Computers and Electronics in Agriculture 209 (2023) 107833

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Cost-efficient coupled learning methods for recovering near-infrared information from RGB signals: Application in precision agriculture

Alexandros Gkillas *, Dimitrios Kosmopoulos, Kostas Berberidis

Computer Engineering and Informatics Department, University of Patras, Greece

ARTICLE INFO

Keywords:
Near-infrared images
RGB images
Coupled dictionary learning
Coupled autoencoders
Multispectral imaging
Hyperspectral imaging

ABSTRACT

Multispectral imaging and the derived spectral analysis offer useful tools for revealing beneficial information for a variety of applications, e.g., precision agriculture, medical imaging and autonomous driving. Contrary to mainstream RGB cameras that can capture information derived only from three bands within the visible spectrum, the multispectral cameras can offer better spectral resolution by utilizing the underlying information in the visible and the near-infrared spectrum. However, the cost of the multispectral cameras is very high and their mobility is limited due to their weight and their need for special hardware equipment. Considering the aforementioned limitations, we propose two low-cost and efficient methods to infer detailed spectral information outside the visible spectrum range by employing only an RGB camera. The proposed methods require significantly less training data, containing approximately **99.8%** less parameters compared to the competing deep learning approaches and can be deployed on various edge devices with computational and power constraints, e.g., mobile phones or unmanned drones for addressing problems in precision agriculture under real-field settings. Extensive numerical results demonstrate the efficacy of the proposed models to reconstruct images outside the visible spectrum. Additionally, the reconstructed images can be utilized to estimate the normalized difference vegetation index (NDVI), which can reveal valuable information concerning the health of the monitored plants without the need of a multispectral camera.

1. Introduction

In recent years, multispectral imaging and the associated spectral analysis constitute remarkable tools for revealing and capturing information for a variety of applications across numerous domains. Different from the mainstream conventional RGB imaging systems that are able to capture information derived only from three spectral bands within the visible electromagnetic spectrum, multispectral cameras can provide much more detailed spectral resolution, by utilizing the underlying information that lies both in the visible and the near-infrared spectrum. Thus, this detailed spectral information can be beneficial to numerous challenging problems ranging from remote sensing, medical imaging to autonomous driving and precision agriculture (Kussul et al., 2017; Calvini et al., 2019; Charte et al., 2018).

As far as the precision agriculture domain is concerned, the employment of multispectral imaging systems can be beneficial for both the environment and the farmers. The spectral signatures of the soil and the plants can be investigated in different wavelengths of the electromagnetic spectrum at different production phases, thus allowing crop monitoring, while reducing the use of excess water, pesticides, and fertilizers (Garrett et al., 2006; Sladojevic et al., 2016; Chakraborty et al., 2000). Advances in signal processing and learning have provided an opportunity to extend and ameliorate automated systems so as to monitor plant health and identify pathogens at the earliest stages of their manifestation (Arnal Barbedo, 2013; Mahlein, 2015). In view of this, multispectral (Mahlein, 2015) imaging provides new insights into the complicated pathogen–host system, enabling researchers to investigate the reflectance properties of the plants in various wavelengths of the electromagnetic spectrum (visible, near-infrared), thereby leading to more efficient methods for plant disease detection from the laboratory to greenhouses (Mahlein, 2015; Fahrentrapp et al., 2019).

As mentioned before multispectral sensors can capture images from numerous spectral bands inside the electromagnetic spectrum. Concerning the plant monitoring applications, spectral bands inside the visible spectral range (i.e., 400 nm – 700 nm) provide adequate information regarding the plant health. In more detail, due to the chlorophyll, a healthy plant exhibits the strongest reflectance in the green spectral band, and absorbs more radiation in the blue and red bands in order to create more chlorophyll (Ustin and Jacquemoud, 2020). However, one of the most informative spectral bands with the strongest reflectance

* Corresponding author.
E-mail addresses: st1003586@ceid.upatras.gr (A. Gkillas), dkosmo@upatras.gr (D. Kosmopoulos), berberid@ceid.upatras.gr (K. Berberidis).

https://doi.org/10.1016/j.compag.2023.107833
Received 28 September 2022; Received in revised form 26 February 2023; Accepted 2 April 2023
Available online 24 April 2023

properties for monitoring the vegetation are those inside the near-infrared spectrum. In the case, when a plant is attacked by a pest and it is under stress, a strong variation of the reflectance of these bands can be observed (Ustin and Jacquemoud, 2020). Finally, near-infrared bands along with the red bands can be employed for estimating vegetation indices, e.g., the most widely used vegetation indices is the so-called normalized difference vegetation index (NDVI) (Jiang et al., 2006), which can reveal valuable information concerning the plant health (Mahlein et al., 2013).

Despite the high utility of multispectral technology, it goes along with several complications that limit its application in precision agriculture (Gkillas et al., 2021). To be more specific, multispectral imaging cameras are expensive, and they exhibit various mobility limitations due to their weight and size and the need for specialized hardware equipment, requiring expertise for its operation. Additionally, due to the fact that the multispectral sensors capture multiple spectral bands, in order to maintain a high signal-to-noise-ratio of each band, the exposure time is long. Furthermore, compared to the conventional RGB sensors, multispectral cameras often suffer from spatial and temporal resolution issues. This can be attributed to the fact that the high spectral resolution of the multispectral cameras reduces the density of the photons at the sensor (Akhtar and Mian, 2020). Overall, these limitations may constitute an obstacle for applications outside the controlled environment of a laboratory.

Over the last years, great efforts have been devoted to the problem of reconstructing spectral bands inside the visible electromagnetic spectrum (i.e., $400-700$ nm) using only RGB images. The literature on this problem is rich ranging from studies that employ shallow learning models to studies that employ very deep learning architectures. On the contrary, very few attempts have been made towards reconstructing information outside the visible spectrum (i.e., near-infrared) based on RGB signals only. Actually, as will be shown later, such a reconstruction would be particularly useful to many practical applications. In this section, we first provide a representative sample of methods dealing with the visible spectrum and then we discuss the limited literature concerned with the NIR spectrum. Particularly, Jia et al. (2017) a radial basis network was employed to model the mapping function between the RGB and multispectral signals. Arad and Ben-Shahar (2016) a dictionary learning approach was proposed to recover hyperspectral images from RGB measurements. Moreover, Wu et al. (2017) propose a shallow learning method employing an anchored neighborhood regression model to recover hyperspectral information.

In recent years, various deep learning approaches have been proposed (Alvarez-Gila et al., 2017; Zhu et al., 2021b; Li et al., 2020; Peng et al., 2020; Kaya et al., 2019; Zhu et al., 2021a; Yan et al., 2020; Zhao et al., 2020; Li et al., 2020). Shi et al. (2018) employ effectively a residual convolutional neural network for the hyperspectral reconstruction problem. Zhao et al. (2020) introduce a hierarchical regression network to recover hyperspectral information from RGB signals. Li et al. (2020) propose an adaptive weighted attention model for recovering multispectral information using the sensitivity function of the RGB camera. Yan et al. (2020) propose a U-net-based model to reconstruct spectral signals from RGB images.

It should be emphasized that the above studies aim only to recover spectral bands inside the visible spectrum. Nonetheless, in various settings and real-world applications, e.g., plant health monitoring (Jiang et al., 2006; Fahrentrapp et al., 2019) and medical imaging (Lu and Fei, 2014), the spectral bands that belong in the near-infrared spectrum contain the most valuable information, whereas the information derived from the bands inside the visible spectral range is limited. In the literature, the problem of reconstructing near-infrared information from RGB signals is highly underexplored. To our knowledge, only few studies investigate the estimation of spectral bands outside the visible spectrum. In more detail, in preliminary study (Brown and Süsstrunk, 2011), the authors modified a conventional SLR camera by removing the infrared blocking filter in order to capture near infrared information. The near infrared response along with the response of camera in the visible spectrum improved significantly the performance of a SIFT descriptor in a scene-recognition application. Moreover, in study (Sharma and Hefeeda, 2020) a deep learning approach was employed that focuses on a medical application regarding the vein visualization problem. In study (Aslahishahri et al., 2021) based on the cGAN architecture (Isola et al., 2017) predicted near infrared images from RGB aerial images of crops. Additionally, in study (Zhang et al., 2022) a two-step GAN based methodology was used comprised of a ResUp-Down block to reconstruct multispectral images and generate NDVI indices. Finally, in Abady et al. (2020) based on the Nice-GAN methodology (Chen et al., 2020), a deep learning methodology was employed to recover spectral information using RGB signals. Note that all these methods along with the proposed ones do not learn the mapping from RGB to NIR data based on any spectroscopic assumptions. The images in these papers have strong self-similarities properties, which enables the mapping from RGB to NIR in a data-driven fashion. Aslahishahri et al. (2021)

As the information captured by RGB cameras does not practically extend to the near-infrared range, it would be desirable to have methods to reconstruct the near-infrared bands from RGB. These methods should take into account the fact that in many real-field agriculture applications, e.g., in plant monitoring in a greenhouse, the use of advanced equipment/hardware (e.g., GPUs) may be cumbersome and not very practical. Furthermore, it would also be desirable to have methods that do not require vast amounts of data, like the deep learning approaches (see, e.g., Sharma and Hefeeda, 2020; Zhang et al., 2022; Abady et al., 2020).

Such methods ideally would be efficiently deployed to various edge devices with rather limited computational and storage capabilities, e.g., mobile phones, or unmanned aerial vehicles for addressing real-field problems in precision agriculture (O'Grady et al., 2019). In this work we propose two such methods. In Gkillas et al. (2021) we presented such a shallow method that employs two coupled dictionaries to model the fundamental relation between the two spaces, where the first dictionary aims to represent the signals in the RGB space and the second dictionary seeks to represent the signals in the spectral space. Here we elaborate further on that method providing more insight as well as additional experiments and, moreover, we introduce a new auto-encoder based method.

Overall, the proposed methods not only offer competitive or better results compared to the very deep approach of Sharma and Hefeeda (2020), Zhang et al. (2022) and Abady et al. (2020), but also offer significant practical benefits. More specifically, our methods require significantly less training data, without demanding advanced hardware for the training and inference processes, thus rendering our methods ideal for edge computing applications with limited resources.

The key objectives of the proposed work can be summarized in the following:

- To propose a novel lightweight dictionary learning-based method to provide effective spectral reconstruction of spectral bands beyond the visible spectrum by relying only on visible (RGB) data. A preliminary version of this work appeared in Gkillas et al. (2021).
- To propose a second lightweight method to solve the same problem using a shallow autoencoder architecture.
- To apply the proposed methods to obtain spectral data for a plant disease detection task in a real greenhouse setting by calculating the NDVI index, which has proved to be very useful for monitoring plant health.

