



Deep convolutional neural network for damaged vegetation segmentation from RGB images based on virtual NIR-channel estimation



Artzai Picon ^{a,*}, Arantza Bereciartua-Perez ^a, Itziar Eguskiza ^b, Javier Romero-Rodriguez ^c, Carlos Javier Jimenez-Ruiz ^c, Till Eggers ^d, Christian Klukas ^d, Ramon Navarra-Mestre ^d

^a TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Tecnológico de Bizkaia, C/ Geldo. Edificio 700, E-48160 Derio - Bizkaia, Spain

^b University of the Basque Country, Plaza Torres Quevedo, 48013 Bilbao, Spain

^c BASF Espanola S.L. Carretera A376, 41710 Utrera Sevilla, Spain

^d BASF SE, Speyererstrasse 2, 67117 Limburgerhof, Germany

ARTICLE INFO

Article history:

Received 30 June 2022

Received in revised form 12 September 2022

Accepted 12 September 2022

Available online 24 September 2022

Keywords:

Vegetation indices estimation

Vegetation coverage map

Near infrared estimation

Convolutional neural network

Deep learning

ABSTRACT

Performing accurate and automated semantic segmentation of vegetation is a first algorithmic step towards more complex models that can extract accurate biological information on crop health, weed presence and phenological state, among others. Traditionally, models based on normalized difference vegetation index (NDVI), near infrared channel (NIR) or RGB have been a good indicator of vegetation presence. However, these methods are not suitable for accurately segmenting vegetation showing damage, which precludes their use for downstream phenotyping algorithms. In this paper, we propose a comprehensive method for robust vegetation segmentation in RGB images that can cope with damaged vegetation. The method consists of a first regression convolutional neural network to estimate a virtual NIR channel from an RGB image. Second, we compute two newly proposed vegetation indices from this estimated virtual NIR: the infrared-dark channel subtraction (IDCS) and infrared-dark channel ratio (IDCR) indices. Finally, both the RGB image and the estimated indices are fed into a semantic segmentation deep convolutional neural network to train a model to segment vegetation regardless of damage or condition. The model was tested on 84 plots containing thirteen vegetation species showing different degrees of damage and acquired over 28 days. The results show that the best segmentation is obtained when the input image is augmented with the proposed virtual NIR channel ($F_1=0.94$) and with the proposed IDCR and IDCS vegetation indices ($F_1=0.95$) derived from the estimated NIR channel, while the use of only the image or RGB indices lead to inferior performance (RGB($F_1=0.90$) NIR($F_1=0.82$) or NDVI($F_1=0.89$) channel). The proposed method provides an end-to-end land cover map segmentation method directly from simple RGB images and has been successfully validated in real field conditions.

© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vegetation coverage map estimation is of great importance as a first stage of more complex algorithms aiming to automatically assess crop status, measure the effect of nutrients, evaluate the stress situation in a crop or quantify the effect of existing agricultural practices (Bendig et al., 2015; Picon et al., 2022a). Generating an accurate vegetation segmentation map serves other algorithms and models to perform subsequent and precise assessments over this segmented segmentation maps such as damage estimation (Picon et al., 2019a), weeds analysis (Picon et al., 2022a) or presence of plagues (Bereciartua-Pérez et al., 2022) among others.

Traditionally, vegetation index calculations (Bannari et al., 1995) have been used for vegetation coverage estimation. Leaves present

particularly low reflectance on the visible light (450–750 nm) range except for the fairly small window of the visible spectrum which is the green color, the signature reflectance of chlorophyll, around 550 nm. The rest of the visible wavelengths have minor representation. This encouraged researchers to define vegetation indexes to find indicators by combining more spectral wavelengths than just the ones on the visible range. In their 2014 research review Li et al. (2014) claim that the leaf mesophyll –that we can imagine as the leaf fleshy tissue– reflects low light in the visible spectrum, but has a major contribution to near-infrared (700–1200 nm). Moreover, they say that NIR radiation can penetrate the vegetal canopy from the upper leaves to the lower ones, which makes the actual structure determinant to the final NIR reflectance. The canopy structure is composed of several factors such as leaf thickness, overlapping, height and growth habit among others. That is the main reason why NIR is considered to be best suited for estimating vegetal biomass. In fact, many Vegetation Indices (VIs) involve the

* Corresponding author.

E-mail address: artzai.picon@tecnalia.com (A. Picon).

combination of NIR reflectance with other light spectra and channels. A widely-used channel for vegetation-soil discrimination is Normalized Difference Vegetation Index -NDVI- (Rouse et al., 1974), which is a combination of NIR and the red -R- channel of the usual RGB color codification for visual cameras. These vegetation indexes not only serve to infer plant coverage map or biomass. Biomass-related Vegetation Indices such as green biomass reported by Gitelson et al. (2003), can provide information about leaf cover, leaf area index, chlorophyll per ground area and intercepted fraction of radiation with combinations of NIR and RGB channels. Those combinations included the widely known NDVI but also Simple Ratio (SR), Red Edge (λ_{RE}), Photochemical reflectance index (PRI), Structural Independent Pigment Index (SIPPI) and others. They noticed that by measuring such indices of water level, pigment, biomass, etc. they could infer if plants suffered disease, had a certain risk of fire or salt-stress among other situations.

Several works have tried to correlate the NDVI and other Vegetation Indices (VIs) based on light channels with vegetation coverage estimation (Price, 1992; Huete et al., 1997; Wu et al., 2007; Zhengwei et al., 2009; Roth and Streit, 2018; Devia et al., 2019; Ren and Zhou, 2019). Most of them are based on acquiring experimental data on soil and leaf reflectance of different light channels and correlate those reflectance with combinations of such channels (which are called Vegetation Indices). They also try to correlate those Vegetation Indices with the actual biomass measurements (fresh weight, dry weight) by means of linear and non-linear regressive analysis. The aforementioned research works correlate the different VIs with the actual biomass weighting by means of pixel-wise linear or non-linear regression analysis.

An additional problem is the necessity of the NIR image channel in addition to the standard RGB channels to obtain these vegetation indices. This presents two important drawbacks: First, high price and low accessibility of these cameras which are often accompanied by lack of availability of specific knowledge for the end-user. And second, the impossibility of addressing specific use cases where standard, low-cost devices or light acquisition devices are required. This is of special relevance on Unmanned Aerial Vehicles -UAVs- such as drones or cell-phone based applications (Johannes et al., 2017) among others.

However, these indices lack the capability of generating complex models (Hemming and Rath, 2001) as they are based on single pixel information. Image processing based methods integrate spatial information over RGB images or vegetation index channels that can cope with complicated tasks such as specie identification (Hemming and Rath, 2001), disease classification (Johannes et al., 2017; Huddar et al., 2012), insect counting (Bereciartua-Pérez et al., 2022) among others. However, it is with the advent of deep learning techniques when image processing based models have become capable to perform complex image identification tasks with equal or higher performance than humans (Picon et al., 2019a) in scenarios such as radiology (Yala et al., 2019), board games (Granter et al., 2017) among others. In precision agriculture, deep neural networks have been successful for pest classification (Picon et al., 2019a; Picon et al., 2019b; Argüeso et al., 2020), crop and weed segmentation (Milioto et al., 2018; Sa et al., 2018; Picon et al., 2022a), insect counting (Bereciartua-Pérez et al., 2022). The combination of vegetation indices with image processing and analysis algorithms have been successfully used for more complex applications such as forest dynamics analysis (Sader and Winne, 1992), irrigated rice mapping (Nguyen et al., 2012), environment quality analysis (Fung and Siu, 2000) or crop identification (Jakubauskas et al., 2002). Although these RGB based models have demonstrated capable of performing vegetation segmentation (Hassanein et al., 2018; Netto et al., 2018), they have not been tested to segment damaged vegetation.

2. Related works

One of the research lines to overcome the need for specific multispectral acquisition devices is the generation of algorithms capable for virtually estimating the near infrared channel from a RGB image. In

this sense, several authors have worked on this field for the last few years. Most of the authors only use pixel information and infer the NIR channel based on the values of the Red, Green and Blue channels of that pixel (Rabaté et al., 2011a; Rabaté et al., 2011b). In this sense, Arai et al. (2016) found a linear correlation between NIR and Green color channel, allowing them to estimate NIR reflectance using a conventional RGB camera through a regression analysis. Furthermore they calculated the NDVI index with images from a visible camera mounted on a drone using this method where other authors have analysed the hyperspectral endmembers to develop a pixel based method to estimate NIR images from RGB data (de Lima et al., 2019). These methods, although simple and fast, do not exploit spatial information contained on the pixel neighborhood. These intensity level relationships contain information on visual shapes and textures (Picón et al., 2009; Picon et al., 2011) that allows a more accurate estimation of the NIR channel and better performance of image processing methods.

