Original papers

Complementary chemometrics and deep learning for semantic segmentation of tall and wide visible and near-infrared spectral images of plants

Puneet Mishra a,*, Roy Sadeh b, Ehud Bino b, Gerrit Polder c, Martin P. Boer d, Douglas N. Rutledge e, f, Ittai Herrmann b

a Wageningen University & Research, Food and Biobased Research, P.O. Box 17, 6700AA Wageningen, the Netherlands
b The Robert H. Smith Institute for Plant Sciences and Genetics in Agriculture, Faculty of Agriculture, Food and Environment, The Hebrew University of Jerusalem, P.O. Box 12, Rehovot 7610001, Israel
c Wageningen University & Research, Greenhouse Horticulture Group, P.O. Box 644, 6700AP Wageningen, the Netherlands
d Wageningen University & Research, Biometris, P.O. Box 16, 6700 AA Wageningen, the Netherlands
e Université Paris-Saclay, INRAe, AgroParisTech, UMR SayFood, 75005 Paris, France
f National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia

ARTICLE INFO

Keywords:
U-net
Semantic segmentation
Phenotyping
Artificial intelligence
Hyperspectral imagery

ABSTRACT

Close range spectra imaging of agricultural plants is widely performed to support digital plant phenotyping, a task where physicochemical changes in plants are monitored in a non-destructive way. A major step before analyzing the spectral images of plants is to distinguish the plant from the background. Usually, this is an easy task and can be performed using mathematical operations on the combinations of selected spectral bands, such as estimating the normalized difference vegetative index (NDVI). However, when the background of plants contains objects with similar spectral properties as plant then the segmentation based on the threshold of NDVI images can suffer. Another common approach is to train pixel classifiers on spectra extracted from selected locations in the spectral image, but such an approach does not take the spatial information about the plant structure into account. From a technical perspective, plant spectral imaging for digital phenotyping applications usually involves imaging several plants together for a comparative purpose, hence, the imaging scene is relatively big in terms of memory. To solve the challenge of plant segmentation and handling the memory challenge, this study proposes a novel approach, which combines chemometrics with advanced deep learning (DL) based semantic segmentation. The approach has four key steps. As a first step, the spectral image is pre-processed to reduce illumination effects present in the close-range spectral images of plants resulting from the interaction of light with complex plant geometry. Different chemometric pre-processing methods were explored to find possible improvements in the segmentation performance of the DL model. The second step was to perform a principal components analysis (PCA) to reduce the dimensionality of the images, thus drastically reducing their size so that they can be handled more easily using the available computer memory during the training of the DL model. As the third step, small random images (128 × 128) were subsampled from the tall and wide image matrices to generate the training and validation sets for training the DL models. In the last step, a U-net based deep semantic segmentation model was trained and validated on the sub-sampled spectral images. The results showed that the proposed approach allowed efficient handling and training of the DL segmentation model. The intersection over union (IoU) scores for the segmentation was 0.96 for the independent test set image. The segmentation based on variable sorting for normalization and standard normal variate pre-processed data achieved the highest IoU scores. A combination of chemometrics and DL led to an efficient segmentation of tall and wide spectral images which otherwise would have given out-of-memory errors. The developed method can facilitate digital phenotyping tasks where close-range spectral imaging is used to estimate the physicochemical properties of plants.

* Corresponding author.
E-mail address: puneet.mishra@wur.nl (P. Mishra).

https://doi.org/10.1016/j.compag.2021.106226
Received 15 April 2021; Received in revised form 20 May 2021; Accepted 23 May 2021
Available online 29 May 2021
0168-1699/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Introduction

Spectral imaging is a popular sensing technique that captures the spatially resolved spectral properties of samples (Gowen et al., 2007; Mishra et al., 2020a; Polder and Gowen, 2020). Spectral imaging is the combination of two sensing modalities i.e., imaging and spectroscopy where the imaging modality captures the spatial properties while the spectroscopy captures the spectral properties (Amigo et al., 2013; Mobarak and Amigo, 2018). Spectral imaging is widely used for non-destructive and rapid analysis of materials ranging from foods to high-end pharmaceutical tablets (Mobarak and Amigo, 2018). Furthermore, based on the desired application, spectral imaging can be performed using spectral information ranging from high energy X-rays (Du et al., 2019) to low energy terahertz wavelengths (Gowen et al., 2012).

