Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ., 91*(3–4), 332–344. https://doi.org/10.1016/j.rse.2004.03.014

Cheriyadat, A. M. (2014). Unsupervised feature learning for aerial scene classification. *IEEE Trans. Geosci. Remote Sens., 52*(1), 439–451. https://doi.org/10.1109/TGRS.2013.2241444

Cubuk, E. D., Zoph, B., Shlens, J., & Le, Q. V. (2019). RandAugment: Practical automated data augmentation with a reduced search space. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2020-June, 3008–3017. http://arxiv.org/abs/1909.13719. Accessed 25 October 2020.

Cuturi, M., & Blondel, M. (2017). Soft-DTW: a Differentiable Loss Function for Time-Series. In 34th International Conference on Machine Learning, ICML 2017 (Vol. 2, pp. 1483–1505). International Machine Learning Society (IMLS). http://arxiv.org/abs/1703.01541. Accessed 24 November 2020.

Dou, Z., Gao, K., Zhang, X., Wang, H., & Wang, J. (2020). Hyperspectral unmixing using orthogonal sparse prior-based autoencoder with hyper-laplacian loss and data-driven outlier detection. *IEEE Trans. Geosci. Remote Sens., 58*(9), 6550–6564. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2020.2977819

Emami, H., Aliabadi, M. M., Dong, M., & Chinnam, R. B. (2019). SPA-GAN: Spatial Attention GAN for Image-to-Image Translation. *IEEE Trans. Multimedia*, 1 Accessed 25 October 2020 http://arxiv.org/abs/1908.06616.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., & Muller, P.-A. (2018). Deep learning for time series classification: A review. *Data Min. Knowl. Disc., 33*(4), 917–963. https://doi.org/10.1007/s10618-019-00619-1

Foerster, S., Kaden, K., Foerster, M., & Itzerott, S. (2012). Crop type mapping using spectral-temporal profiles and phenological information. *Comput. Electron. Agric., 89*, 30–40. https://doi.org/10.1016/j.compag.2012.07.015

Gebbers, R., & Adamchuk, V. I. (2010, February 12). Precision agriculture and food security. Science. American Association for the Advancement of Science. 10.1126/science.1183899.

Girin, L., Leglaive, S., Bie, X., Diard, J., Hueber, T., & Alameda-Pineda, X. (2020). Dynamical Variational Autoencoders: A Comprehensive. *Review.* Accessed 25 October 2020 http://arxiv.org/abs/2008.12595.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2020). A Review on Generative Adversarial Networks: Algorithms. *Theory, and Applications, 14*(8) Accessed 25 October 2020 http://arxiv.org/abs/2001.06937.

Gulcu, A., & Kus, Z. (2020). Hyper-Parameter Selection in Convolutional Neural Networks Using Microcanonical Optimization Algorithm. *IEEE Access, 8*, 52528–52540. https://doi.org/10.1109/Access.628763910.1109/ACCESS.2020.2981141

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2020). Deep Learning for 3D Point Clouds: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell., 1–1*. https://doi.org/10.1109/tpami.2020.3005434

Han, P., Li, G., Skulstad, R., Skjong, S., & Zhang, H. (2020). A Deep Learning Approach to Detect and Isolate Thruster Failures for Dynamically Positioned Vessels Using Motion Data. *IEEE Trans. Instrum. Meas., 1–1*. https://doi.org/10.1109/tim.2020.3016413

Hasanzadeh, A., Hajiramezanali, E., Duffield, N., Narayanan, K., Zhou, M., & Qian, X. (2019). Semi-Implicit Graph Variational Auto-Encoders. https://github.com/sigvae/SIGraphVAE. Accessed 25 October 2020.