Some more modern research integrates the spatial information distribution from the RGB pixels to infer the NIR channel. For example, Khan et al. (2018); Lima et al., 2019 propose a method to estimate several vegetation indices by means of a neural network. However, they use a neural network that do not estimate the vegetation indexes at pixel level resolution, only the average vegetation index for the whole input tile with an average coefficient of determination of $R^2 = 0.92$. Recent methods have taken advantage of convolutional neural networks for NIR channel estimation. For example, Aslahishahri et al. (2021) and de Lima et al. (2022) employed pix2pix (Isola et al., 2017) to accurately estimate NIR channel from RGB images from UNet based generators (Ronneberger et al., 2015). These pix2pix methods have been successfully extended on the medical domain (Picon et al., 2021; Picon et al., 2022b) by employing more efficient loss functions and generator architectures based on fully convolutional DenseNet (Jégou et al., 2017) which is more parameter efficient than traditional UNet.

In this work, we propose and validate an end-to-end method for robust damaged vegetation segmentation over RGB images without the need for an infrared capable camera. This method can estimate a virtual NIR channel from an RGB image and use it to feed a vegetation segmentation neural network with an extended image obtaining better results than RGB based segmentation models regardless its damage condition.

This method provides the following contributions:

- The definition of two new vegetation indices: Infrared-Dark-Channel-Subtractive index (IDCS) and the Infrared-Dark-Channel Ratio index (IDCR) which are sensitive to vegetation coverage map estimation on situations of plant damage presence.
- A convolutional semantic regression network to estimate near infrared channel from RGB image (RGB2NIR) that can optionally incorporate an adversarial loss.
- A vegetation biomass coverage estimation semantic segmentation network that takes as input a multichannel image composed by the R, G, B channels and the (estimated) NIR, and vegetation indexes.
- A end to end method that takes a RGB image, estimates the required indices and segments the vegetation coverage map of the image.

3. Materials

In order to develop and validate the proposed models, thirteen vegetal species were selected: three crops: GLXMA (*Glycine max*), HELAN (*Helianthus annuus*), ZEAMX (*Zea mays*), seven broad leaf species: ABUTH (*Abutilon theophrasti*), AMARE (*Amaranthus retroflexus*), CHEAL (*Chenopodium album*), DATST (*Datura stramonium*), POLCO (*Fallopia convolvulus*), POROL (*Portulaca oleracea*), SOLNI (*Solanum nigrum*) and three grass species: DIGSA (*Digitaria sanguinalis*), ECHCG (*Echinochloa crus-galli*), SETVE (*Setaria verticillata*). 84 plots were planted combining the presence of the different species. 24 plots contained GLXMA, 24 HELAN and 24 ZEAMX, whereas the weeds were randomly distributed

among the main plots. Each crop field (plot) followed a different weed control treatment to generate damage on the different species.

Each plot image was acquired with a Micasense RedEdge MX camera from a 2 meters height. This camera has five different sensors: blue(450 nm), green(560 nm), red(650 nm), red-edge(730 nm) and NIR(840 nm) and delivered 1280x920px images. As each image band was taken by a different monochrome sensor, acquired images were co-registered by applying the affine transformation that minimized the Mattes mutual information between channels (Klein et al., 2009; Shamonin et al., 2014) following the same method we performed in Picon et al. (2022b). Images were taken at different days after crop seeding (DAS = 14, 16, 32, 35, 38, 42, 44, 49). About 5% of the images were not correctly co-registered due to the short acquisition distance and were removed from the data set leading to 504 RGB-NIR registered images from the 84 plots. From these images, 358 images were randomly chosen, and the vegetation coverage was manually segmented using CVAT annotation tool. Fig. 1 shows some of the acquired plots.

To avoid bias, the distribution of images across the training, validation and test datasets was selected by plots. This means that each crop field (plot) was assigned an identification number and all images belonging to the same crop field (plot) were assigned to the same subset of data (train, validation, test). This ensures that images from the same plot taken on consecutive days are assigned to the same set, avoiding contamination of the training, validation and test sets. Eighty percent of the crop fields (plots) were randomly chosen for training, while the remaining 20% were distributed into validation and test sets, and all images were incorporated into the set determined by their crop field (plot) number resulting into 24 plots for training, 2 for validation and 3 for testing.

4. Proposed method

In our approach, we propose the use of a semantic regression neural network to estimate a virtual NIR channel from an RGB image. This virtual channel is then used to enrich the original RGB image to generate a multi-channel image that is used to train a multispectral vegetation segmentation convolutional neural network. This multichannel image includes not only the original red, green and blue channels and the estimated virtual NIR channel but also different multi-spectral indices derived from this four channels.

The intuition behind this is based on the fact that NIR channel is a good estimator for vegetation which is relatively robust to vegetation damage as it can be appreciated in Figs. 8 and 9 where the damaged parts of the plant present similar NIR response. Complementing this virtual NIR information into the original RGB image might help on vegetation segmentation.

Proposed method is depicted in Fig. 2. An RGB image passes through the RGB2NIR network, which is responsible for estimating the virtual near-infrared channel of the image. This virtual channel is used, together with the original RGB image to generate the additional image channels containing the vegetation indices (NDVI, IDCS and IDCRI). These generated channels are aggregated in a multichannel image. This enhanced and more informative image feeds a semantic segmentation neural network responsible for estimating the vegetation coverage map of the image.

Below, we present and detail the different modules for the end-to-end method for robust vegetation segmentation over RGB images.

4.1. RGB2NIR: estimation of near infrared channel from RGB images

The first module of the proposed method (rgb2nir module on Fig. 2) is responsible for estimating NIR information from RGB images. To this end, we employ a fully convolutional DenseNet architecture Jégou et al. (2017) similar to the one we used in Picon et al. (2021). This network combines the descriptive power from traditional segmentation networks based on fully convolutional versions such as SegNet (Badrinarayanan et al., 2017) where the accuracy for the border detection is provided by the skip connections on the U-Net segmentation network (Ronneberger et al., 2015).

Concretely, the proposed fully convolutional DenseNet network (Jégou et al., 2017) was set to an input size of 224x224 pixels. Architecture follows original paper implementation where the number of initial convolution filter was set to 48. The encoder was composed by 5 transition down blocks (TD) with 4 convolution layers each with a growing rate of 16. For the decoder part, we use 5 transition up (TU) layers each of them is linked with their corresponding encoder block. This allows recovering the input image high level details as these skip connections transfer the low level features and spatial information from the source domain into the detailed reconstruction of the target domain.

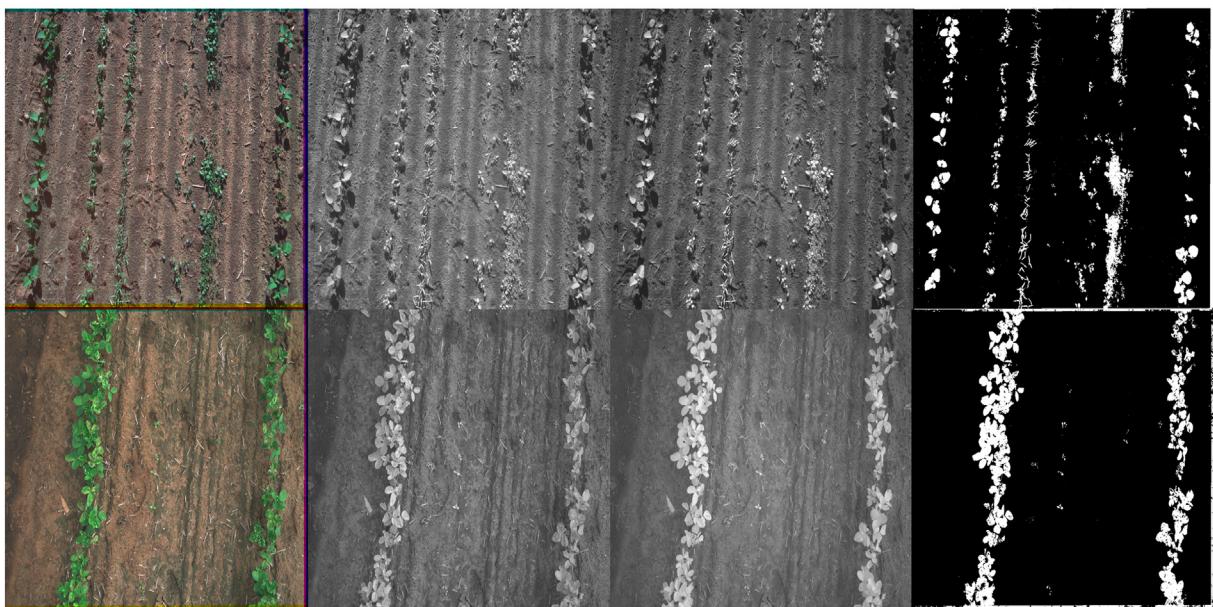


Fig. 1. Examples of the generated images: a) RGB Image, b) red-edge image, c) near-infrared image, d) Ground-truth of plant coverage.