Recently, interest has grown in the use of close-range spectral imaging for agricultural plants (Mishra et al., 2017; Mishra et al., 2020a; Mishra et al., 2020c), that is whole plants and not just plucked leaves (Mishra et al., 2020a). Close-range spectral imaging of plants should not be confused with satellite and airborne (Hank et al., 2019) spectral cameras which have been used for many years to image vegetation scenes (Mishra et al., 2020a). Unlike remotely sensed spectral imaging, close-range spectral imaging is very recent and only a few applications exist showing its proper usage for whole plant analysis, especially for digital plant phenotyping (Asaari et al., 2018; Mishra et al., 2019a; Mishra et al., 2020a; Mishra et al., 2019b), where the aim is to monitor morpho-physiological traits of a variety of genotypes under challenging environmental conditions in order to select the best performing genotypes for eventual commercial growth (Costa et al., 2019; Fiersuschka and Schurr, 2019; Roitsch et al., 2019; Yang et al., 2020). Such monitoring is to be performed on the same plant throughout the growth cycle to understand how the plant evolves, which means that destructive plant organ sampling is not the preferred approach as it can affect the experiment by inducing stress in the plants (Rahaman et al., 2019; Roitsch et al., 2019). Hence, a non-invasive, non-destructive technique like spectral imaging is preferable as it allows to measure plants physico-chemical properties while minimally influencing the experiment (Mishra et al., 2017; Mishra et al., 2020a).

Spectral imaging in the visible (Vis; 400–700 nm) and near-infrared (NIR; 700–2500 nm) spectral range is of particular interest for the physicochemical as well as the morphophysiological analysis of agricultural plants (Mishra et al., 2017; Mishra et al., 2020a; Mishra et al., 2020c). The Vis spectral region captures the colour of the key photosynthetic pigments (i.e., chlorophyll, anthocyanins and carotenoids), present in the plants (Féret et al., 2021; Jacquemoud and Baret, 1990; Jay et al., 2016). The NIR spectral region characterizes the internal structure of the leaves by the light scattering, and the concentration of chemical components such as lignin, water, cellulose, proteins, sugars, and fats, by the absorption of NIR light by functional groups such as OH, CH, NH, and SH (Féret et al., 2021; Jacquemoud and Baret, 1990). Applications of Vis-NIR spectral imaging digital plant phenotyping can be found related to both quantitative (Briglia et al., 2019; Bruning et al., 2019; Ge et al., 2019; Liu et al., 2019; Mazis et al., 2020) and qualitative analysis (Asaari et al., 2019; Asaari et al., 2018; Mishra et al., 2019a; Mishra et al., 2020c; Mishra et al., 2019b).

The exploitation of Vis-NIR spectral imaging data requires advanced chemometrics or machine learning approaches (Amigo et al., 2013; Herrmann et al., 2013). The primary step of the spectral imaging data processing is to extract the pure plant spectra by distinguishing them from the non-relevant background present in the scene (Asaari et al., 2019; Asaari et al., 2018; Mishra et al., 2017). Once the spectra related to plants is extracted efficiently then classical chemometrics and machine learning approaches can be utilized for qualitative and quantitative analysis (Mishra et al., 2017). Usually, this segmentation is an easy task and can be performed by setting thresholds on the normalized difference vegetation index (NDVI) images. Alternatively, a classifier can be trained on a manual selection of spectra from the plant and the background, and then used to classify each pixel of the image (Asaari et al., 2019; Asaari et al., 2020; Mishra et al., 2017). However, these approaches are good for cases where the background has spectra that are quite different from the plants. In many practical cases, as will be presented later in this study, the soil was partly covered by green algae and moss that have spectra like those of the plants. Hence, a better solution would be to use approaches that simultaneously model the spatial and spectral information in the data to jointly learn from the plant physical shape as well as its spectral characteristics to eliminate the non-relevant background.