Hang, R., Zhou, F., Liu, Q., & Ghamisi, P. (2021). Classification of Hyperspectral Images via Multitask Generative Adversarial Networks. *IEEE Transactions on Geoscience and Remote Sensing, 59*(2), 1424–1436. https://doi.org/10.1109/TGRS.2020.3003341

Herrmann, I., Shapira, U., Kinast, S., Karnieli, A., & Bonfil, D. J. (2013). Ground-level hyperspectral imagery for detecting weeds in wheat fields. *Precis. Agric., 14*(6), 637–659. https://doi.org/10.1007/s11119-013-9321-x

Hoshen, Y. (2018). Non-adversarial mapping with VAES. In Advances in Neural Information Processing Systems (Vol. 2018-Decem, pp. 7528–7537).

Im, D. J., Ahn, S., Memisevic, R., & Bengio, Y. (2015). Denoising Criterion for Variational Auto-Encoding Framework. 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2059–2065. http://arxiv.org/abs/1511.06406. Accessed 26 October 2020.

Imani, M., & Ghassemian, H. (2020). An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges. *Information Fusion, 59*, 59–83. https://doi.org/10.1016/j.inffus.2020.01.007

Iwana, B. K., Frinken, V., & Uchida, S. (2020). DTW-NN: A novel neural network for time series recognition using dynamic alignment between inputs and weights. *Knowl.-Based Syst., 188*, 104971. https://doi.org/10.1016/j.knosys.2019.104971

Jia, Xiaojun, & Liu, Zihao (2020). Element extraction and convolutional neural network-based classification for blue calico. *Textile Research Journal, 91*(3), 261–277. https://doi.org/10.1177/0040517520939573

Jiang, T., Li, Y., Xie, W., & Du, Q. (2020). Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens., 58*(7), 4666–4679. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2020.2965961

Julien, Y., & Sobrino, J. A. (2019). Optimizing and comparing gap-filling techniques using simulated NDVI time series from remotely sensed global data. *Int. J. Appl. Earth Obs. Geoinf., 76*, 93–111. https://doi.org/10.1016/j.jag.2018.11.008

Kamir, E., Waldner, F., & Hochman, Z. (2020). Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. *ISPRS J. Photogramm. Remote Sens., 160*, 124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008

Kang, Z., Lu, X., Liang, J., Bai, K., & Xu, Z. (2020). Relation-Guided Representation Learning. Neural Networks, 131, 93–102. http://arxiv.org/abs/2007.05742. Accessed 26 October 2020.

Karim, F., Majumdar, S., Darabi, H., & Harford, S. (2018). Multivariate LSTM-FCNs for Time Series Classification. *Neural Networks, 116*, 237–245. https://doi.org/10.1016/j.neunet.2019.04.014

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational bayes. In 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings. International Conference on Learning Representations, ICLR. https://arxiv.org/abs/1312.6114v10. Accessed 26 October 2020.

Kipf, T. N., & Welling, M. (2016). Variational Graph Auto-Encoders 1 A latent variable model for graph-structured data.

Kolbæk, M., Tan, Z.-H., Jensen, S. H., & Jensen, J. (2019). On Loss Functions for Supervised Monaural Time-Domain Speech Enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process., 28*, 825–838. https://doi.org/10.1109/TASLP.2020.2968738

Kong, D., Zhang, Y., Gu, X., & Wang, D. (2019). A robust method for reconstructing global MODIS EVI time series on the Google Earth Engine. *ISPRS J. Photogramm. Remote Sens., 155*, 13–24. https://doi.org/10.1016/j.isprsjprs.2019.06.014

Lai, X., Wu, X., Zhang, L., Lu, W., & Zhong, C. (2019). Imputations of missing values using a tracking-removed autoencoder trained with incomplete data. *Neurocomputing, 366*, 54–65. https://doi.org/10.1016/j.neucom.2019.07.066

Li, S., Song, W., Fang, L., Chen, Y., Ghamisi, P., & Benediktsson, J. A. (2019). Deep learning for hyperspectral image classification: An overview. *IEEE Trans. Geosci. Remote Sens., 57*(9), 6690–6709. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2019.2907932

Li, Y., Tarlow, D., Brockschmidt, M., & Zemel, R. (2015). Gated Graph Sequence Neural Networks. 4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings, (1), 1–20. http://arxiv.org/abs/1511.05493. Accessed 26 October 2020.