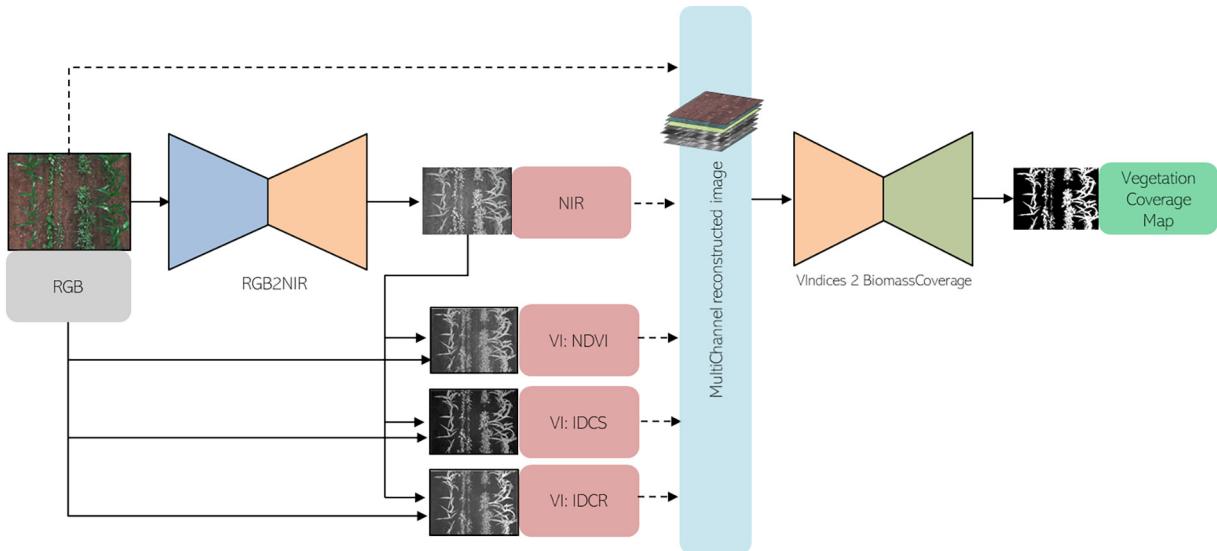


Fig. 2. Coverage estimation diagram Infrared-Dark-Channel Ratio index model diagram: An RGB image goes through the RGB2NIR transformation network and a virtual NIR channel is calculated. NDVI, Infrared-Dark-Channel Subtractive (IDCS), Infrared-Dark-Channel Ratio index (IDCR) channels are estimated by their corresponding formula from the R,G,B and estimated NIR channels. All these channels are used to generate a multi-channel image that is fed into a vegetation segmentation convolutional neural network (see Section 4.3).

The last layer of this network has been substituted by a sigmoid activation function and the loss function has been replaced by a mean absolute error loss in order to learn a pixel-wise regression transformation that translates the image from the source to the target domain. The last layer consists of a 224x224x1 that performs NIR reconstruction. The network is trained by minimizing a mean absolute error loss function which can be enhanced by an adversarial perceptual loss function following a pix2pix architecture (Isola et al., 2017). The addition of the perceptual adversarial loss functions ensures that the generated image is visually plausible and the estimated image cannot be distinguished from a real image by a specifically trained network. Similar approach has been followed by Aslahishahri et al. (2021); de Lima et al., 2022 to include a perceptual adversarial loss. However, they use UNet architecture as generator which increases the number of trainable parameters. Proposed fully convolutional DenseNet network has 2,3 M parameters whereas its UNet counterpart has 24 M instead.

4.2. Vegetation indices for vegetation estimation

Existing vegetation indices for vegetation segmentation fails under the presence of plant damage or direct illumination. This is caused by the fact that the presence of damage on the plant normally increases the reflectance on particular visible wavelengths that makes indices such as NDVI or negative CIE-a (Johannes et al., 2017) channels to fail on appropriately segmenting non-healthy vegetation. Additionally, changes on illumination intensity reduce the robustness for other biomarkers such as NIR channel due to intensity scale changes between the dark and the illuminated areas of the image. In order to overcome this issue, we propose two new indices that are vaguely inspired on the Dark Channel Prior (He et al., 2010; Galdran et al., 2015) method. This method is used to estimate haze on the images for image restoration purposes. It estimates the haze level by considering the minimum value of the R, G and B channels taking advantage of the whitish halo created by fog on a image. Dark channel is calculated as the minimum of the red, green and blue channels on the spatial neighborhood of each pixel. However, for our approach, we will just consider the minimum value for the R, G and B channels for each pixel without considering their neighborhood (Section 4.3).

$$DC(R, G, B) = \min(R, G, B). \quad (1)$$

In this work, we adapt this formulation to propose two new vegetation indices: the Infrared-Dark-Channel Subtractive (IDCS) (Eq. (3)) index, that reflects the relative intensity of the NIR channel by subtracting the dark channel and the Infrared-Dark-Channel Ratio index (IDCR) (Eq. (2)), that reflects the intensity ratio of the NIR channel against the dark channel. this formulation can be extended for multispectral or hyperspectral images just getting the minimum for all the spectral range or for specific spectral ranges.

$$IDCS = NIR - \min(R, G, B, NIR). \quad (2)$$

$$IDCR = NIR / (\min(R, G, B, NIR) + \epsilon). \quad (3)$$

If we analyse Fig. 3, the proposed indices correlate the presence of vegetation on a plot image and are more robust to the existence of damage on the vegetation. Fig. 3 shows an example of a RGB image of a plot and its corresponding NIR channels and NDVI values together with the proposed IDCS and IDCR indices. It can be appreciated that the proposed IDCS and IDCR vegetation indices can separate better between vegetation and non-vegetation pixels even for the unhealthy part of the plants. In order to get quantitative metrics for the vegetation-soil separation capabilities of the proposed indices, proposed indices were compared against other indices and color channels: r, g and b channels, CIE-L, CIE-a and CIE-b color channels from CIELab Zhang and Wandell (1997), NDVI Rouse et al. (1974) and NIR channel. On one hand, we calculate the intersection of the probability density function (Lee et al., 2005) from the indices values between the vegetation and the non-vegetation classes. This metric measures the existing overlap between the distribution of the intensity values for the two classes (soil and vegetation), showing a intersection of 0.0 a perfect separability among the classes whereas a value of 1.0 indicates a full overlap among classes. On the other hand, we also measure the Area Under the Curve Metric (Fawcett, 2006) of a hypothetical Naïve Bayes classifier applied over the vegetation indices for the different classes. This curve measures the true positives rate against the false positives rate. An ideal classifier will present a ROC value of 1.0 whereas a random classifier will arise to a ROC value of 0.5. Table 1 shows the average results and standard deviation obtained with each vegetation index. It can be appreciated that NIR, NDVI, and CIE-a channels are appropriate vegetation indices for vegetation estimation with AuC values of 0.916, 0.989 and 0.937 respectively.

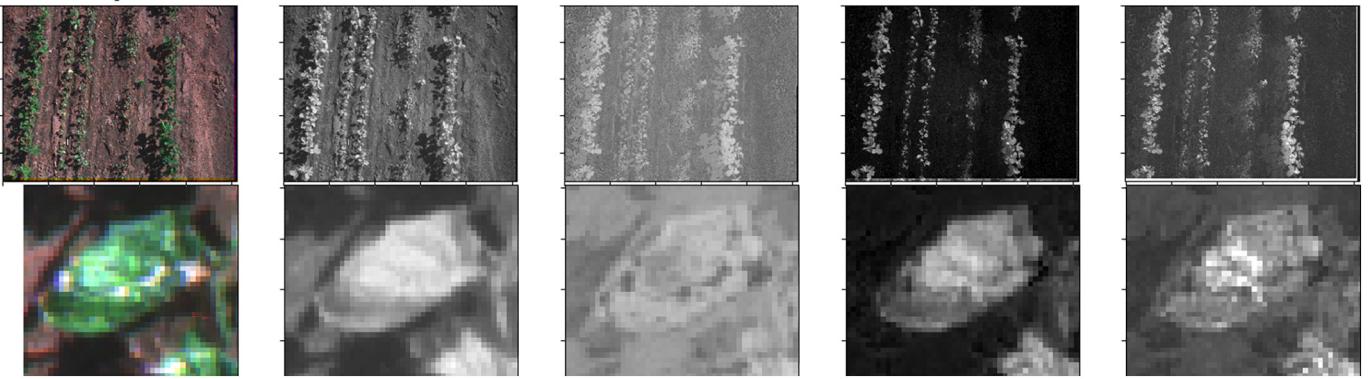


Fig. 3. Plot image: a) RGB Image, b) NIR image, c) NDVI, d) Proposed IDCS, e) Proposed IDCR. Bottom close-up of an unhealthy leaf.

However, best results are obtained by the proposed IDCS and IDCR estimated channels that obtain an AuC value of 0.998 and 0.997 with an histogram intersection value of 0.166 and 0.198. This can be explained as this vegetation index is more robust to the damage of the plants and to non-homogeneous image illumination. Figs. 4 and 5 depict the probability density distribution of the intensity values of the pixels containing vegetation against the pixels that contains other elements for a given image.

4.3. Vegetation segmentation convolutional neural network

A second network with similar structure as the one defined in Section 4.1 is used for plant coverage estimation. In this case, the input layer of the network size is $M \times N \times K$, where M and N represents the height and the width of the input image and K the number of channels used. In our case the number of input channels is $K = 7$. These channels are composed by the R, G, B channels of the original image and all the estimated channels by the previous methods (NIR, NDVI, IDCS and IDCR). The final layer of this network is composed by two output channels of size M and N ($M \times N \times 2$) resembling the original size of the image. one of the output channels maps the estimation for the plant coverage segmentation meanwhile the other contains the other classes. A softmax activation layer that ensures the mutually exclusiveness of the two classes. This network is minimized over a categorical cross-entropy loss function.