In the next section we present the proposed methods to achieve these objectives. In Section 3 we demonstrate the experimental results in comparison to other methods. In Section 4 we present our conclusions.

2. Methods

2.1. Problem formulation

Consider an RGB image, denoted as $\boldsymbol{I}_{rgb} \in \mathbb{R}^{n1\times n2\times 3}$. The goal of this study is to recover accurately and efficiently a near-infrared image $\boldsymbol{I}_{nir} \in \mathbb{R}^{n1\times n2\times B}$ corresponding to the same natural scene, where B corresponds to the spectral dimension (wavelength channels). Thus, our focus is to enhance the spectral span of the RGB image by estimating the more informative bands in the near-infrared spectrum, which provide valuable information concerning the health of the examined plants. Concretely, the above problem can be formulated as follows:

$$\boldsymbol{I}_{nir} = \mathcal{F}(\boldsymbol{I}_{rgb}), \tag{1}$$

where $\mathcal{F}$ stands for the mapping function that describes the fundamental relationship between the RGB space (i.e., source space) and the near-infrared space (i.e., target space).

Having two different but strongly related (coupled) image spaces (i.e., the RGB and the near-infrared space), this scenario can be considered as a image-to-image translation problem (Pang et al., 2022) aiming to effectively model the unknown mathematical function $\mathcal{F}(\cdot)$ defined in (1) that couples the two spaces. To this end, two novel image-to-image translation approaches are proposed based on a coupled dictionary and a coupled autoencoder learning framework, respectively to tackle this challenging problem.

In the following, two coupled learning methods are proposed to provide effective spectral reconstruction for spectral bands beyond the visible spectrum based on RGB measurements. In particular, utilizing the inherent sparsity of the natural images, the first method, called **CDLNR** (Coupled Dictionary Learning Near-infrared Reconstruction) employs two coupled dictionaries to model the fundamental relation between the two spaces, where the first dictionary aims to represent the signals in the RGB space and the second dictionary seeks to represent the signals in the spectral space. On the other hand, the second method called **CANR** (Coupled Autoencoder Near-infrared Reconstruction) employs two coupled autoencoders in order to reveal and capture the complex relations between the RGB and spectral space in a deep and non-linear manner.

2.2. Coupled dictionary learning — preliminaries

The coupled dictionary learning framework can be considered as a domain adaptation process (or image-to-image translation process for image data) aiming to model the signals that belong in the source (i.e., RGB data) and target domain (i.e., near-infrared data), so that the sparse representation of the data in the RGB domain can be used to describe the signals in the target domain. Formally, the coupled dictionary learning problem can be defined as the process of learning a pair of correlated dictionaries (overcomplete matrices), in such a way that the data in the RGB domain and the data in the spectral domain can be described by the respective dictionary and a common sparse representation matrix as shown by the following equations

$$\boldsymbol{X} = \boldsymbol{D}_x\boldsymbol{G} \quad \& \quad \boldsymbol{Y} = \boldsymbol{D}_y\boldsymbol{G}, \tag{2}$$

where $\boldsymbol{X}$ and $\boldsymbol{Y}$ denote the data in the RGB and spectral space, $\boldsymbol{D}_x$, $\boldsymbol{D}_y$ are the corresponding dictionaries and $\boldsymbol{G}$ stands for the sparse representation matrix. In other words, the common sparse representation matrix $\boldsymbol{G}$ can be interpreted as the bridge between the two spaces, thus enabling the transferring of valuable information from the source domain to the target domain.

2.2.1. The CDLNR method

Consider an RGB dataset $\boldsymbol{X} : \{x_i\}_{i=1}^N$, where $x_i \in \mathbb{R}^d$ is a raw RGB patch (size $p\times p\times 3$ and held as a column vector of length $d = p\times p\times 3$), and the corresponding near-infrared dataset $\boldsymbol{Y} : \{y_i\}_{i=1}^N$, where $y_i \in \mathbb{R}^l$ is a raw near-infrared patch (size $p\times p\times B$ and held as a column vector of length $l = p\times p\times B$). To effectively learn the optimal coupled dictionaries $\boldsymbol{D}_x \in \mathbb{R}^{d\times K}$ and $\boldsymbol{D}_y \in \mathbb{R}^{l\times K}$ for the RGB and near-infrared data, respectively, we propose an optimization problem that involves the required data consistency terms and the sparsity-promoting $L1$ norm. Hence, the problem is given by

$$\min_{\boldsymbol{D}_x,\boldsymbol{D}_y,\boldsymbol{G}} \left\|\boldsymbol{X} - \boldsymbol{D}_x\boldsymbol{G}\right\|_F^2 + \left\|\boldsymbol{Y} - \boldsymbol{D}_y\boldsymbol{G}\right\|_F^2 + \lambda \left\|\boldsymbol{G}\right\|_1, \tag{3}$$

where $G \in \mathbb{R}^{K\times N}$ stands for the sparse representation matrix.

Nonetheless, taking into account the fact that in most cases the RGB information does not extend to the near-infrared spectrum, recovering spectral bands beyond the visible spectrum is a highly ill-posed problem. In view of this, to effectively reveal and model the complex relationships between the near-infrared and RGB signals, we employed a cluster-based framework, as shown in Fig. 1. In greater detail, instead of learning only one pair of coupled dictionaries (i.e., $\boldsymbol{D}_x$, $\boldsymbol{D}_y$) to describe the whole dataset, namely the RGB and near-infrared data, in this study, we group the dataset into clusters and we learn multiple smaller coupled dictionaries containing more compact which represent the specific structure of the data of each cluster. In other words, each cluster contains similar near-infrared signals and the corresponding RGB data, and thus, we learn coupled dictionaries for each cluster independently.

In particular, let $\boldsymbol{X}^c$ and $\boldsymbol{Y}^c$ denote the raw RGB and the near-infrared data in the cth cluster. The corresponding coupled dictionaries (i.e., $\boldsymbol{D}_x^c$ and $\boldsymbol{D}_y^c$) can be learnt by solving the proposed optimization problem in (3). Nevertheless, instead of learning the dictionaries jointly, we start with the RGB data $\boldsymbol{X}^c$ constructing the RGB dictionary $\boldsymbol{D}_x^c$ based on the classic single feature space dictionary learning problem,

$$\min_{\boldsymbol{D}_x^c,\boldsymbol{G}_c} \left\|\boldsymbol{X}^c - \boldsymbol{D}_x^c\boldsymbol{G}_c\right\|_F^2 + \lambda \left\|\boldsymbol{G}_c\right\|_1 \tag{4}$$

The problem in relation (4) can be effectively tackled via the K-SVD dictionary learning algorithm (Aharon et al., 2006), whereas the sparse coding stage can be performed using some basis pursuit algorithm (Rubinstein et al., 2008). Having learnt the dictionary $\boldsymbol{D}_x^c$ and sparse coding matrix $\boldsymbol{G}^c$, the corresponding spectral dictionary $\boldsymbol{D}_y^c$ is computed by solving a simple linear least-squares problem

$$\min_{\boldsymbol{D}_y^c} \left\|\boldsymbol{Y}^c - \boldsymbol{D}_y^c\boldsymbol{G}_c\right\|_F^2 \tag{5}$$

The closed-form solution of problem (5) is given by

$$\boldsymbol{D}_y^c = \boldsymbol{Y}^c\boldsymbol{G}_c^+ = \boldsymbol{Y}^c\boldsymbol{G}_c^T(\boldsymbol{G}_c\boldsymbol{G}_c^T)^{-1}. \tag{6}$$

Fig. 2a illustrates the training process.

2.2.2. Near-infrared image recovery

Considering a new testing RGB image, say $\boldsymbol{X}_{new}$, the following procedure is employed to estimate the Near-infrared image that corresponds to the same scene. In particular, the RGB image is divided into non-overlapping patches size $\sqrt{p}\times\sqrt{p}\times 3$ (held as a column vectors of length d). For each patch x_i encountered in the RGB image, its nearest cluster in the training set is found and the corresponding RGB dictionary is employed to represent it. Then, all patches in the examined image that belong to the cth cluster can be expressed as

$$\boldsymbol{X}_{new}^c = \boldsymbol{D}_x^c\boldsymbol{G}_c, \tag{7}$$

where $\boldsymbol{G}_c$ denote the sparse representation matrix for the testing RGB data in the cth cluster. Note that the sparse representation matrix can be calculated by using some sparse coding algorithm such as batch-OMP (Rubinstein et al., 2008). However, taking into account that

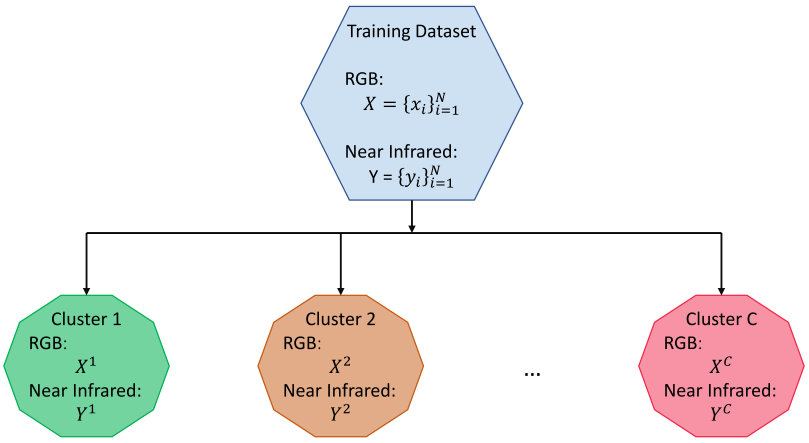
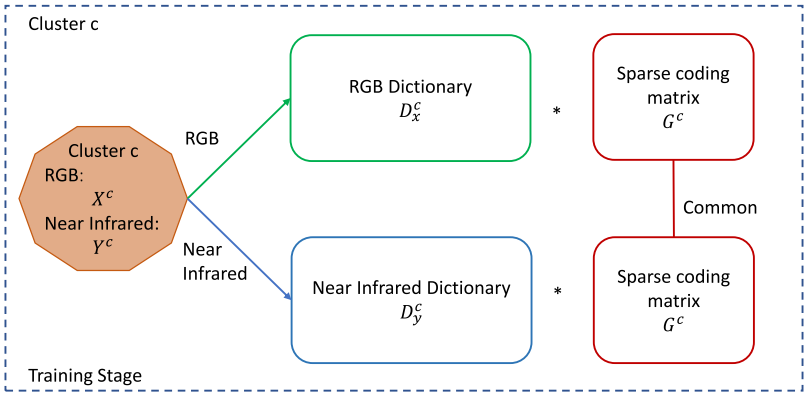
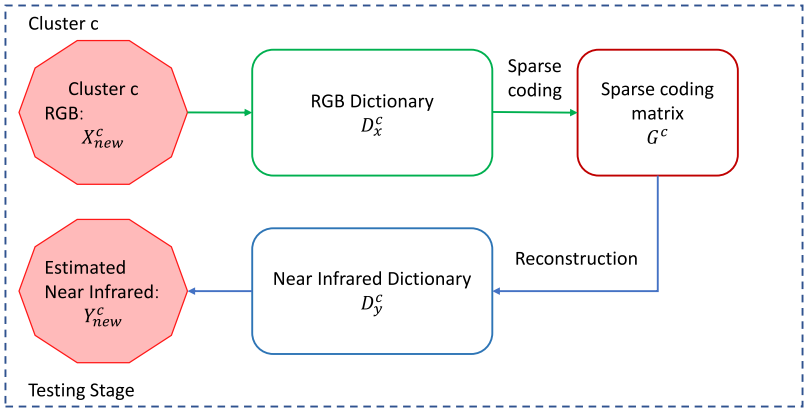