Ma, L., Liu, Y.u., Zhang, X., Ye, Y., Yin, G., & Johnson, B. A. (2019). June 1). Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.. Elsevier B.V., 152*, 166–177. https://doi.org/10.1016/j.isprsjprs.2019.04.015

Mao, J., Wang, X., & Li, H. (2019). Interpolated Convolutional Networks for 3D Point Cloud Understanding. Proceedings of the IEEE International Conference on Computer Vision, 2019-October, 1578–1587. http://arxiv.org/abs/1908.04512. Accessed 26 October 2020.

Mou, L., & Zhu, X. X. (2020). Learning to Pay Attention on Spectral Domain: A Spectral Attention Module-Based Convolutional Network for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens., 58*(1), 110–122. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2019.2933609

Palsson, B., Ulfarsson, M. O., & Sveinsson, J. R. (2019). Convolutional Autoencoder for Spatial-Spectral Hyperspectral Unmixing (pp. 357–360). Institute of Electrical and Electronics Engineers (IEEE). 10.1109/igarss.2019.8900297.

Patterson, N. K., Lane, B. A., Sandoval-Solis, S., Pasternack, G. B., Yarnell, S. M., & Qiu, Y. (2020). A hydrologic feature detection algorithm to quantify seasonal components of flow regimes. *J. Hydrol., 585*, 124787. https://doi.org/10.1016/j.jhydrol.2020.124787

Peng, X., Zhu, H., Feng, J., Shen, C., Zhang, H., & Zhou, J. T. (2019). Deep Clustering With Sample-Assignment Invariance Prior. *IEEE Trans. Neural Networks Learn. Syst., 1–12*. https://doi.org/10.1109/tnnls.2019.2958324

Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). PointNet: Deep learning on point sets for 3D classification and segmentation. In Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 (Vol. 2017-January, pp.

77–85). Institute of Electrical and Electronics Engineers Inc. 10.1109/CVPR.2017.16.

Qian, Y., Xiong, F., Qian, Q., & Zhou, J. (2020). Spectral Mixture Model Inspired Network Architectures for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens., 58*(10), 7418–7434. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2020.2982490

Richardson, A. D., Anderson, R. S., Arain, M. A., Barr, A. G., Bohrer, G., Chen, G., et al. (2012). Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program Site Synthesis. *Glob. Change Biol., 18*(2), 566–584. https://doi.org/10.1111/j.1365-2486.2011.02562.x

Rubwurm, M., & Korner, M. (2017). Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (Vol. 2017-July, pp. 1496–1504). 10.1109/CVPRW.2017.193.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic Routing Between Capsules. Advances in Neural Information Processing Systems, 2017-December, 3857–3867. http://arxiv.org/abs/1710.09829. Accessed 26 October 2020.

Shekhar, C. (2016). On simplified application of multidimensional Savitzky-Golay filters and differentiators. In AIP Conference Proceedings (Vol. 1705, p. 020014). American Institute of Physics Inc. 10.1063/1.4940262.

Shi, Y., Davaslioglu, K., Sagduyu, Y. E., Headley, W. C., Fowler, M., & Green, G. (2019). Deep Learning for RF Signal Classification in Unknown and Dynamic Spectrum Environments. 2019 IEEE International Symposium on Dynamic Spectrum Access Networks, DySPAN 2019. http://arxiv.org/abs/1909.11800. Accessed 26 October 2020.