5. Results

5.1. NIR channel estimation from RGB image

The two models for the estimation of the near infrared channel from RGB images defined in Section 4.1 were trained over the training set of

Table 1
AuC and histogram intersection values obtained by the different image vegetation indices or channels for each image of the entire dataset.

Image Channel	AuC (average) \pm std	Intersection \pm std
r	0.872 \pm 0.085	0.463 \pm 0.132
g	0.666 \pm 0.108	0.551 \pm 0.230
b	0.736 \pm 0.110	0.601 \pm 0.117
CIE-L	0.678 \pm 0.113	0.586 \pm 0.172
CIE-a	0.937 \pm 0.049	0.257 \pm 0.173
CIE-b	0.740 \pm 0.089	0.298 \pm 0.213
NIR	0.916 \pm 0.071	0.551 \pm 0.189
NDVI	0.989 \pm 0.008	0.261 \pm 0.081
IDCS	0.998 \pm 0.002	0.166 \pm 0.126
IDCR	0.997 \pm 0.004	0.198 \pm 0.120

described in Section 3 for 100 epochs on the mae loss approach and 40 epochs for the pix2pix approach. Training was performed over 224x224 pixel tiles that were extracted randomly from the full-size images. As size of the image is 1280x920 pixels, the tiling process ensure an equivalent number of approximately 12096 tiles for training considering no tile overlap. In order to generate more variability on the input dataset, several augmentation techniques were performed on the images such as shifting, rotating and scaling. To simulate light changing conditions, RGB and NIR intensity channels are multiplied by the random constant factor simulating changes on light intensity coherent with dichromatic reflection model (Tominaga, 2020). These tiles are fed into the neural network as described in Section 4.1. Adam optimiser was employed for training and the learning rate was set to 10^{-5} . A reduction of the learning rate is performed when the loss function value on the validation set raises or stagnates.

Table 2 shows inference results for the proposed models. Both Pearson's coefficient and the mean absolute error are shown as performance metrics. We have compared the proposed DenseNet based model with Aslahishahri et al. (2021) based UNet model. We can appreciate that metrics are slightly better when using the proposed DenseNet architecture rather than the UNet architecture. This might be caused by the better parameter efficiency of the DenseNet model. Loss evolution and regression graph are depicted for the winning model in Figs. 7 and 6 respectively. It is noteworthy to remark that, in the pix2pix related model, only mean average error loss is depicted as the adversarial loss is based on competition among the discriminator and adversarial loss part in the generator. Obtained regression results (Fig. 6) show the correlation graph between the real NIR values and the estimated values.

The use of fully convolutional DenseNet provides better reconstruction performance than using UNet architecture. It can be also appreciated that the use of a pix2pix (Isola et al., 2017) based adversarial loss contributes not only to generate more plausible images but also to reduce the error between the predictions and the real NIR images that present a lower error rate (5%) and better correlation coefficient ($r = 0.96$). This endorses the pix2pix based design approach (Aslahishahri et al., 2021; de Lima et al., 2022). Figs. 8 and 9 show examples of the estimation of image tiles for both configurations. It can be appreciated that the NIR estimation is accurate, even for damaged part of the plants, which consist of whitish necrosis spots along the image.

5.2. Vegetation coverage map estimation

A semantic segmentation network as defined in Section 4.3 is used to estimate the plant coverage map. For that, the network was trained over the training set of described in Section 3. Training was performed over 224x224 pixel tiles that were extracted randomly from the full-size images during 30 epochs. The tiles are fed into the near infrared estimation

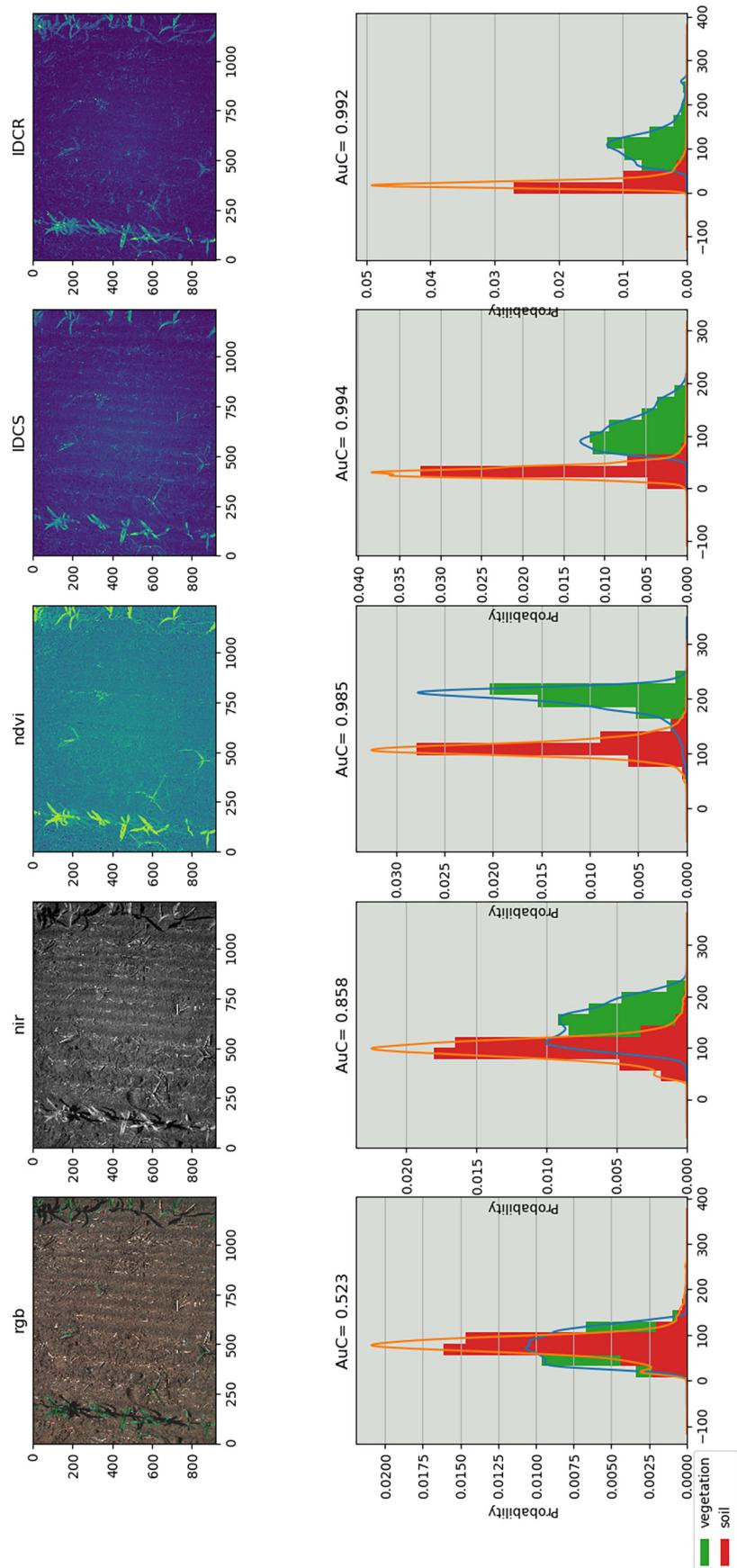


Fig. 4. Top) Field images representing the target vegetation index; bottom) Probability distribution of the pixels containing vegetation and not vegetation for their respective vegetation index for their corresponding image.