Recent advances in computer vision, and in deep learning (DL) approaches, have opened up new possibilities for the analysis of spectral images (Audebert et al., 2019; Paoletti et al., 2019). Of particular interest are convolutional neural networks (CNNs) which allow joint modelling of both the spatial and spectral dimensions (Audebert et al., 2019; Paoletti et al., 2019; Polder et al., 2019). CNNs can also be used for image segmentation and, in the domain of computer vision, the task is commonly referred to as semantic segmentation (Garcia-Garcia et al., 2018). There are several DL models available for semantic segmentation of RGB images, such as fully connected neural networks (Long et al., 2015), U-net (Ronneberger et al., 2015), Deep Lab (Chen et al., 2018) and their variants (Garcia-Garcia et al., 2018; Goodfellow et al., 2016). However, applications of DL semantic segmentation models are still lacking for close-range spectral image segmentation. Several reasons can account for this, but the two main ones are the lack of large close-range spectral image data sets and the presence of many bands which leads to out of memory errors during modelling. There is also currently a lack of a link between DL methods and the basic chemometric knowledge about spectroscopy which could play a complementary role in producing high accuracy models. For example, in the case of close-range spectral images of plants, spectral normalization and derivative techniques from the chemometrics domain can be used to correct for illumination effects (Asaari et al., 2019; Asaari et al., 2018; Mishra et al., 2019a; Mishra et al., 2020b; Mishra et al., 2019b) thus, reducing the additive and multiplicative effects in the spectra of plants, thus improving the performance of models. Hence, combining chemometric pre-processing with the DL approaches can be highly relevant and, in the following sections, this work explores the effects of spectral pre-processing on the performance of DL models. Furthermore, a solution from the chemometric domain to deal with the fact that spectral images consist of hundreds of bands that are often highly correlated (Amigo et al., 2013), is to perform a spectral decorrelation by principal components analysis (PCA) (Amigo et al., 2013; Mobarak and Amigo, 2018). Performing a principal components analysis (PCA) before the DL modelling reduces the size of the spectral image and allows the DL models to be trained (Alkhayat et al., 2020) without out-of-memory errors during the modelling.

This study proposes a combination of chemometrics and DL for semantic segmentation of close-range spectral images of plants. The approach has four key steps. As a first step, the spectral image is preprocessed to reduce the illumination effects present in the close-range spectral images of plants caused by the interaction of light with complex plant geometry (Mishra et al., 2020a; Mishra et al., 2020b). Different chemometric pre-processing methods are explored to improve the segmentation performance of the DL model. Then, a PCA (Bro and Smilde, 2014) is performed to reduce the dimensionality of the images, so that they could be easily handled within the limits of the computer memory while training the model. As the band dimensionality is low, random images are subsampled from the tall and wide images to generate the training and validation sets for training the DL models. Finally, as the last step, a U-net (Ronneberger et al., 2015) based deep semantic segmentation model is trained and, as a performance evaluation, is tested on the independent test set spectral image. The performance of this approach is demonstrated in the case of close range spectral imaging data of Arabidopsis thaliana potted plants within the framework of a digital phenotyping experiment.
2. Materials and methods

2.1. Data set

Two spectral image scenes, each consisting of eighteen Arabidopsis thaliana plants, were imaged with a line-scan spectral camera (HySPEX VNIR-1800, Norsk Elektro Optikk, Oslo, Norway) in the visible near-infrared spectral range i.e., 407–997 nm with a sampling interval of 3.26 nm. For imaging, plants were illuminated with two tungsten halogen bulbs (12v each) mounted at 45° elevation from the stage surface and opposite to each other to minimize shadowing. Prior to imaging, the plants were grown in a light (300 μmol m⁻² s⁻¹) and dark cycle of 16 h and 8 h, respectively. The temperature of the plant growth chamber was kept at 25°C. The dimensions of the final imaged scene consisting of 36 plants were 3958 × 1800 × 186, where the first two dimensions are the spatial pixels while the third dimension is the spectral bands. The size of the image in the computer physical memory was ~10 Giga Bytes. For imaging, the camera was positioned in nadir, 1 m above the top of the plant surface. HySPEXRef software (Norsk Elektro Optikk, Oslo, Norway) was used to estimate the radiance and later the reflectance, based on a white reference panel imaged prior to each tray scan using HySPEXRef software (Norsk Elektro Optikk, Oslo, Norway).