Fig. 1. Considering the challenging nature of the problem of reconstructing Near-Infrared Information from RGB signals, the first stage of the proposed methods is a cluster based procedure. In more detail, the training dataset of the raw RGB and Near-Infrared patches is grouped into clusters, where each cluster contains similar near-infrared signals and the corresponding RGB data. After that for each cluster we learn a different model independently based on the proposed CDLNR and CANR methods.



(a) Training stage for the c-th cluster.



(b) Testing stage for the c-th cluster.

Fig. 2. An illustration of the proposed CDLNR model for reconstructing the Near-Infrared data of the cth cluster. (a) Training: For each cluster. two compact coupled dictionaries are learnt to represent the RGB and Near-Infrared data of the respective cluster along with the common sparse coding matrix. (b) Testing: The above procedure is repeated for all clusters depicted in Fig. 1.

the method can be employed for processing large volumes of data, we employ a more efficient block-processing sparse coding algorithm presented in Gkillas et al. (2021), which reduces notably the required computational complexity, thus making our method ideal for online applications. In particular, since the patches that belong to the same cluster exhibit strong similarities properties, this structure can be utilized in order to determine the sparse representation of the similar patches in a block-processing manner. In the case of interest, where the RGB patches of c cluster X_{new}^c satisfy the self-similarity property, notably computational benefits could be derived by considering that the same support set S^c i.e., the same atoms from the respective dictionary $\boldsymbol{D}_x^c$ can be used to represent the signals in the cluster. Hence, instead of calculating the support set of each signal independently, we first compute the average/centroid signal of the cluster

$$x_c = \frac{1}{N_c}\sum_{i=1}^{N_c} x_i \qquad x_i \in \mathbb{R}^d, \tag{8}$$

and then employ some sparse coding algorithm (e.g., the OMP, or Lasso) to sparsely encode the vector x_c using the given dictionary. The required support S^c is derived as the set of atoms employed in the representation of the centroid signal x_c. Having estimated a proper support set, a simple optimization problem is formulated to compute the optimal weights for each data-vector in $\boldsymbol{X}^c_{new}$

$$\arg\min_{\boldsymbol{G}_{c,S^c}} \left\| \boldsymbol{X}^c_{new} - \boldsymbol{D}^c_{x,S^c} \boldsymbol{G}_{c,S^c} \right\|_F^2 \tag{9}$$

where $\boldsymbol{D}^c_{x,S^c}$ is the matrix that results from the dictionary $\boldsymbol{D}$ after keeping only the columns/atoms indexed by the set S^c, and $\boldsymbol{G}_S$ is the corresponding matrix of representation weights. It should be noted that matrix $\boldsymbol{G}_{c,S^c}$ is not sparse, since it contains only the nonzero elements of the sparse coding matrix $\boldsymbol{G}_c$. The above procedure is repeated for all the clusters.

Given the computed sparse representation matrix $\boldsymbol{G}_c$, the corresponding near-infrared data of the cth cluster can be estimated by

$$\boldsymbol{Y}^c_{new} = \boldsymbol{D}^c_y \boldsymbol{G}_c. \tag{10}$$

Thus, the overall near-infrared image can be recovered by gathering the reconstructed data from all the $c^C_{i=1}$ clusters. Fig. 2b illustrates the testing process.

2.3. Coupled autoencoders — preliminaries

Autoencoders is an efficient unsupervised feature learning approach, which aims to model the hidden structure of the examined data, i.e., the intrinsic hidden representations of the input data (Ca et al., 2010). To capture such hidden representations, the autoencoder aim to copy its input to its output. On the other hand, coupled Autoencoders are a new promising concept, similar to the coupled dictionary learning problem. Formally, the coupled autoencoder-based model can be defined as learning a pair of autoencoders in order to reveal the intrinsic representations of the so-called source and target spaces along with a coupled mapping function which can model the linear/non-linear relationships between the intrinsic representations of the signals that belong to the source space and the signals of the target space.

In this study, we demonstrate the effectiveness of the coupled autoencoders to reconstruct near-infrared data (target signals) from their corresponding RGB measurements (source signals). Considering that the autoencoders stand out in preserving the local consistency of the data (Wang et al., 2014), we propose a different framework compared to the existing literature. Particularly, following a similar procedure as in Section 2.2.1, instead of learning two global coupled autoencoders and one mapping function for a large dataset, we cluster the training dataset and learn a bank of coupled autoencoders and their respective mapping functions for each cluster separately. Thus, the proposed coupled autoencoder-based model (i.e., CANR) is more compact and effective in representing the spectral and RGB signals.

2.3.1. The CANR method

Given the RGB training set (source) $\boldsymbol{X}$ and the corresponding Near-Infrared training set (target) $\boldsymbol{Y}$, we cluster again the training data into C clusters via the Kmeans++ algorithm so that each cluster again contains similar pairs of Near-infrared and RGB patches. Thus, having grouped the data into clusters, the coupled autoencoder-based model is employed to each cluster independently.

In more detail, let $\boldsymbol{X}^c \in \mathbb{R}^{d \times N_c}$ and $\boldsymbol{Y}^c \in \mathbb{R}^{l \times N_c}$ denote the RGB and the Near-Infrared training data in the cth cluster, where N_c is the number of members in this cluster. The RGB (source) and the spectral (target) autoencoders that learn the hidden intrinsic representations of the RGB and Near-Infrared patches, respectively, can be derived by minimizing the reconstruction errors

$$\left\| \boldsymbol{X}^c - \hat{\boldsymbol{X}}^c \right\|_F^2 = \left\| \boldsymbol{X}^c - D^c_s(\mathcal{E}^c_s(\boldsymbol{X}^c)) \right\|_F^2$$

$$\left\| \boldsymbol{Y}^c - \hat{\boldsymbol{Y}}^c \right\|_F^2 = \left\| \boldsymbol{Y}^c - D^c_t(\mathcal{E}^c_t(\boldsymbol{Y}^c)) \right\|_F^2, \tag{11}$$

where $\hat{\boldsymbol{X}}^c$ and $\hat{\boldsymbol{Y}}^c$ denote the calculated through autoencoders RGB and near-infrared patches respectively, $\mathcal{E}^c_s(\cdot)$, $D^c_s(\cdot)$ stand for the encoder and decoder of the RGB domain and $\mathcal{E}^c_t(\cdot)$, $D^c_t(\cdot)$ are the encoder and decoder of spectral domain. The intrinsic representations of the RGB data are derived via the encoder process

$$\boldsymbol{S}^c = \mathcal{E}^c_s(\boldsymbol{X}^c)$$

or equivalently,

$$\begin{aligned} \boldsymbol{S}^c_1 &= \varphi(\boldsymbol{W}^c_{s,1} \boldsymbol{X}^c + b^c_{s,1}) \\ \boldsymbol{S}^c_2 &= \varphi(\boldsymbol{W}^c_{s,2} \boldsymbol{S}^c_1 + b^c_{s,2}) \\ &\cdots \\ \boldsymbol{S}^c &= \varphi(\boldsymbol{W}^c_{s,L} \boldsymbol{S}^c_{L-1} + b^c_{s,L}), \end{aligned} \tag{12}$$

where $\boldsymbol{W}^c_{s,i}$, $b^c_{s,i}$ $(i = 1, \ldots, L)$ denote the weight matrices and the bias terms for the encoding layers of the RGB (source) autoencoder, $\varphi(.)$ is the activation function, L denotes the number of hidden layers, $\boldsymbol{S}^c \in \mathbb{R}^{k \times N_c}$ is the output of the source encoder $\mathcal{E}^c_s(.)$ and $k \ll d$.

Accordingly, the intrinsic representations of the Near-Infrared data are given by the following encoding procedure $\mathcal{E}^c_t(\cdot)$

$$\boldsymbol{T}^c = \mathcal{E}^c_t(\boldsymbol{Y}^c)$$

or equivalently,

$$\begin{aligned} \boldsymbol{T}^c_1 &= \varphi(\boldsymbol{W}^c_{t,1} \boldsymbol{Y}^c + b^c_{t,1}) \\ \boldsymbol{T}^c_2 &= \varphi(\boldsymbol{W}^c_{t,2} \boldsymbol{T}^c_1 + b^c_{t,2}) \\ &\cdots \\ \boldsymbol{T}^c &= \varphi(\boldsymbol{W}^c_{t,L} \boldsymbol{T}^c_{L-1} + b^c_{t,L}), \end{aligned} \tag{13}$$

where $\boldsymbol{W}^c_{t,i}$, $b^c_{t,i}$ $(i = 1, \ldots, L)$ denote the weight matrices and the bias terms for the encoding layers of the RGB (source) autoencoder, $\varphi(.)$ is the activation function, L denotes the number of hidden layers, $\boldsymbol{T}^c \in \mathbb{R}^{k \times N_c}$ is the output of the source encoder $\mathcal{E}^c_t(.)$ and $k \ll p$.

Similarly, the corresponding reconstructed RGB and Near-Infrared signals can be computed as follows

$$\hat{\boldsymbol{X}}^c = D^c_s(\boldsymbol{S}^c) \tag{14}$$

$$\hat{\boldsymbol{Y}}^c = D^c_t(\boldsymbol{T}^c), \tag{15}$$

where $D^c_s(.)$ denotes the decoding process of the RGB autoencoder and $D^c_t(.)$ is the decoding process of the spectral autoencoder. Note that both decoders consist also of L fully-connected layers.