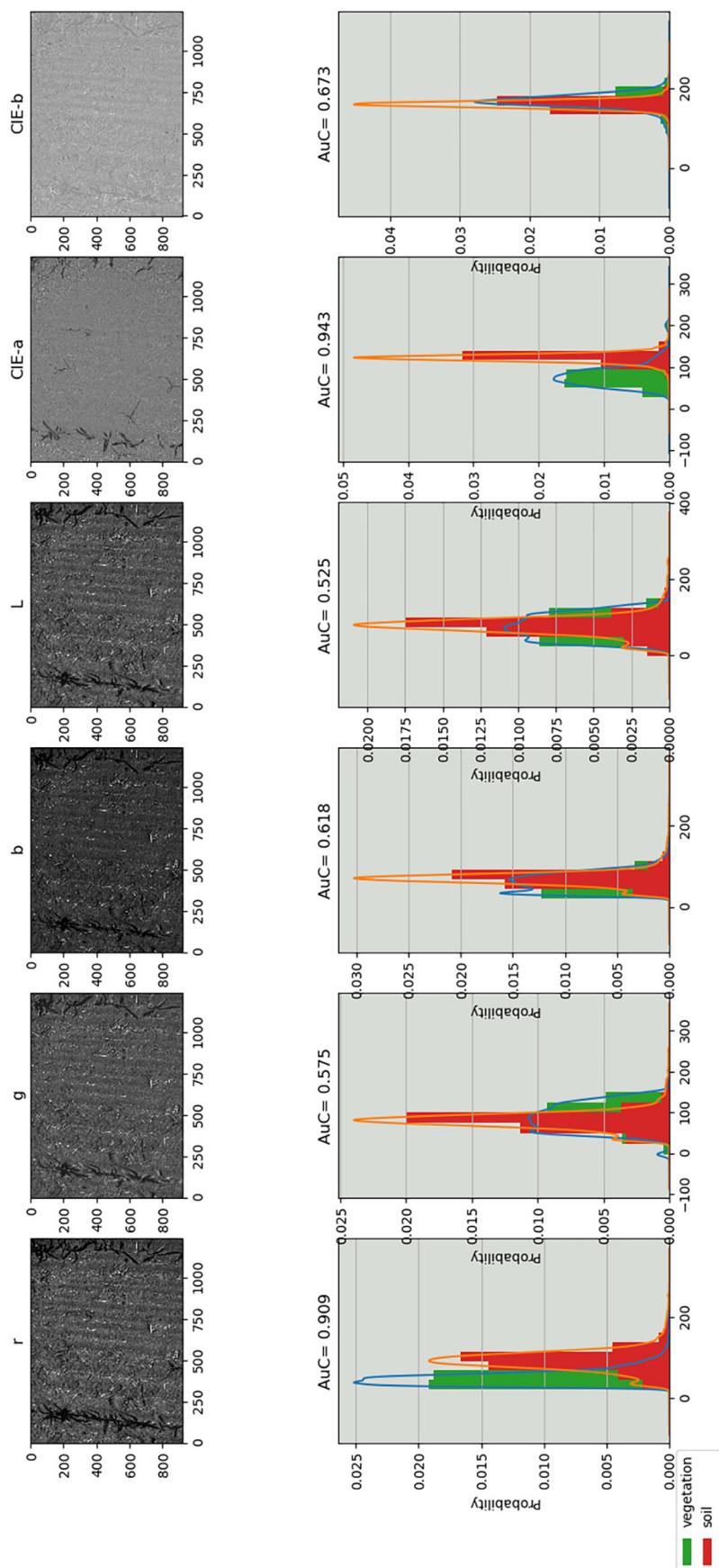


Fig. 5. Top) Images representing the target vegetation index, bottom) Probability distribution of the pixels containing vegetation and not vegetation for their respective vegetation index.

Table 2

Performance of the two NIR estimation models given by their Pearson coefficient r and the mean absolute error (MAE).

Algorithm name	Network	Loss	r	MAE
RGB2NIR (ours)	Fully Convolutional DenseNet	mean average error (MAE)	0.93	0.05
RGB2NIR pix2pix (ours)	Fully Convolutional DenseNet	MAE with adversarial loss (pix2pix)	0.96	0.04
RGB2NIR	UNet Aslahishahri et al. (2021)	mean average error (MAE)	0.91	0.06
RGB2NIR pix2pix	UNet Aslahishahri et al. (2021)	MAE with adversarial loss (pix2pix)	0.94	0.05

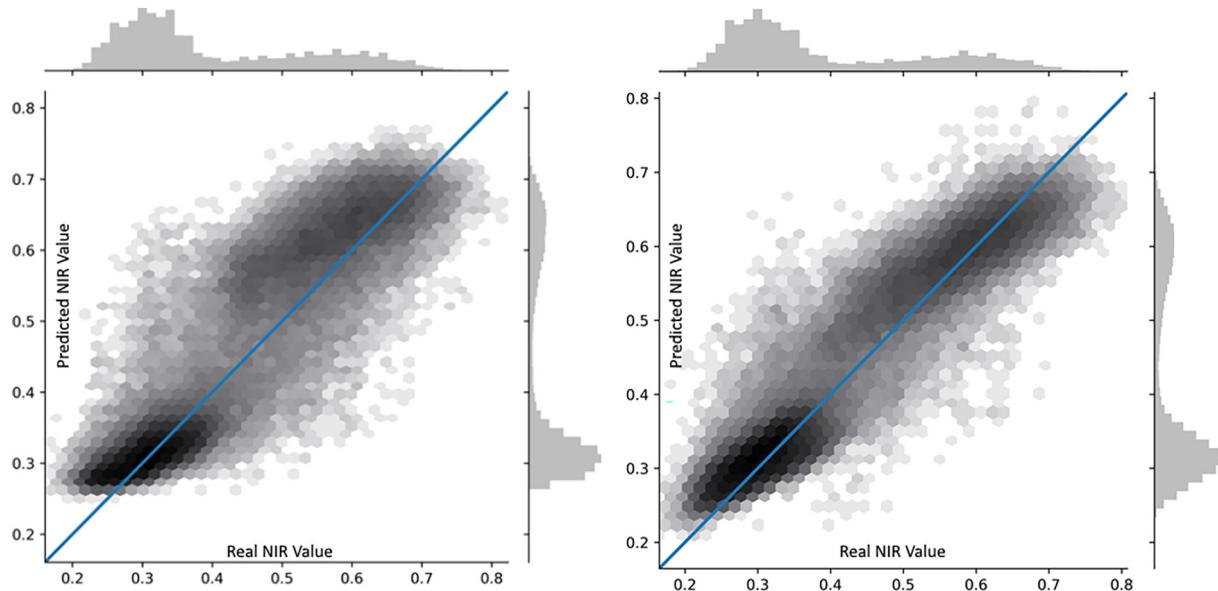


Fig. 6. Regression graphs between real and predicted NIR values for the proposed model. Left) Mean Average Error loss, Right) Mean Average Error + Adversarial loss.

neural network to get the estimated NIR channel. Additional vegetation indices (NDVI, IDCS and IDCRI) are calculated from the R, G, B and the estimated NIR channels. Training is performed over the training set of the dataset described in 3. After training, the validation subset of the dataset was used to calculate the optimal thresholds values that maximized the balanced accuracy (BAC). These thresholds were applied over the testing set. In order to measure the effect of using estimated NIR channels instead the real ones, specific experiments using real NIR channel have been performed.

We performed an ablation study showing different channel combinations for the multi-spectral input image that feeds the vegetation segmentation neural network. Results of the vegetation coverage map estimation over the testing set are depicted on Table 3.

Performance of the different algorithms is analysed based on two common metrics for semantic segmentation: 1) The balanced accuracy (BAC), which represents the average value between true positive rate and true negative rate (Eq. (4)), and 2) the F-Score (Eq. (7)) that returns the geometric average among the precision (Eq. (6)) and recall (Eq. (5))

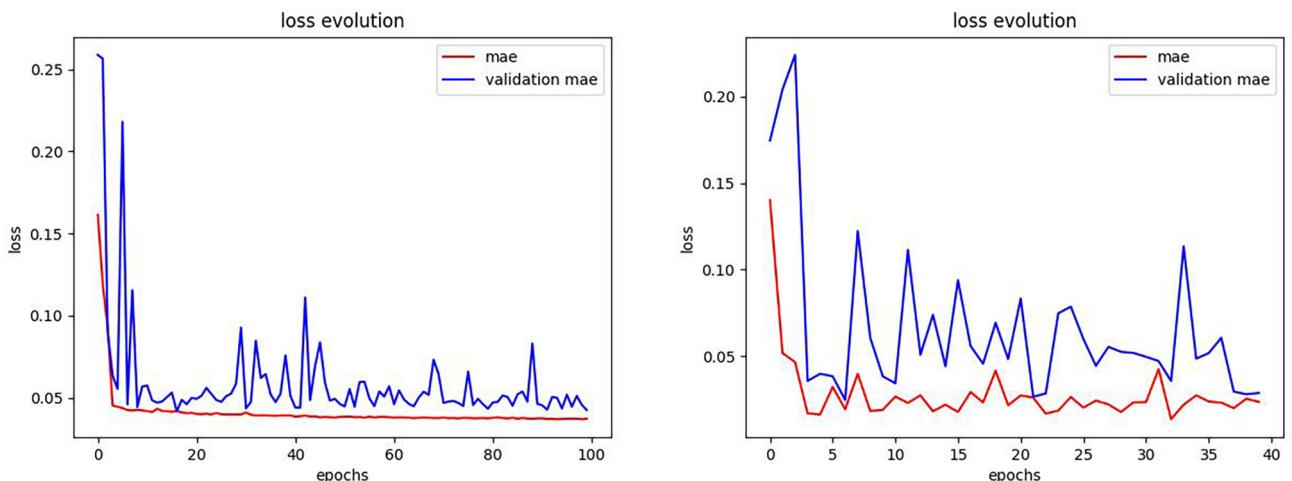


Fig. 7. Evolution of Mean Average Error (MAE) and validation Mean Average Error for the proposed model. Left) Mean Average Error trained model, Right) Mean Average Error (MAE) + Adversarial loss trained model. In adversarial model only MAE loss is plotted.