Sandfort, Veit, Yan, Ke, Pickhardt J, Perry, & Summers M, Ronald (2019). Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports, 9*(11), Article 16884. https://doi.org/10.1038/s41598-019-52737-x

Su, Y., Li, J., Plaza, A., Marinoni, A., Gamba, P., & Chakravortty, S. (2019). DAEN: Deep Autoencoder Networks for Hyperspectral Unmixing. *IEEE Trans. Geosci. Remote Sens., 57*(7), 4309–4321. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2018.2890633

Tian, Y., Peng, X., Zhao, L., Zhang, S., & Metaxas, D. N. (2018). CR-GAN: Learning Complete Representations for Multi-view Generation. IJCAI International Joint Conference on Artificial Intelligence, 2018-July, 942–948. http://arxiv.org/abs/1806.11191. Accessed 26 October 2020.

Tschannen, M., Bachem, O., & Lucic, M. (2018). Recent Advances in Autoencoder-Based Representation Learning. http://arxiv.org/abs/1812.05069. Accessed 26 October 2020.

Wang, H., Wang, J., Wang, J., Zhao, M., Zhang, W., Zhang, F., et al. (2019). Learning Graph Representation with Generative Adversarial Nets. *IEEE Trans. Knowl. Data Eng.*. https://doi.org/10.1109/TKDE.2019.2961882

Wang, Y., Sun, Y., Liu, Z., Sarma, S. E., Bronstein, M. M., & Solomon, J. M. (2018). Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics, 38(5), Article 146. http://arxiv.org/abs/1801.07829. Accessed 26 October 2020.

Weil, G., Lensky, I. M., & Levin, N. (2017). Using ground observations of a digital camera in the VIS-NIR range for quantifying the phenology of Mediterranean woody species. *Int. J. Appl. Earth Obs. Geoinf., 62*, 88–101. https://doi.org/10.1016/j.jag.2017.05.016

Xiang, M., Yu, Q., & Wu, W. (2019). From multiple cropping index to multiple cropping frequency: Observing cropland use intensity at a finer scale. *Ecol. Ind., 101*, 892–903. https://doi.org/10.1016/j.ecolind.2019.01.081

Xie, W., Lei, J., Yang, J., Li, Y., Du, Q., & Li, Z. (2020). Deep Latent Spectral Representation Learning-Based Hyperspectral Band Selection for Target Detection. *IEEE Trans. Geosci. Remote Sens., 58*(3), 2015–2026. https://doi.org/10.1109/TGRS.3610.1109/TGRS.2019.2952091

Yan, L., & Roy, D. P. (2020). Spatially and temporally complete Landsat reflectance time series modelling: The fill-and-fit approach. *Remote Sens. Environ., 241*, 111718. https://doi.org/10.1016/j.rse.2020.111718

Yang, L., Huang, W.u., & Sun.. (2019). Weighted Double-Logistic Function Fitting Method for Reconstructing the High-Quality Sentinel-2 NDVI Time Series Data Set. *Remote Sens., 11*(20), 2342. https://doi.org/10.3390/rs11202342

Zeng, Linglin, Wardlow, Brian D., Xiang, Daxiang, Hu, Shun, & Li, Deren (2020). A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sens. Environ., 237*, 111511. https://doi.org/10.1016/j.rse.2019.111511

Zhang, Zhiwen, Duan, Feng, Sole-Casals, Jordi, Dinares-Ferran, Josep, Cichocki, Andrzej, Yang, Zhenglu, et al. (2019). A Novel Deep Learning Approach with Data Augmentation to Classify Motor Imagery Signals. *IEEE Access, 7*, 15945–15954. https://doi.org/10.1109/ACCESS.2019.2895133

Zhao, Kaiguang, Wulder, Michael A., Hu, Tongxi, Bright, Ryan, Wu, Qiusheng, Qin, Haiming, et al. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sens. Environ., 232*, 111181. https://doi.org/10.1016/j.rse.2019.04.034

Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F., & Fraundorfer, F. (2017, December 1). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine. Institute of Electrical and Electronics Engineers Inc. 10.1109/MGRS.2017.2762307.