Having estimated the intrinsic representation of the RGB and spectral data for the cth cluster according to Eqs. (12), (13) a neural network is employed to learn the mapping function between the intrinsic representations $\boldsymbol{S}^c$ and $\boldsymbol{T}^c$, hence transferring beneficial information from the RGB to spectral domain. Mathematically, this can be expressed as follows:

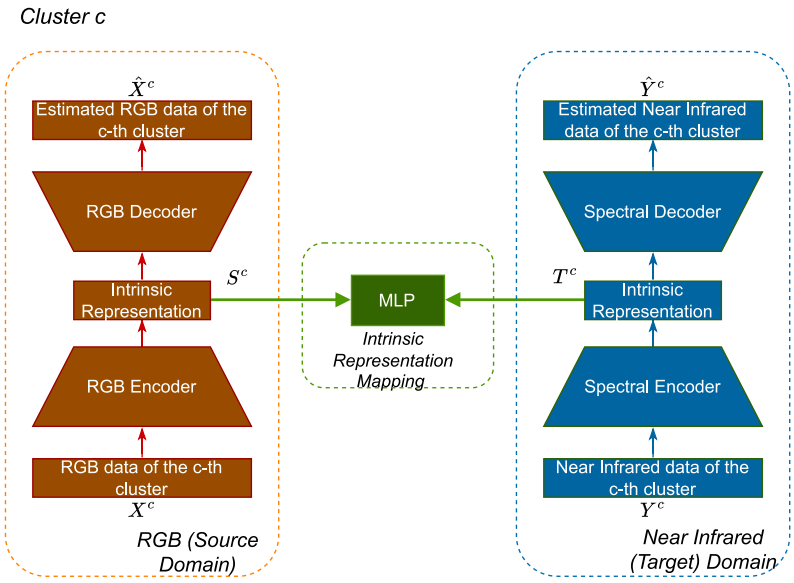
$$\boldsymbol{T}^c = \mathcal{F}(\boldsymbol{W}^c_m \boldsymbol{S}^c + b^c_m) \tag{16}$$

where $\boldsymbol{W}^c_m$ is the weight matrix and b^c_m denotes the bias term. Thus, the network can be optimized solving the following loss function

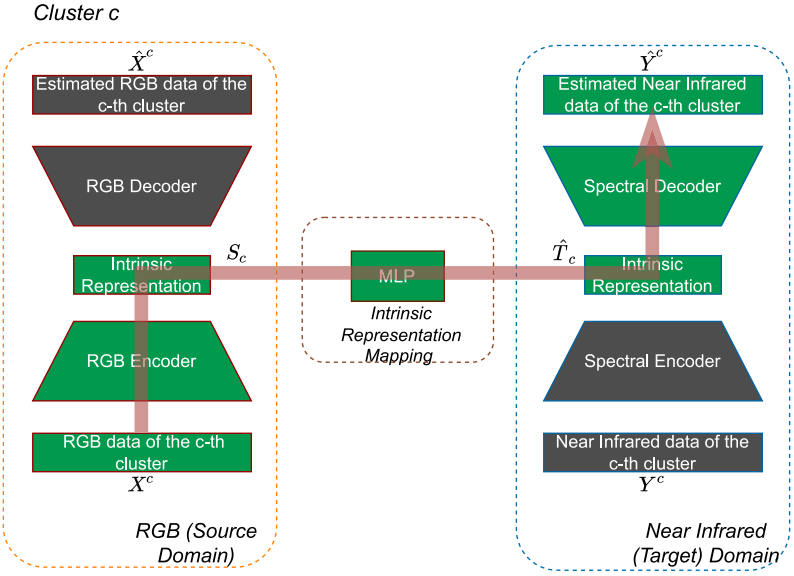
$$\mathcal{L}_{mlp} = \left\| \boldsymbol{T}^c - \hat{\boldsymbol{T}}^c \right\|_F^2 = \left\| \boldsymbol{T}^c - \mathcal{F}(\boldsymbol{W}^c_m \boldsymbol{S}^c + b^c_m) \right\|_F^2. \tag{17}$$

2.3.2. Coupled learning: End-to-end training

The most crucial part of the proposed method is to optimize and couple the autoencoders along with the mapping function for each training cluster. However, by training the autoencoders first and then the mapping function (based on the estimated intrinsic representations) could result in sub-optimal results, due to the fact that the autoencoders and the mapping network are trained independently. Thus, there is no transfer or coupled learning between the source space (i.e., the



(a) Initialization stage for the c-th cluster: First, the autoencoders are trained independently and then the MLP network is trained to learn a mapping between the intrinsic representations.



(b) Coupled Learning stage for the c-th cluster.

Fig. 3. An illustration of our proposed CANR model for reconstructing the Near-Infrared data of the cth cluster. (a) Initialization: First the autoencoders are trained independently to learn the intrinsic representations of the RGB and Near-Infrared data, respectively, of the cth cluster (stage 1) and then a mapping function (MLP) is learnt between these representations (stage 2). (b) Coupled Learning: since the autoencoders were trained independently and there was no transfer learning across domains, a coupled autoencoder is employed in order to jointly optimize all the active parts of the autoencoders i.e., the RGB encoder, the MLP network and the Spectral decoder) involved in the reconstruction of Near-infrared data from the corresponding RGB data of the cth cluster (stage 3). The above procedure is repeated for all the clusters depicted in Fig. 1.

RGB patches) and the target space (i.e., the near-infrared patches). Nonetheless, this procedure can be employed as initialization process of the model.

To proceed further, consider the two autoencoders and the neural network (mapping function) for the cth cluster. Taking into consideration that the primary aim of the CANR method is to accurately recover the near-infrared data $\boldsymbol{Y}^c$, the objective function for optimizing jointly the two autoencoders (RGB and spectral) and the mapping network is given by:

$$\left\|\boldsymbol{Y}^c-\hat{\boldsymbol{Y}}^c\right\|_F^2 \overset{(15)}{\Longrightarrow} \left\|\boldsymbol{Y}^c-D_t^c(\hat{\boldsymbol{T}}^c)\right\|_F^2 \overset{(16)}{\Longrightarrow}$$
$$\left\|\boldsymbol{Y}^c-D_t^c(\mathcal{F}(\boldsymbol{S}^c)\right\|_F^2 \overset{(12)}{\Longrightarrow} \left\|\boldsymbol{Y}^c-D_t^c(\mathcal{F}(\mathcal{E}_s^c(\boldsymbol{X}^c)\right\|_F^2. \quad (18)$$

Note that in relation (18) the RGB encoder $\mathcal{E}_s^c(.)$, the neural network $\mathcal{F}(.)$ and the spectral decoder $D_t^c(.)$ all explicitly participate in the estimation of the corresponding spectral data $\hat{\boldsymbol{Y}}^c$. *Thus, the coupling between the two autoencoders and the mapping function $\mathcal{F}$ is derived by designing a stacked architecture where its first module is the RGB encoder $\mathcal{E}_s^c(\cdot)$, the second module is the mapping function and the final module is the spectral decoder $D_t^c(\cdot)$.* Fig. 3 demonstrates the proposed coupled architecture. Then, an end-to-end training procedure is employed in order to minimize the objective function in (18).

2.3.3. Efficient near-infrared recovery

Given a new testing RGB image, say $\boldsymbol{I}_{new}$, the following methodology is used to recover the corresponding near-infrared image of

the scene. For each RGB patch x_i encountered in the RGB image, its nearest cluster is found. Then, for all patches in the examined image that belong to the cth cluster, say $\boldsymbol{X}^c_{new}$, their intrinsic representations $\boldsymbol{S}^c_{new}$ are estimated through the RGB encoder based on Eq. (12). Then the corresponding near-infrared intrinsic representations $\boldsymbol{T}^c_{new}$ can be estimated via the mapping neural net via Eq. (16), and finally the near-infrared patches $\boldsymbol{Y}^c_{new}$ of the cth cluster are reconstructed via the spectral decoder (15). Thus, the overall near-infrared image can be recovered by gathering the reconstructed data from all the C clusters.

3. Results and discussion

To validate the efficacy and applicability of the two proposed methods[1] (i.e., CDLNR and CANR) extensive experiments were conducted in the context of the precision agriculture. In particular, our goals are to highlight:

- that the proposed methods are able to accurately reconstruct near-infrared images using only RGB signals;
- that the reconstructed spectral bands can be effectively employed to estimate the NDVI index without the need of a multispectral camera.
- that the proposed shallow learning methods exhibit competitive or better performance against the deep learning approaches presented in Sharma and Hefeeda (2020), Abady et al. (2020) and Zhang et al. (2022), offering at the same time much lower computational cost.
- that there is low dependency on data availability compared to competing deep learning methods.

3.1. Experimental setup

Dataset: To validate the efficacy of the proposed architectures, we employ a custom dataset that pertains a variety of tomato plants in a greenhouse, captured under real field conditions. The considered dataset consists of 100 multispectral images and their corresponding RGB images with spatial resolution 1770×2368. In particular, we employed the MUSES9-MS- PL multispectral camera (Spectricon, Greece), which features 4–6 Megapixels C-MOS @ 25 f/s in the 370–1100 nm spectral range. The camera contains two sensors i.e., a conventional RGB sensor and a multispectral sensor. In total the camera provides 8 channels. The first 3 channels are acquired with the conventional RGB sensor while the rest are captured via the multi-spectral sensor. The multispectral sensor provides 3 channels inside the visible spectrum, namely bands at 460 nm (near blue), 540 nm (near green), and 630 nm (near red), and other two spectral bands in the near-infrared spectrum at 850 nm and 950 nm. The width of each spectral band of the multispectral sensor of the considered camera is given in Table 1.

Using the visible bands derived from the multispectral sensor of the camera, the goal is to reconstruct the near-infrared band at 850 nm (i.e., the spectral dimension is equal to $B = 1$). Fig. 4 illustrates a small sample of RGB images derived from the considered dataset. More details regarding the dataset along with the setting of the greenhouse can be found in study (Georgantopoulos et al., 2023).

Parameter Settings: To determine the appropriate parameters and assess the performance of the proposed architectures, a 5-fold cross validation process was used, thus splitting the dataset into training, validation and test set. Concerning the **CDLNR** the dictionaries contained $K = 68$ atoms, while the sparsity level was set to 9. Furthermore, the RGB and the near-infrared images (840 nm) were processed into non-overlapping patches with size $6 \times 6 \times 3$ (held as a column vector with length $d = 108$) and $6 \times 6 \times 1$ (held as a column vector with length $l = 36$) respectively. The number of clusters was set to $C = 40$ clusters.