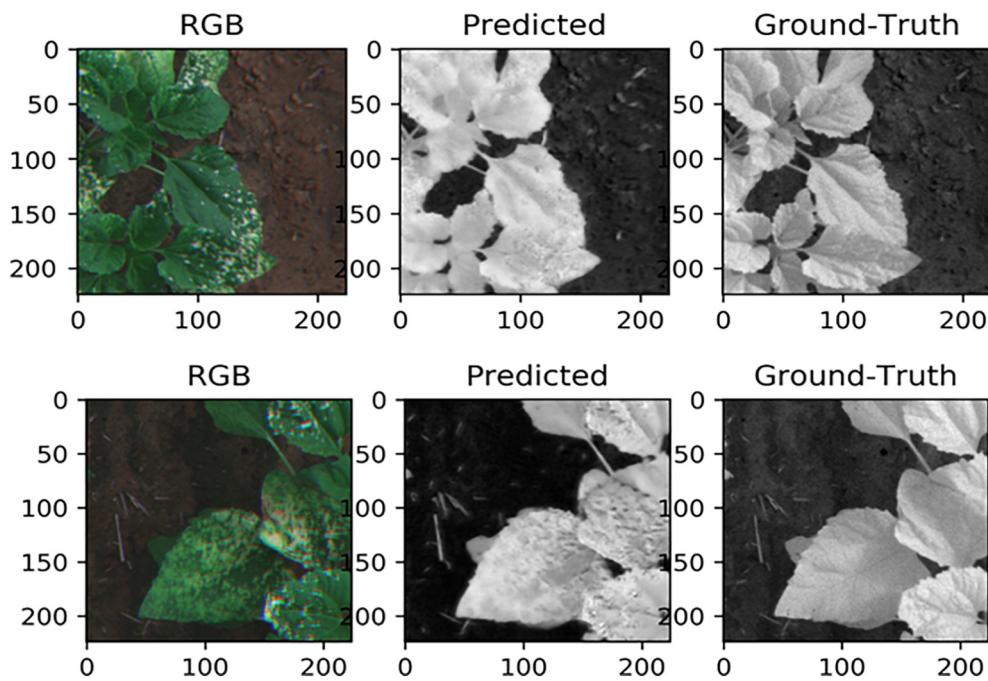


Fig. 8. Examples of NIR channel prediction with the RGB2NIR algorithm. NIR reconstruction can be appreciated for the healthy and unhealthy (white necrosis spots) of the plant.

of the model. The precision is calculated as the number of true vegetation pixels divided by the number of all predicted vegetation pixels whereas the recall is the number of true positive vegetation pixels divided by the number of all true vegetation pixels. For unbalanced datasets in semantic segmentation, F-Score is normally preferred to BAC as it ignores the effect of the true negative class (non vegetation pixels). TP, TN, FN and FP refers to true positives, true negatives, false negatives and false positives respectively.

$$BAC = (((TP/(TP + FN) + (TN/(TN + FP)))/2. \quad (4)$$

$$Recall = TP/(TP + FN) \quad (5)$$

$$Precision = TP/(TP + FP) \quad (6)$$

$$F_1 = 2 * (Precision * Recall)/(Precision + Recall) \quad (7)$$

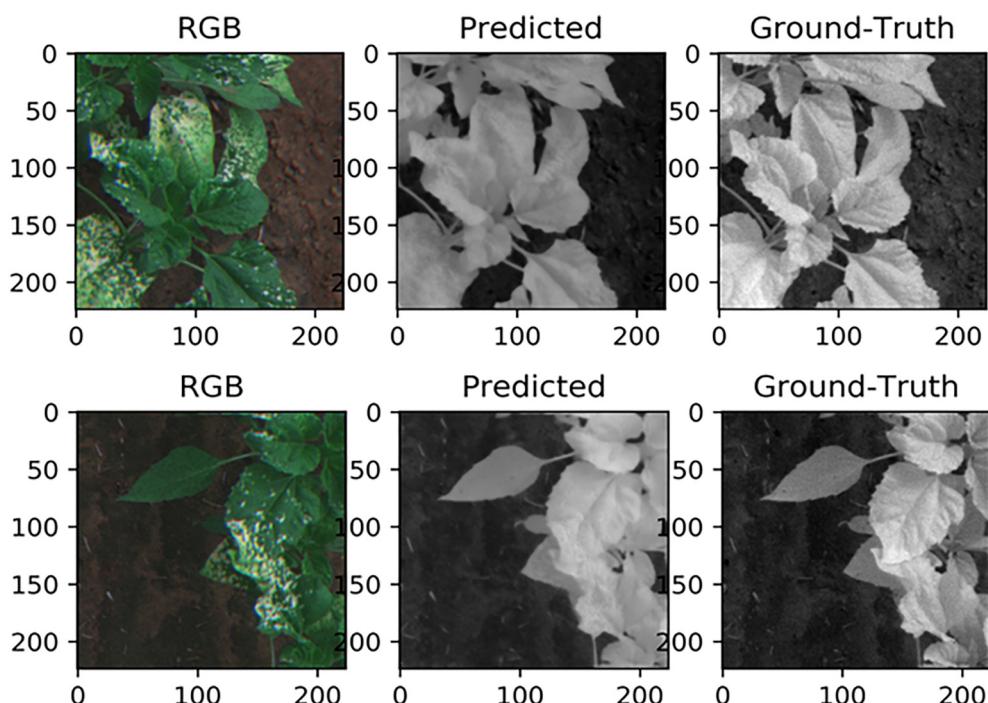


Fig. 9. Examples of NIR channel prediction with the RGB2NIR pix2pix algorithm. NIR reconstruction can be appreciated for the healthy and unhealthy (white necrosis spots) of the plant.

Table 3

Results for the different algorithm combination for vegetation coverage map estimation.

Image channels	# Channels	AuC	BAC	Sensitivity	Specificity	NPV	PPV	F1
NIR (Real)	1	0.981	0.937	0.916	0.925	0.979	0.769	0.836
NIR	1	0.977	0.924	0.926	0.922	0.981	0.741	0.823
NDVI	1	0.991	0.958	0.967	0.950	0.992	0.823	0.889
IDCS	1	0.996	0.974	0.982	0.966	0.995	0.874	0.925
IDCR	1	0.992	0.964	0.972	0.956	0.993	0.841	0.902
RGB	3	0.992	0.962	0.967	0.958	0.992	0.848	0.904
RGB + NIR (Real)	4	0.995	0.976	0.978	0.979	0.997	0.903	0.939
RGB + NIR	4	0.997	0.978	0.982	0.974	0.996	0.901	0.940
RGB + NDVI	4	0.997	0.977	0.980	0.973	0.995	0.900	0.939
RGB + IDCS	4	0.995	0.967	0.964	0.969	0.992	0.878	0.919
RGB + IDCR	4	0.998	0.981	0.986	0.974	0.997	0.903	0.943
RGB + NIR + NDVI + IDCS + IDCR	7	0.998	0.980	0.980	0.996	0.914	0.913	0.946
RGB + NIR + IDCS + IDCR	6	0.998	0.984	0.988	0.980	0.997	0.919	0.952

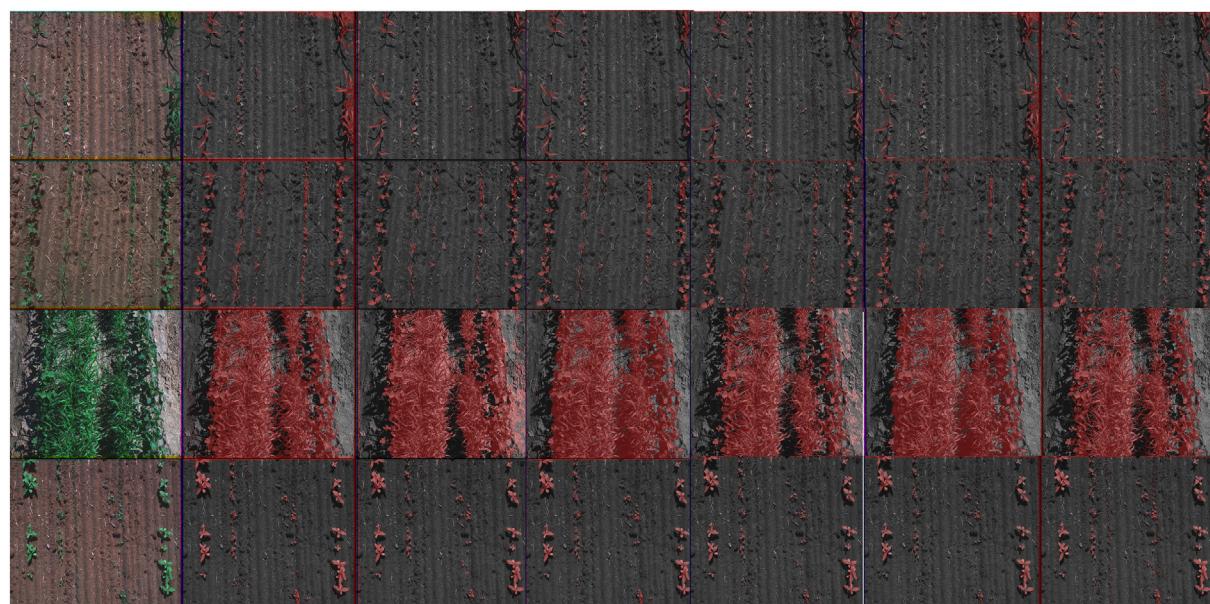


Fig. 10. This figure shows examples of coverage map estimation for the different combinations. a) Original image, b) RGB, c) NIR, d) NDVI, e) IDCS, f) IDCR, g) GroundTruth.