[1] Source code of the proposed methods: https://github.com/alexandrosgk/coupled-learning-methods-for-Recovering-Near-Infrared-Images.

Table 1
Wavelength regions supported by the MUSES9-MS-PL camera and those used in current study. (FWHM: Full Width at Half Maximum.)

Multispectral sensor bands	Lower limit (nm)	Upper limit (nm)	Current study (nm)	FWHM (nm)
Infrared	800	1000	980,850	50
Red	600	700	630	40
Green	500	600	540	30
Blue	400	500	460	30

For **CANR** a patch with size of $10 \times 10 \times 3$ and 10×10 was employed to split the RGB and near-infrared images into non-overlapping patches accordingly. The coupled autoencoders consisted of one layer with latent dimension equal to 30 and 10 for the RGB and spectral autoencoder respectively. Moreover, the size of the one-layer neural network was set to 30. The number of clusters was set to $C = 25$. During the initialization training stage the learning rate was set to $1e-03$ and the number of epochs was 300. For the coupled learning stage, the number of epochs was 350 and the learning rate was set to $1e-04$. Finally, the Adam optimizer was used for the training of the second method and the ReLU was employed as activation function.

Regarding the clustering procedure, we explored various clustering algorithms without observing any significant impact on the performance of the pro- posed methods. To this end, we used a well-established algorithm i.e., the Kmeans++ for its performance and simplicity. In particular, the kmeans++ algorithm was employed on the raw RGB and near infrared data using as distance metric the cosine distance, which provided the best results among other distance functions.

Evaluation metrics: To quantify the performance of our models, we compared the reconstructed near-infrared images with the corresponding ground truth images in terms of the Peak Signal to Noise Ratio (PSNR), the Structural Similarity Index (SSIM) and the Relative Root Mean Square Error (RMSE), normalized by the ground truth luminance. Note that higher values of the PSNR and SSIM metrics indicate better reconstruction performance.

Compared methods: To highlight the benefits of the shallow learning proposed models (CDLNR and CANR), we compare them with the very deep learning approach presented in Sharma and Hefeeda (2020) and two GAN-based approaches (Abady et al., 2020; Zhang et al., 2022) . Regarding the approach in Sharma and Hefeeda (2020), we employ the official implementation[2] provided by the authors. For the method in Abady et al. (2020), we use the official implementation[3] of the Nice GAN based methodology. Regarding the study in Zhang et al. (2022), we should clarify that since the original implementation (source code) is not publicly available, we perform the comparison based on a similar GAN based architecture, called cycle GAN (Zhu et al., 2017).[4] The optimal parameters for the above methods were optimized based on the original papers.

3.2. Performance evaluation

Table 2 summarizes the quantitative performance of the proposed shallow learning architectures (i.e., CDLNR and CANR) against the deep learning methodologies proposed in Sharma and Hefeeda (2020), Abady et al. (2020) and Zhu et al. (2017), which is similar to method (Zhang et al., 2022). It is evident that CANR outperforms the other methods, maintaining high PSNR and SSIM values, and low error values in terms of nRMSE metric. The above results are consistent with the visual analysis provided in Figs. 5 and 6. In greater detail, in Fig. 5 the reconstructed near-infrared images from the considered approaches

[2] https://github.com/nehasharma512/vein-visualization.
[3] https://github.com/alpc91/NICE-GAN-pytorch.
[4] https://github.com/junyanz/pytorch-CycleGAN-and-pix2pix.

Fig. 4. A small sample of the RGB images of the considered dataset.

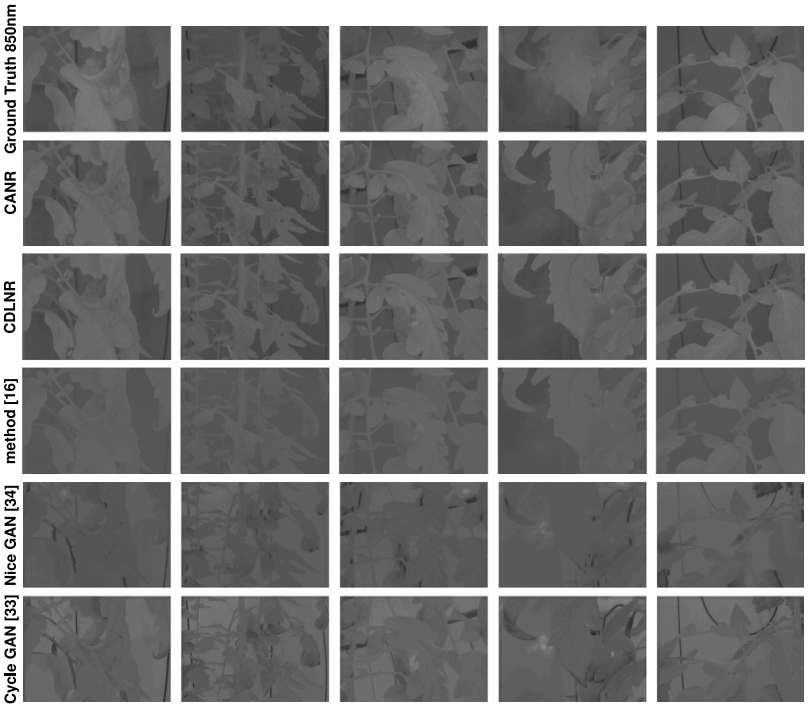


Fig. 5. The ground-truth and reconstructed near-infrared images (850 nm). In first row we present the ground truth image, in second row the reconstructed near-infrared image employing the CANR method, in third three we depict the reconstructed image using the CDL method, in fourth row the reconstructed image using method (Sharma and Hefeeda, 2020), in fifth row the reconstructed image using Nice GAN method in Abady et al. (2020) and in sixth row reconstructed image using Cycle GAN method in Zhu et al. (2017).

are depicted, whereas in Fig. 6 the corresponding averaged absolute difference value between the ground truth and the reconstructed spectral images are demonstrated. From these figures, we can deduce the high quality of the results derived from the CANR method and the competitive performance of the coupled dictionary learning approach (CDLNR).

Comparison against the deep learning methods (Sharma and Hefeeda, 2020; Abady et al., 2020; Zhu et al., 2017)**:** Interestingly, although the deep learning approach developed in Sharma and Hefeeda (2020) exhibits good performance, the proposed coupled autoencoder based method (CANR) provides better results, while the coupled dictionary learning approach constitutes a competitive alternative. Additionally, the proposed models notably outperform the two GAN-based approaches in Abady et al. (2020) and Zhu et al. (2017). This remark can be attributed to the fact that the core of the proposed methods is the clustering procedure. In particular, by grouping the data into clusters and learning models for each cluster separately, this process enables the involved models to be more compact and concise, since each model (either dictionary or autoencoder) is focused on capturing detailed features of the data of each cluster only. Thus, although the proposed methods consist of shallow learning models, they are able to exhibit competitive performance against the very deep learning approaches (Sharma and Hefeeda, 2020; Abady et al., 2020; Zhu et al., 2017), which require a huge amount of data in order to capture the relationships between the signals.

Apart from the quantitative analysis, our models provide some essential practical advantages for real-world precision agriculture applications. Considering that in real-field scenarios and settings, the available data and the specialized equipment are limited, the proposed shallow learning approaches outperform the deep learning approaches, since they require significantly less data for the training without the need of advanced hardware equipment for the training and inference procedures. In particular, although the existing GAN-based methodologies (Abady et al., 2020; Zhu et al., 2017) can be characterized by high representation capacity, they do not provide competitive results due to the fact that they contain a vast amount of learnable parameters, thus

Table 2
Quantitative performance of the proposed shallow learning architectures (i.e., CDLNR and CANR) and comparison with the deep learning methodology proposed in Sharma and Hefeeda (2020) and two GAN based methodologies in Abady et al. (2020) and Zhu et al. (2017).

Metrics	CDLNR Section 2.2.1	CANR Section 2.3.1	method (Sharma and Hefeeda, 2020)	Nice GAN (Abady et al., 2020)	Cycle GAN (Zhu et al., 2017)
PSNR	32.01	32.46	32.09	30.05	29.45
SSIM (0-1)	0.9603	0.9619	0.9607	0.9475	0.9325
nRMSE (0-1)	0.0242	0.0250	0.0252	0.0304	0.0472

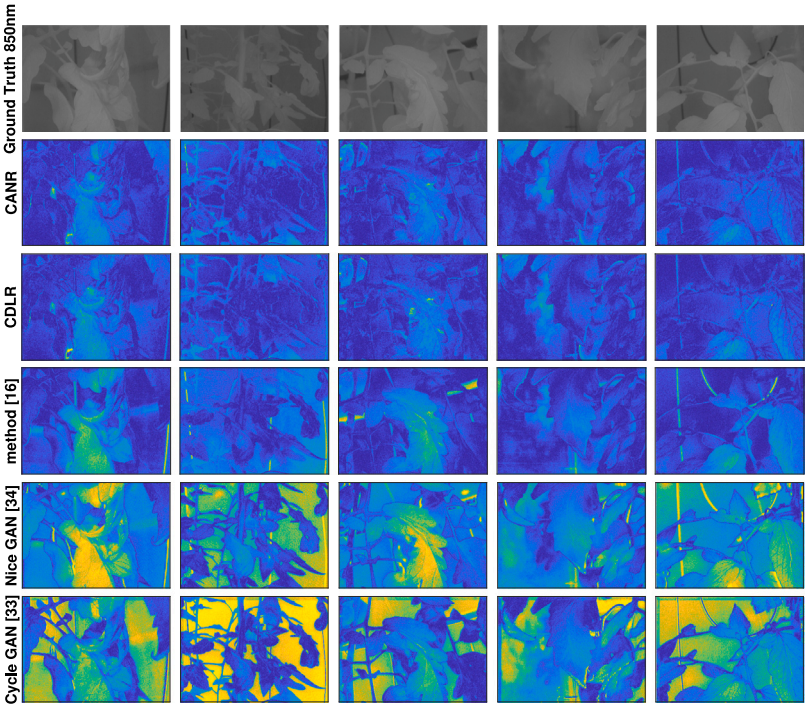


Fig. 6. The ground-truth images and the corresponding averaged absolute difference value between the ground truth and the reconstructed images (error maps). In first row we illustrate the ground-truth images, in second row the error maps of the CANR method, in three row the error maps of the CDLNR method, in fourth row the error maps of method in Sharma and Hefeeda (2020), in fifth row the error maps of Nice GAN method in Abady et al. (2020) and in sixth row the error maps of Cycle GAN method in Zhu et al. (2017). It should be noted that the error maps are depicted on a scale of [0–0.1] instead of the standard scale [0–1], since the errors are very small for the proposed models.

requiring more training data in order to exhibit better performance. Table 3 validates the above remark. In particular, from this table, we can deduce that the proposed approaches require significantly fewer parameters to achieve competitive performance against the other deep learning approaches. Thus, the proposed models can be efficiently deployed to edge devices, e.g., smart phones or sensors with limited computational resources installed inside greenhouses.

Comparison between the proposed methods: Focusing on the proposed methods, the CANR exhibits better performance compared to the CDLNR approach. This remark can be justified by the fact that the coupled dictionary learning method is limited to capture only the linear relationships existing among the RGB and the near-infrared signals. However, the computation of near-infrared bands using only RGB signals is in fact a really challenging ill-posed problem, since the information captured by RGB cameras does not practically extend to the IR range. Taking into account this, the proposed coupled autoencoder based methodology (CANR) is able to provide more accurate results, as it can reveal and capture more complex and non-linear relationships that describe the RGB and near-infrared domain. Nonetheless, since the CDLNR methodology is a linear approach, it requires less computational resources during the training procedure over the CANR method.

3.3. Ablation study

In this section, an ablation study was conducted in order to investigate the sensitivity of the proposed methods to the selection of the optimal parameters.

3.3.1. Impact of the parameters on the performance of the CDLNR method

Fig. 7 demonstrates the impact of the parameters, i.e., the number of clusters (C), the dictionary size (number of atoms) and the sparsity level on the reconstruction performance of the coupled dictionary learning method. In particular, the number of clusters influence significantly the performance of the method, providing a huge dimensionality reduction on the size of the coupled dictionaries and thus enabling the learning of multiple compact dictionaries that describe more accurate the data of each cluster. Thus, the best results occurred when the number of clusters was set to $C = 40$ and the number of atoms of the coupled dictionaries was set to $K = 68$. Regarding the sparsity level, it is evident that it slightly impacts the performance of the method. Considering that high values of the sparsity level would affect only the computational complexity of the method, smaller sparsity levels were preferred (e.g., sparsity level equal to 9).

3.3.2. Impact of the parameters on the performance of the CANR method

Similar to the previous method, the number of clusters constitutes a vital factor that impacts the performance of the coupled autoencoder based architecture. Thus, in this experiment the influence of the number of clusters along with the dimension of the latent representations of the RGB and near-infrared signals is examined. Fig. 8 illustrates the results. Again, the best results occurred when the number of clusters was high, i.e., $C = 25$. This remark can be attributed to the fact that the grouping of the dataset into many clusters $C > 1$ enables the learning of multiple coupled autoencoders, which are able to capture

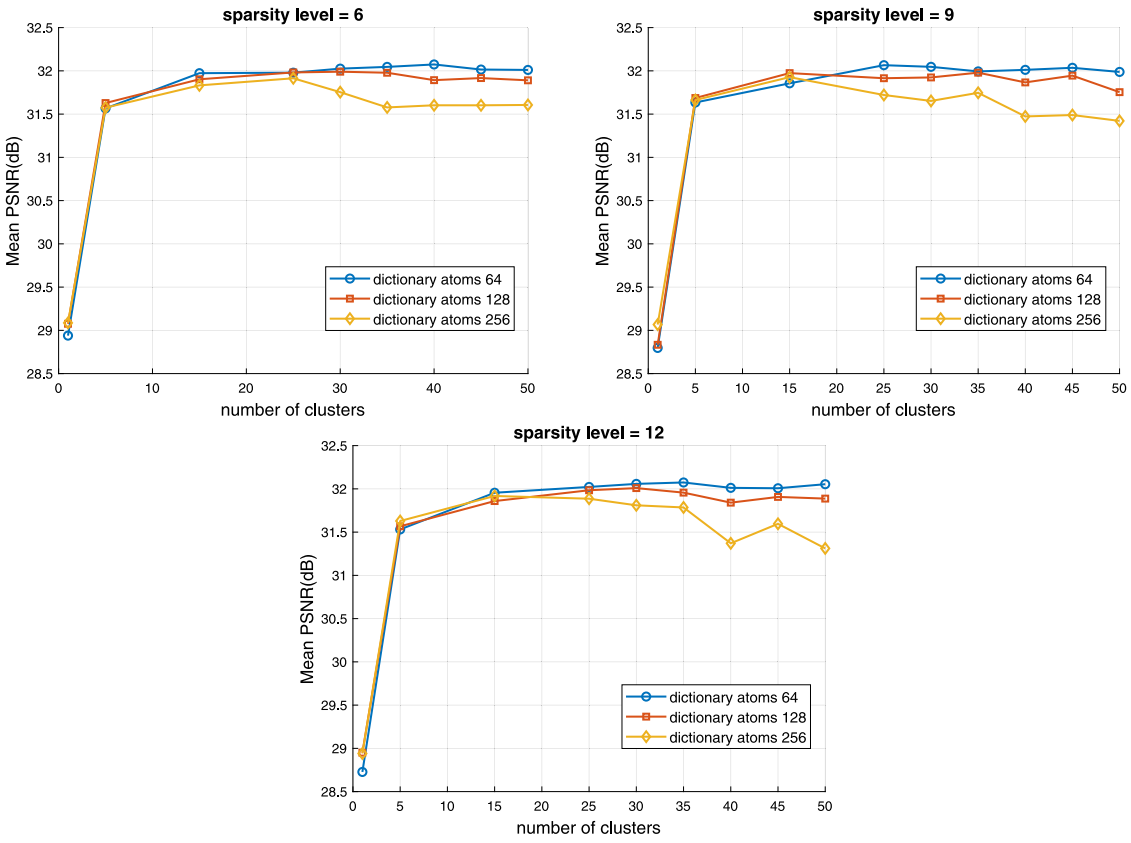


Fig. 7. Impact of the number of clusters and the number of atoms of the dictionary on the performance of the coupled dictionary learning method (CDLNR) employing sparsity level equal to 6, 9 and 12.

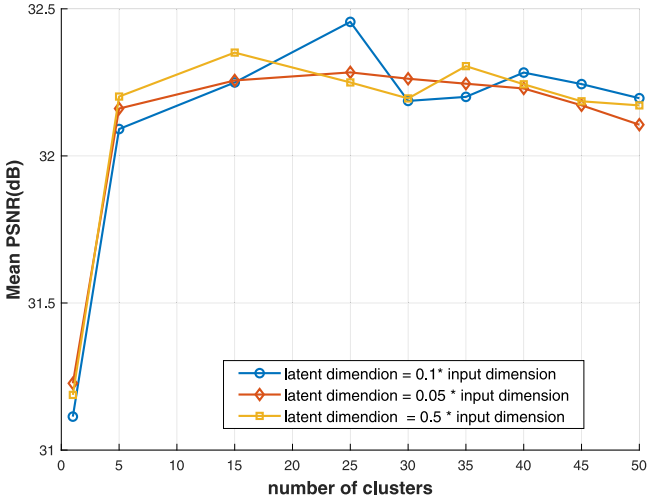


Fig. 8. Impact of the number of clusters and the dimension of the latent representations of the RGB and NIR signals on the performance of the coupled autoencoder based method (CANR). Since the optimal patch size was set to 10×10, the input dimension of the RGB signals is $d = 300(10 \times 10 \times 3)$ and the input dimension of NIR signals is $p = 100(10 \times 10)$. Thus, the examined latent dimensions of the RGB autoencoder was $\{15, 30, 150\}$, whereas the corresponding latent dimension of the spectral autoencoder was $\{5, 10, 50\}$.

more concrete low-level features of the data of each cluster; hence they empower the method to model efficiently the complex relationships existing among the RGB and Near-Infrared signals. On the other hand, when there is a single cluster ($C = 1$), the performance of the method drops significantly, indicating that the learning of one pair of coupled autoencoders is limited to model only high-level features of the whole dataset, without the required detail for such a challenging problem.

3.3.3. Impact of the coupled learning on the performance of the CANR method

One of the most critical aspects of the CANR is the coupled learning stage. To validate the claims of Section 2.3.2 and highlight the impact of the coupled learning on the reconstruction performance of the method, experiments were performed with and without the coupled learning during the training of the models. Table 4 summarizes the results. It is clear that the coupled learning stage significantly enhances the performance of the CANR method. The above remarks can be justified by the fact that during the coupled learning stage the two autoencoders along with the mapping function are optimized jointly, thus enabling the models to capture more effectively the complex relationships between the RGB and near-infrared signals.

3.3.4. Impact of the number of training data on the performance of the compared methods

In this section, we investigate how the number of the available training data affects the performance of the proposed models and the very deep learning approaches. In more detail, we used the 100%, the 50% and the 10% of the training data to optimize the proposed models and the deep learning models in Sharma and Hefeeda (2020), Abady et al. (2020) and Zhu et al. (2017), whereas the testing set remained the same in all cases. Fig. 9 summarizes the results. It is evident that the performance of the proposed methods, is only slightly affected by the limited training data compared to the method in Sharma and Hefeeda (2020) and the two GAN-based methods in Abady et al. (2020) and Zhu et al. (2017). where their performance heavily degraded due to the

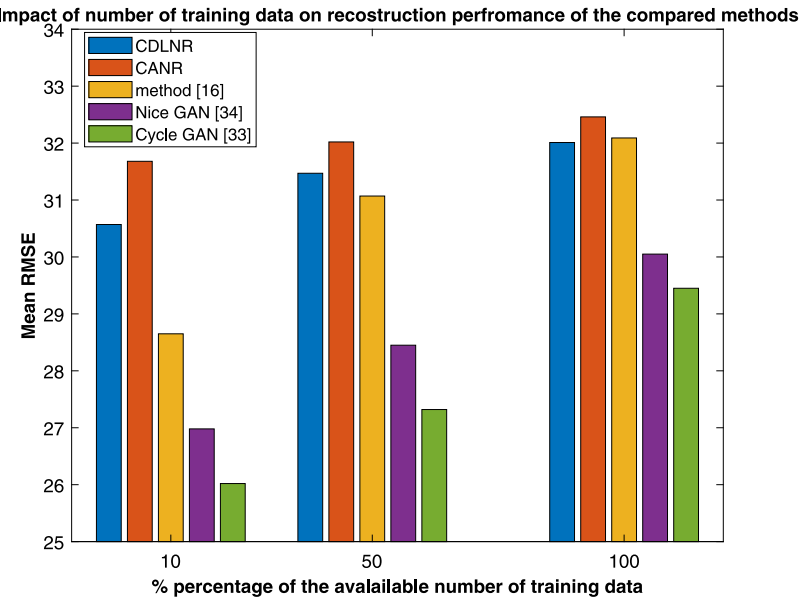


Fig. 9. Impact of the number of training data on the reconstruction performance of the compared models.