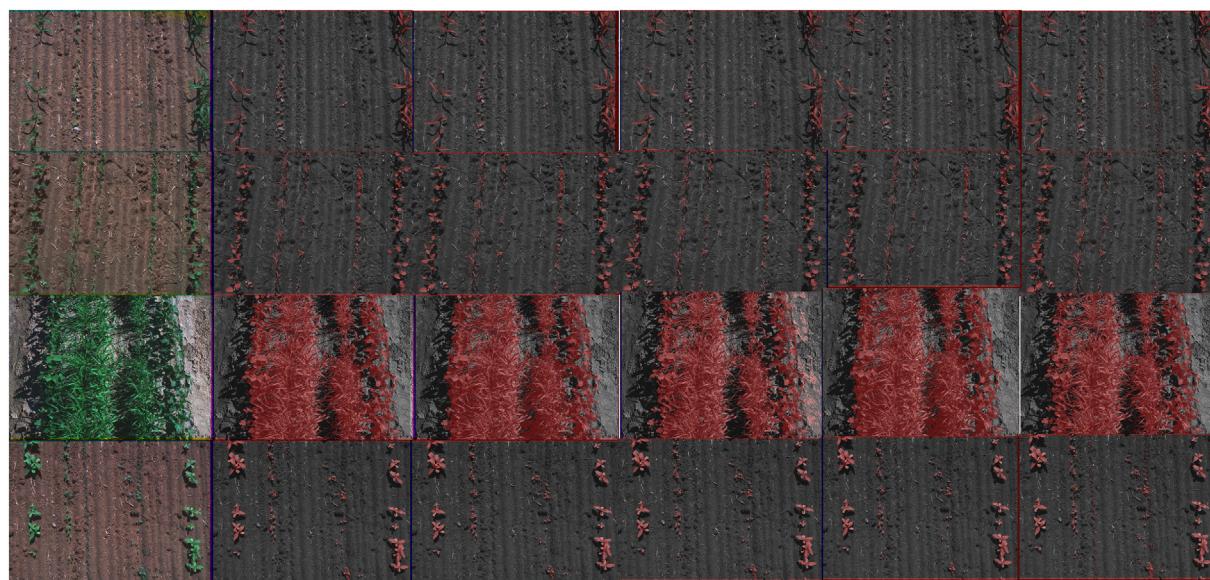


Fig. 11. This figure shows examples of coverage map estimation for the different combinations. a) Original image, b) RGB + NIR, c) RGB + NDVI, d) RGB + IDCS, e) RGB + IDCR, f) GroundTruth.

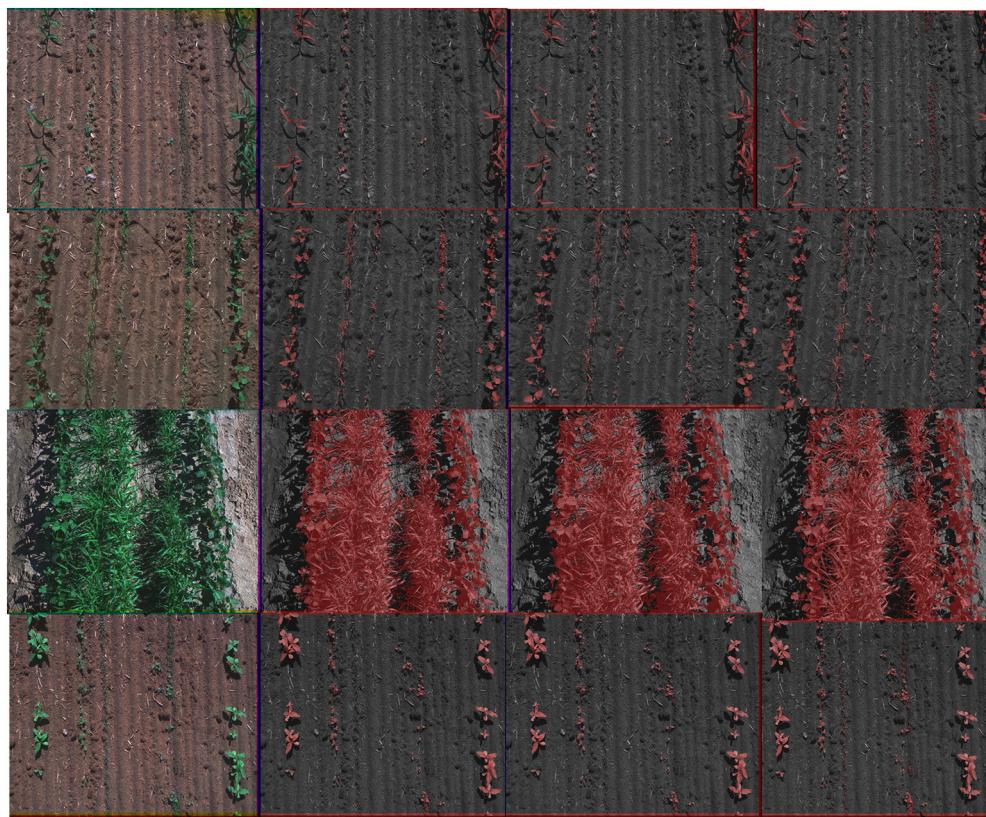


Fig. 12. This figure shows examples of coverage map estimation for the different combinations. a) Original image, b) RGB + NDVI + NIR + IDCS + IDC, c) RGB + NIR + IDCS + IDC, d) GroundTruth.

As we can appreciate in Table 3, the use of the RGB image alone produces a F-score of 0.904. When using just one channel/index to generate vegetation map, we can appreciate that the use of already existing vegetation indices such as NDVI and NIR offer a reduced performance (F-score of 0.823 and 0.889 respectively). However, using just one of the proposed indices for vegetation estimation raises better results IDCS (F-Score=0.925), IDC (F-Score=0.902). If we analyse the effect of using real NIR channels or estimated ones, we can appreciate that it is almost equivalent to use estimated or real NIR channels. Fig. 10 shows examples of segmented fields under the different combinations.

Fig. 11 shows segmentation results for the combination of RGB with one of the vegetation indices. Results on Table 2 show that the combination of RGB channels with the proposed indices deal to better results than the RGB baseline. The combination of RGB with the proposed IDC index obtains the best F-Score with a value of 0.943. However, it is with the combination with several of the proposed vegetation indices when better results are achieved. The best results (see Fig. 12) are obtained with the combination of RGB + NIR + IDCS + IDC achieving a BAC=0.984 and a F-score=0.952.

6. Conclusions

In this work, we have presented an end-to-end method for robust vegetation segmentation over RGB images which is able to appropriately segment vegetation even when vegetation presents damaged conditions without the need of an infrared capable camera.

We have proposed a convolutional semantic regression network to estimate a virtual near infrared channel from an RGB image (RGB2NIR) that can optionally incorporate an adversarial loss. With this adversarial loss, the proposed network can accurately estimate the NIR channel ($p_{value} = 0.96$, RMS=4%). This demonstrates that the adversarial loss contributes to generate more efficient estimation for the NIR channel than just employing a convolutional semantic regression network.

We have introduced two novel vegetation indices such as Infrared-Dark-Channel-Subtractive index (IDCS) and the Infrared-Dark-Channel Ratio index (IDC). We have proven that these indices have good separability properties to differentiate vegetation regardless its damage. These vegetation indices can be used independently for vegetation segmentation purposes independently.

We have generated a vegetation segmentation network to segment damaged vegetation. When feeding the shown model using only RGB image achieves a F1-score of 0.90. This segmentation performance increases when the RGB image is extended with the proposed virtual NIR channel (F1=0.94) or the with proposed vegetation indices (F1=0.95) that are derived from the estimated NIR channel.

The proposed NIR estimation method could be adapted in the future to be applied not only for vegetation coverage estimation but to other agricultural use cases where NIR information might be relevant. This will reduce the need for for without the need for expensive IR cameras that could extend the application range for these methods.

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.