Table 3
Model complexities comparison of our proposed method, the deep learning method in Sharma and Hefeeda (2020) and the two GAN based methodologies in Abady et al. (2020) and Zhu et al. (2017).

Metrics	CDLNR	CANR	Method of (Sharma and Hefeeda, 2020)	Nice GAN (Abady et al., 2020)	Cycle GAN (Zhu et al., 2017)
PSNR	32.01	32.46	32.09	30.05	29.45
# of Param($\times 10^6$)	0.025	0.01	0.8	9.5	21.17

Table 4
The performance of the CANR method with and without the Coupled learning stage.

Metrics	CANR without Coupled learning	CANR with Coupled learning
PSNR	30.26	32.46
SSIM	0.9584	0.9619
nRMSE(0-1)	0.0307	0.0250

Table 5
The impact of different sensor on the performance of the proposed methods.

Sensor	Metrics	CDLNR	CANR
Multispectral-based RGB	PSNR	32.01	32.46
RGB	PSNR	31.97	32.32

small number of available data. This pivotal finding is mostly attributed to the fact that the proposed shallow learning methods contain significantly less parameters compared to the other deep learning approaches, as we can see in Table 3. Since, the two GAN-based methods (Abady et al., 2020; Zhu et al., 2017) contain most of the parameters, they are more affected by the scarcity of the training data. Considering that in many real-world precision agriculture applications the available data can be limited, or expensive to acquire, the proposed models are able to effectively surmount this challenging problem providing accurate results due to their lightweight architecture.

3.3.5. *Impact of different sensors on the performance of the proposed methods*

As we mentioned in Section 3.1, the considered multispectral camera contains two sensors or modes. In more detail, it contains a conventional RGB sensor and a multispectral sensor which provides 5 spectral bands i.e., three spectral bands inside the visible spectrum, namely, the 460 nm, the 540 nm and 630 nm, along with a spectral band in the near-infrared spectrum, i.e., the 840 nm. The wavelengths of the visible bands of the multispectral sensor are typically close to the ones used in the RGB camera, however the response of the multi-spectral sensor is denser around these frequencies than the response of the RGB camera. During the experiments, we used the visible data from the multispectral sensor to train the proposed models and reconstruct the near-infrared band at 850 nm.

To explore the impact of the sensitivity function of different sensors, in this experiment we used the data derived from the conventional RGB sensor of the camera. As we can see from Table 5, the performance of the proposed methods is not affected by altering the sensitivity function of the sensor, thus providing competitive results utilizing a conventional RGB sensor.

Based on the above results, we could argue that the proposed models are able to provide accurate results, even if we use as RGB images derived from a low cost conventional sensor e.g., smartphone. Thus, during the training we need to gather pairs of the RGB images from the smartphone and the corresponding near-infrared image from the multispectral camera. In that case, a pre-processing step is required in order to register the paired images. Finally, in the case where the models have been trained using the multispectral-based RGB sensor and during the inference a different sensor is utilized, a fine-tuning processing stage is required in order to train again the proposed models based on the new data.

3.4. *Reconstruction of the NDVI index*

To further illustrate the effectiveness of the proposed methodologies and highlight their applicability on real-world settings, extensive experiments were performed in order to verify that the computation of

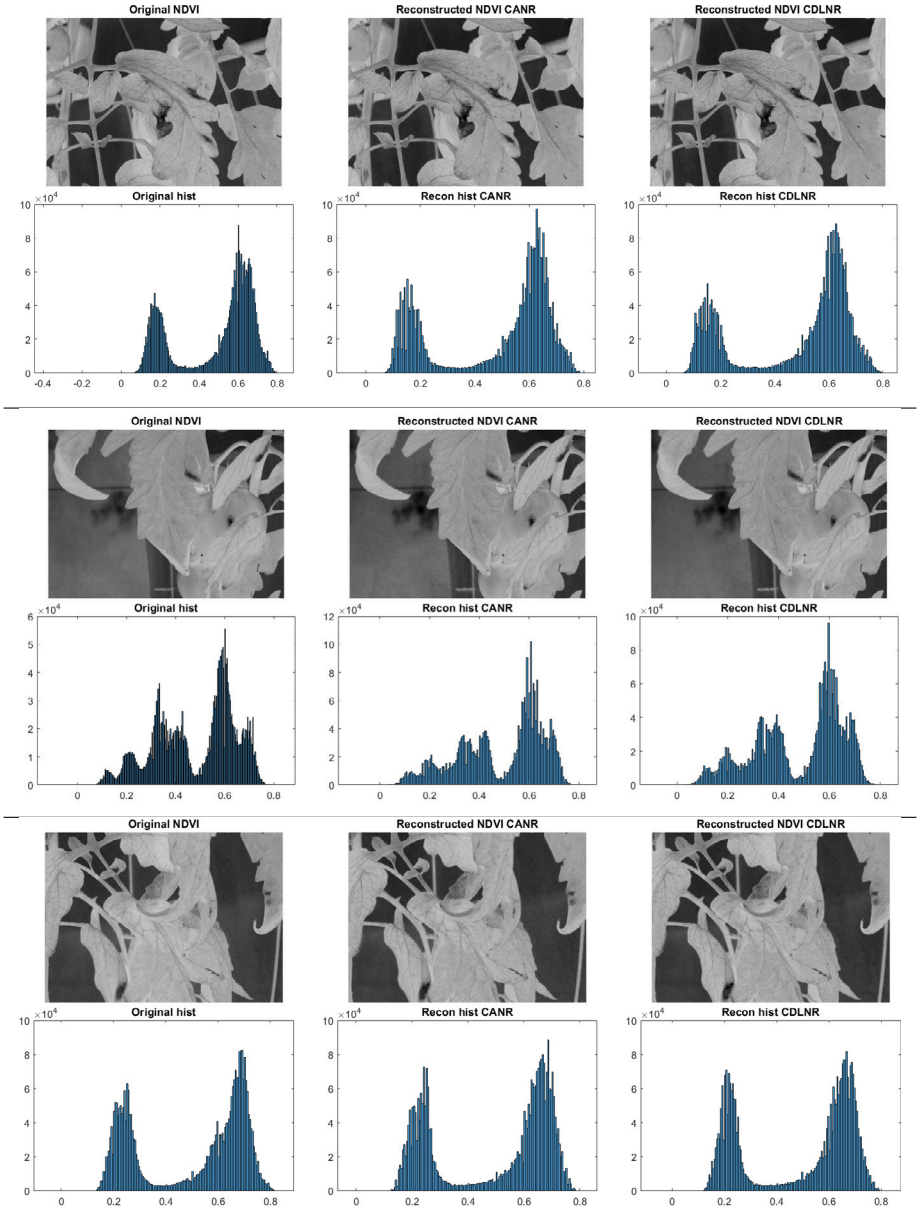


Fig. 10. The ground-truth NDVI index using the spectral bands in the 650 nm and 850 nm captured via the multispectral camera vs the estimated NDVI index given by the corresponding reconstructed spectral bands using the CANR and CDLNR methods. The corresponding histograms of the NDVI indices are shown, along with the NDVI images, to further clarify the quality of the reconstructed NDVI index.

the normalized difference vegetation index (NDVI) is feasible without the need of a multispectral or hyperspectral camera. To compute the NDVI index, we employed the red band (650 nm) of the RGB camera along with the reconstructed near-infrared spectral band at 850 nm.

Fig. 10 exemplifies the original NDVI indices employing the near-infrared band captured via the multispectral camera and the corresponding reconstructed NDVI indices computed through the estimated near-infrared bands from the proposed models. To better assess the quality of the reconstructed NVDI indices, their histograms as well as the histograms of the original NDVI indices are also depicted. According to Fig. 10 the proposed models are able to accurately compute the NDVI index. It is noteworthy that the derived results are comparable with the results provided from the multispectral camera. This remark validates our claims that the proposed architectures can efficiently deployed in agriculture applications under real-field conditions (O'Grady et al., 2019), without the need of an expensive specialized equipment e.g., multispectral camera.

4. Conclusions

In this study the problem of reconstructed near-infrared images from RGB information was considered. Considering the importance of this problem in real-field precision agriculture applications, we developed two resource efficient methods for accurately estimating near-infrared signals from an RGB camera. Contrary to state of the art deep learning methods, the proposed models are much less computationally demanding and require significantly less training data and can be deployed to edge devices with computational and power constraints such as unmanned drones. Additionally, we illustrated the applicability of the shallow learning models in a precision agriculture application by precisely computing the normalized difference vegetation index (NVDI), and demonstrating that they can provide similar results with a multispectral camera. In the future we aim to examine the efficacy and applicability of the proposed models on different datasets concerning other agriculture applications.

CRediT authorship contribution statement

Alexandros Gkillas: Conceptualization, Data curation, Methodology, Investigation, Formal analysis, Validation, Funding acquisition, Writing – original draft, Writing – review & editing. **Dimitrios Kosmopoulos:** Conceptualization, Data curation, Methodology, Formal analysis, Validation, Writing – original draft, Writing – review & editing. **Kostas Berberidis:** Conceptualization, Data curation, Methodology, Formal analysis, Validation, Writing – original draft, 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.

Data availability

Data will be made available on request.

Acknowledgments

This work is partially supported by the Greek Secretariat for Research and Innovation and the EU, Project SOUP: SOilless culture UPgrade, T1EDK-04171 within the framework of "Competitiveness, Entrepreneurship and Innovation" (EPAnEK) Operational Programme 2014–2020. All authors have read and approved the final manuscript.

References

Abady, L., Barni, M., Garzelli, A., Tondi, B., 2020. GAN generation of synthetic multispectral satellite images. 11533, 122–133. https://doi.org/10.1117/12.2575765. URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11533/115330L/GAN-generation-of-synthetic-multispectral-satellite-images/10.1117/12.2575765.full https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11533/115330L/GAN-generation-of-synthetic-multispectral-satellite-images/10.1117/12.2575765.short.

Aharon, M., Elad, M., Bruckstein, A., 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. 54 (11), 4311–4322. http://dx.doi.org/10.1109/TSP.2006.881199.

Akhtar, N., Mian, A., 2020. Hyperspectral recovery from RGB images using Gaussian processes. IEEE Trans. Pattern Anal. Mach. Intell. 42 (1), 100–113. http://dx.doi.org/10.1109/TPAMI.2018.2873729.