References

- Arai, K., Gondoh, K., Shigetomi, O., Miura, Y., 2016. Method for nir reflectance estimation with visible camera data based on regression for ndvi estimation and its application for insect damage detection of rice paddy fields. *Int. J. Adv. Res. Artif. Intell.* 5, 17–22.
- Argüeso, D., Picon, A., Irusta, U., Medela, A., San-Emerito, M.G., Bereciartua, A., Alvarez-Gila, A., 2020. Few-shot learning approach for plant disease classification using images taken in the field. *Comput. Electron. Agric.* 175, 105542.
- 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: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1312–1322.
- Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495.
- Bannari, A., Morin, D., Bonn, F., Huete, A., 1995. A review of vegetation indices. *Remote Sens. Rev.* 13, 95–120.
- Bendig, J., Yu, K., Aasen, H., Bolten, A., Bennertz, S., Broscheit, J., Gnyp, M.L., Bareth, G., 2015. Combining uav-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* 39, 79–87.
- Bereciartua-Pérez, A., Gómez, L., Picón, A., Navarra-Mestre, R., Klukas, C., Eggers, T., 2022. Insect counting through deep learning-based density maps estimation. *Comput. Electron. Agric.* 197, 106933.
- Devia, C.A., Rojas, J.P., Petro, E., Martínez, C., Mondragon, I.F., Patino, D., Rebolledo, M.C., Colorado, J., 2019. High-throughput biomass estimation in rice crops using uav multispectral imagery. *J. Intell. Robot. Syst.* 96, 573–589.
- Fawcett, T., 2006. An introduction to roc analysis. *Pattern Recognit. Lett.* 27, 861–874.
- Fung, T., Siu, W., 2000. Environmental quality and its changes, an analysis using ndvi. *Int. J. Remote Sens.* 21, 1011–1024.
- Galdran, A., Pardo, D., Picón, A., Alvarez-Gila, A., 2015. Automatic red-channel underwater image restoration. *J. Vis. Commun. Image Represent.* 26, 132–145.
- Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophys. Res. Lett.* 30.
- Granter, S.R., Beck, A.H., Papke Jr, D.J., 2017. Alphago, deep learning, and the future of the human microscopist. *Arch. Pathol. Lab. Med.* 141, 619–621.
- Hassanein, M., Lari, Z., El-Sheiny, N., 2018. A new vegetation segmentation approach for cropped fields based on threshold detection from hue histograms. *Sensors* 18, 1253.
- He, K., Sun, J., Tang, X., 2010. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* 33, 2341–2353.
- Hemming, J., Rath, T., 2001. Pa-precision agriculture: computer-vision-based weed identification under field conditions using controlled lighting. *J. Agric. Eng. Res.* 78, 233–243.
- Huddar, S.R., Gowri, S., Keerthana, K., Vasanthi, S., Rupanagudi, S.R., 2012. Novel algorithm for segmentation and automatic identification of pests on plants using image processing. 2012 Third International Conference on Computing, Communication and Networking Technologies (ICCCNT'12). IEEE, pp. 1–5.
- Huete, A., Liu, H., Batchily, K., Van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of tm images for eos-modis. *Remote Sens. Environ.* 59, 440–451.
- 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, pp. 1125–1134.
- Jakabauskas, M.E., Legates, D.R., Kastens, J.H., 2002. Crop identification using harmonic analysis of time-series avhrr ndvi data. *Comput. Electron. Agric.* 37, 127–139.
- Jégou, S., Drozdal, M., Vazquez, D., Romero, A., Bengio, Y., 2017. The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 11–19.
- Johannes, A., Picon, A., Alvarez-Gila, A., Echazarra, J., Rodriguez-Vaamonde, S., Navajas, A.D., Ortiz-Barredo, A., 2017. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* 138, 200–209.
- Khan, Z., Rahimi-Eichi, V., Haefele, S., Garnett, T., Miklavcic, S.J., 2018. Estimation of vegetation indices for high-throughput phenotyping of wheat using aerial imaging. *Plant Methods* 14, 20.
- Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2009. Elastix: a toolbox for intensity-based medical image registration. *IEEE Trans. Med. Imaging* 29, 196–205.
- Lee, S., Xin, J., Westland, S., 2005. Evaluation of image similarity by histogram intersectionvol. 30, Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur pp. 265–274.
- Li, L., Zhang, Q., Huang, D., 2014. A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111.
- de Lima, D.C., Saqui, D., Ataky, S., Jorge, L.A.d.C., Ferreira, E.J., Saito, J.H., 2019. Estimating agriculture nir images from aerial rgb data. International Conference on Computational Science. Springer, pp. 562–574.
- de Lima, D.C., Saqui, D., Mpinda, S.A.T., Saito, J.H., 2022. Pix2pix network to estimate agricultural near infrared images from rgb data. *Can. J. Remote Sens.* 48, 299–315.
- Lima, D.C.d., Saqui, D., Ataky, S., Jorge, L.A.d.C., Ferreira, E.J., Saito, J.H., 2019. Estimating agriculture nir images from aerial rgb data. International Conference on Computational Science. Springer, pp. 562–574.
- Milioti, A., Lottes, P., Stachniss, C., 2018. Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in cnns. *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, pp. 2229–2235.
- Netto, A.F.A., Martins, R.N., de Souza, G.S.A., de Moura Araújo, G., de Almeida, S.L.H., Capelini, V.A., 2018. Segmentation of rgb images using different vegetation indices and thresholding methods. *Nativa* 6, 389–394.
- Nguyen, T.T.H., De Bie, C., Ali, A., Smaling, E., Chu, T.H., 2012. Mapping the irrigated rice cropping patterns of the mekong delta, vietnam, through hyper-temporal spot ndvi image analysis. *Int. J. Remote Sens.* 33, 415–434.
- Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., Johannes, A., 2019a. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* 161, 280–290.
- Picon, A., Ghita, O., Rodriguez-Vaamonde, S., Iriondo, P.M., Whelan, P.F., 2011. Biologically-inspired data decorrelation for hyper-spectral imaging. *EURASIP J. Adv. Signal Process.* 2011, 1–10.
- Picón, A., Ghita, O., Whelan, P.F., Iriondo, P.M., 2009. Fuzzy spectral and spatial feature integration for classification of nonferrous materials in hyperspectral data. *IEEE Trans. Industr. Inf.* 5, 483–494.
- Picon, A., Medela, A., Sánchez-Peralta, L.F., Cicchi, R., Bilbao, R., Alfieri, D., Elola, A., Glover, B., Saratxaga, C.L., 2021. Autofluorescence image reconstruction and virtual staining for in-vivo optical biopsying. *IEEE Access* 9, 32081–32093.
- Picon, A., San-Emeterio, M.G., Bereciartua-Perez, A., Klukas, C., Eggers, T., Navarra-Mestre, R., 2022a. Deep learning-based segmentation of multiple species of weeds and corn crop using synthetic and real image datasets. *Comput. Electron. Agric.* 194, 106719.
- Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., Echazarra, J., 2019b. Crop conditional convolutional neural networks for massive multi-crop plant disease classification over cell phone acquired images taken on real field conditions. *Comput. Electron. Agric.* 167, 105093.
- Picon, A., Terradillos, E., Sánchez-Peralta, L.F., Mattana, S., Cicchi, R., Glover, B.J., Arbide, N., Velasco, J., Etzezarraga, M.C., Pavone, F.S., et al., 2022b. Novel pixelwise co-registered hematoxylin-eosin and multiphoton microscopy image dataset for human colon lesion diagnosis. *J. Pathol. Inform.* 13, 100012.
- Price, J.C., 1992. Estimating vegetation amount from visible and near infrared reflectances. *Remote Sens. Environ.* 41, 29–34.
- Rabatel, G., Gorretta, N., Labbé, S., 2011a. Getting ndvi spectral bands from a single standard rgb digital camera: a methodological approach. Conference of the Spanish Association for Artificial Intelligence. Springer, pp. 333–342.
- Rabatel, G., Gorretta, N., Labbé, S., 2011b. Getting ndvi spectral bands from a single standard rgb digital camera: a methodological approach. Conference of the Spanish Association for Artificial Intelligence. Springer, pp. 333–342.
- Ren, H., Zhou, G., 2019. Estimating green biomass ratio with remote sensing in arid grasslands. *Ecol. Ind.* 98, 568–574.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.
- Roth, L., Streit, B., 2018. Predicting cover crop biomass by lightweight uas-based rgb and nir photography: an applied photogrammetric approach. *Precision Agric.* 19, 93–114.
- Rouse, J., Haas, R., Schell, J., Deering, D., 1974. Monitoring vegetation systems in the great plains with ertsvol. 351. NASA special publication pp. 309.
- Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R., 2018. Weedmap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. *Remote Sens.* 10, 1423.
- Sader, S., Winne, J., 1992. Rgb-ndvi colour composites for visualizing forest change dynamics. *Int. J. Remote Sens.* 13, 3055–3067.
- Shamonin, D.P., Bron, E.E., Lelieveldt, B.P., Smits, M., Klein, S., Staring, M., 2014. Fast parallel image registration on cpu and gpu for diagnostic classification of alzheimer's disease. *Front. Neuroinform.* 7, 50.
- Tominaga, S., 2020. Dichromatic reflection model. Computer Vision: A Reference Guide, pp. 1–3.
- Wu, J., Wang, D., Bauer, M.E., 2007. Assessing broadband vegetation indices and quickbird data in estimating leaf area index of corn and potato canopies. *Field Crops Res.* 102, 33–42.
- Yala, A., Lehman, C., Schuster, T., Portnoi, T., Barzilay, R., 2019. A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology* 292, 60–66.
- Zhang, X., Wandell, B.A., 1997. A spatial extension of cielab for digital color-image reproduction. *J. Soc. Inform. Display* 5, 61–63.
- Zhengwei, Y., Hu, Z., Liping, D., Yu, G., 2009. A comparison of vegetation indices for corn and soybean vegetation condition monitoring. Geoscience and remote sensing symposium. IGARSS.