Alvarez-Gila, A., Van De Weijer, J., Garrote, E., 2017. Adversarial networks for spatial context-aware spectral image reconstruction from RGB. In: Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, Vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., pp. 480–490. http://dx.doi.org/10.1109/ICCVW.2017.64, arXiv:1709.00265.

Arad, B., Ben-Shahar, O., 2016. Sparse recovery of hyperspectral signal from natural RGB images. In: European Conference on Computer Vision. Springer, Cham, pp. 19–34. http://dx.doi.org/10.1007/978-3-319-46478-7_2.

Arnal Barbedo, J.G., 2013. Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus 2 (1), 660. http://dx.doi.org/10.1186/2193-1801-2-660, URL: http://springerplus.springeropen.com/articles/10.1186/2193-1801-2-660.

Aslahishahri, M., Stanley, K.G., Duddu, H., Shirtliffe, S., Vail, S., Bett, K., Pozniak, C., Stavness, I., 2021. From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops. ICCVW, pp. 1312–1322. http://dx.doi.org/10.1109/ICCVW54120.2021.00152.

Brown, M., Süsstrunk, S., 2011. Multi-spectral SIFT for scene category recognition. In: CVPR 2011. pp. 177–184. http://dx.doi.org/10.1109/CVPR.2011.5995637.

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. Technical Report.

Calvini, R., Ulrici, A., Amigo, J.M., 2019. Growing applications of hyperspectral and multispectral imaging. In: Data Handling in Science and Technology, Vol. 32. Elsevier, pp. 605–629. http://dx.doi.org/10.1016/B978-0-444-63977-6.00024-9.

Chakraborty, S., Tiedemann, A., Teng, P., 2000. Climate change: potential impact on plant diseases. Environ. Pollut. 108 (3), 317–326. http://dx.doi.org/10.1016/S0269-7491(99)00210-9, URL: https://linkinghub.elsevier.com/retrieve/pii/S0269749199002109.

Charte, D., Charte, F., García, S., del Jesus, M.J., Herrera, F., 2018. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. Inf. Fusion 44, 78–96. http://dx.doi.org/10.1016/j.inffus.2017.12.007.

Chen, R., Huang, W., Huang, B., Sun, F., Fang, B., 2020. Reusing discriminators for encoding: Towards unsupervised image-to-image translation. http://dx.doi.org/10.48550/ARXIV.2003.00273, URL: https://arxiv.org/abs/2003.00273.

Fahrentrapp, J., Ria, F., Geilhausen, M., Panassiti, B., 2019. Detection of gray mold leaf infections prior to visual symptom appearance using a five-band multispectral sensor. Front. Plant Sci. 10, http://dx.doi.org/10.3389/fpls.2019.00628, URL: https://www.frontiersin.org/article/10.3389/fpls.2019.00628/full.

Garrett, K.A., Dendy, S.P., Frank, E.E., Rouse, M.N., Travers, S.E., 2006. Climate change effects on plant disease: Genomes to ecosystems. Annu. Rev. Phytopathol. 44 (1), 489–509. http://dx.doi.org/10.1146/annurev.phyto.44.070505.143420, URL: http://www.annualreviews.org/doi/10.1146/annurev.phyto.44.070505.143420.

Georgantopoulos, P., Papadimitriou, D., Constantinopoulos, C., Manios, T., Daliakopoulos, I., Kosmopoulos, D., 2023. A multispectral dataset for the detection of tuta absoluta and leveillula taurica in tomato plants. Smart Agric. Technol. 4, 100146. http://dx.doi.org/10.1016/j.atech.2022.100146, URL: https://www.sciencedirect.com/science/article/pii/S2772375522001101.

Gkillas, A., Ampeliotis, D., Berberidis, K., 2021. Fast sparse coding algorithms for piecewise smooth signals. In: 28th European Signal Processing Conference, EUSIPCO 2020.

Gkillas, A., Kosmopoulos, D., Constantinopoulos, C., Ampeliotis, D., Berberidis, K., 2021. A method for recovering near infrared information from RGB measurements with application in precision agriculture. In: 2021 29th European Signal Processing Conference. EUSIPCO, pp. 616–620. http://dx.doi.org/10.23919/EUSIPCO54536.2021.9615949.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR.

Jia, Y., Zheng, Y., Gu, L., Subpa-Asa, A., Lam, A., Sato, Y., Sato, I., 2017. From RGB to spectrum for natural scenes via manifold-based mapping. In: Proceedings of the IEEE International Conference on Computer Vision, Vol. 2017-October. Institute of Electrical and Electronics Engineers Inc., pp. 4715–4723. http://dx.doi.org/10.1109/ICCV.2017.504.

Jiang, Z., Huete, A.R., Chen, J., Chen, Y., Li, J., Yan, G., Zhang, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. Remote Sens. Environ. 101 (3), 366–378. http://dx.doi.org/10.1016/j.rse.2006.01.003.

Kaya, B., Can, Y.B., Timofte, R., 2019. Towards spectral estimation from a single RGB image in the wild. In: Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019. Institute of Electrical and Electronics Engineers Inc., pp. 3546–3555. http://dx.doi.org/10.1109/ICCVW.2019.00439, arXiv:1812.00805.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14 (5), 778–782. http://dx.doi.org/10.1109/LGRS.2017.2681128, URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85017192157&doi=10.1109%2fLGRS.2017.2681128&partnerID=40&md5=6c296283c35b5bb62ef0e1641ea33be8. cited By 235.

Li, J., Wu, C., Song, R., Li, Y., Liu, F., 2020. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Vol. 2020-June. IEEE Computer Society, pp. 1894–1903. http://dx.doi.org/10.1109/CVPRW50498.2020.00239, arXiv:2005.09305.

Lu, G., Fei, B., 2014. Medical hyperspectral imaging: a review. J. Biomed. Opt. 19 (1), 010901. http://dx.doi.org/10.1117/1.jbo.19.1.010901, URL: /pmc/articles/PMC3895860/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3895860/.

Mahlein, A.-K., 2015. Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping. Plant Disease 100, http://dx.doi.org/10.1094/PDIS-03-15-0340-FE.

Mahlein, A.K., Rumpf, T., Welke, P., Dehne, H.W., Plümer, L., Steiner, U., Oerke, E.C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sens. Environ. 128, 21–30. http://dx.doi.org/10.1016/j.rse.2012.09.019.

O'Grady, M.J., Langton, D., O'Hare, G.M.P., 2019. Edge computing: A tractable model for smart agriculture? Artif. Intell. Agric. 3, 42–51. http://dx.doi.org/10.1016/j.aiia.2019.12.001, URL: https://www.sciencedirect.com/science/article/pii/S2589721719300339.

Pang, Y., Lin, J., Qin, T., Chen, Z., 2022. Image-to-image translation: Methods and applications. IEEE Trans. Multimed. 24, 3859–3881. http://dx.doi.org/10.1109/TMM.2021.3109419.

Peng, H., Chen, X., Zhao, J., 2020. Residual pixel attention network for spectral reconstruction from RGB images. In: IEEE CVPRW, Vol. 2020-June. IEEE Computer Society, pp. 2012–2020. http://dx.doi.org/10.1109/CVPRW50498.2020.00251.

Rubinstein, R., Zibulevsky, M., Elad, M., 2008. Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. CS Tech. 40.

Sharma, N., Hefeeda, M., 2020. Hyperspectral reconstruction from RGB images for vein visualization. In: Proceedings of the 11th ACM Multimedia Systems Conference. MMSys '20, Association for Computing Machinery, New York, NY, USA, pp. 77–87. http://dx.doi.org/10.1145/3339825.3391861.

Shi, Z., Chen, C., Xiong, Z., Liu, D., Wu, F., 2018. HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images. In: IEEE CVPRW, Vol. 2018-June. IEEE Computer Society, pp. 1052–1060. http://dx.doi.org/10.1109/CVPRW.2018.00139.

Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. 2016, http://dx.doi.org/10.1155/2016/3289801.

Ustin, S.L., Jacquemoud, S., 2020. How the optical properties of leaves modify the absorption and scattering of energy and enhance leaf functionality. In: Remote Sensing of Plant Biodiversity. Springer International Publishing, pp. 349–384. http://dx.doi.org/10.1007/978-3-030-33157-3_14/FIGURES/16, URL: https://link.springer.com/chapter/10.1007/978-3-030-33157-3_14.

Wang, W., Cui, Z., Chang, H., Shan, S., Chen, X., 2014. Deeply Coupled Auto-encoder Networks for Cross-view Classification. Technical Report, arXiv:1402.2031v1.

Wu, J., Aeschbacher, J., Timofte, R., 2017. In defense of shallow learned spectral reconstruction from RGB images. In: Proceedings - 2017 IEEE International Conference on Computer Vision Workshops, ICCVW 2017, Vol. 2018-January. Institute of Electrical and Electronics Engineers Inc., pp. 471–479. http://dx.doi.org/10.1109/ICCVW.2017.63.

Yan, L., Wang, X., Zhao, M., Kaloorazi, M., Chen, J., Rahardja, S., 2020. Reconstruction of hyperspectral data from RGB images with prior category information. IEEE Trans. Comput. Imaging 6, 1070–1081. http://dx.doi.org/10.1109/TCI.2020.3000320.

Zhang, Y., Yang, W., Zhang, W., Yu, J., Zhang, J., Yang, Y., Lu, Y., Tang, W., 2022. Two-step ResUp&Down generative adversarial network to reconstruct multispectral image from aerial RGB image. Comput. Electron. Agric. 192, 106617. http://dx.doi.org/10.1016/J.COMPAG.2021.106617.

Zhao, Y., Po, L.M., Yan, Q., Liu, W., Lin, T., 2020. Hierarchical regression network for spectral reconstruction from RGB images. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Vol. 2020-June. IEEE Computer Society, pp. 1695–1704. http://dx.doi.org/10.1109/CVPRW50498.2020.00219, arXiv:2005.04703.

Zhu, Z., Liu, H., Hou, J., Jia, S., Zhang, Q., 2021a. Deep amended gradient descent for efficient spectral reconstruction from single RGB images. IEEE Trans. Comput. Imaging 7, 1176–1188. http://dx.doi.org/10.1109/TCI.2021.3124364.

Zhu, Z., Liu, H., Hou, J., Zeng, H., Zhang, Q., 2021b. Semantic-embedded unsupervised spectral reconstruction from single RGB images in the wild. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. ICCV, pp. 2279–2288.

Zhu, J.-Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. ICCV.