At first the images were manually pre-processed to remove the non-relevant part from the image, i.e., the border of the image carrying non-relevant information was manually cropped, and two images were stitched together which led to a final image size of 3000 × 2796 × 186 for the 36 plants. The image was divided into calibration set (24 plants i.e. 66.66%) and independent test set (12 plants i.e. 33.33%) where all model training was performed on the calibration set (24 plants) and the final model testing was performed on the independent test set (12 plants). For model training and validation, ground truth segmentation masks were created by manually labelling all plant edges in the image scene with the ‘roipoly’ function in MATLAB (The MathWorks, Natick, MA, USA).

2.2. Image preparation for deep learning

Details on each of the three preparation stages (preprocessing, PCA and subsampling) are as follows:

2.2.1. Spectral pre-processing

Close-range spectral imaging of plants suffers from illumination effects where factors such as shadowing, scattering and mixtures of shadowing and scattering introduce additive and multiplicative effects into the spectra of imaged plants. In earlier works, spectral normalisations, such as standard normal variate (SNV) (Asaari et al., 2019; Asaari et al., 2018) and variable sorting for normalisation (VSN) (Mishra et al., 2020a), and 1st and 2nd spectral derivatives (Pandey et al., 2017), have been shown to effectively managed these effects and improve the data modelling. In this study, 8 different pre-processing techniques and their combinations were explored to evaluate their effect on the training and the performance of the DL models. The 8 techniques were 1st derivative (Savitzky and Golay, 1964), 2nd derivative (Savitzky and Golay, 1964), SNV (Barnes et al., 1989), SNV + 1st derivative, SNV + 2nd derivative, VSN (Rabatel et al., 2020), VSN + 1st derivative and VSN + 2nd derivative. The performance of the pre-processing techniques was benchmarked with respect to the modelling of the raw reflectance data. Hence, 9 models were explored i.e., 1 for the raw reflectance data and 8 for the different pre-processed data. The spectral pre-processing was performed pixel-wise by unfolding the image data array in MATLAB and using the MBA-GUI toolbox (Mishra et al., 2020a).

2.2.2. Principal components analysis for dimensionality reduction

After the pre-processing, the PCA was carried out for all 9 different cases separately. For PCA, a cropped image, carrying both the green plants and the background soil, of size 200 × 200 was extracted to reduce the computational load and PCA decomposition was performed in the spectral domain by first unfolding the cropped image. PCA analysis was performed in MATLAB using the ‘pcacov’ function in the ‘Statistics and Machine Learning’ toolbox. The optimal number of principal components (PCs) for each of the 9 cases was decided by the elbow point in the extracted variance plot. The loadings extracted from the PCA on the cropped image were used to transform the complete image by means of an inner multiplication. The final image after the PCA transform had a size of (1976 × 2796 × n) for calibration and (1024 × 2796 × n) for the independent test image, where n is the total number of PCs, which varied with the pre-processing.

2.2.3. Data generation for deep learning with subsampling

Several images are necessary for training the DL network, in order to learn the data patterns. To deal with this, small images were subsampled from the spectral image. This subsampling has two advantages, the first is that it generates a data set suitable for DL modelling and the second is that the smaller subsampled images are within the memory limits during the training of the DL model. In this study, the image subsampling was performed using the integral 'patch2d' function from the ‘feature extraction’ subclass of the sklearn toolkit (https://scikit-learn.org/). Due to memory limits (64 GB RAM), a patch size of 128 × 128 was selected. For model training, a total 5000 randomly sub-sampled images were extracted from the calibration image, while for model tuning, a total of 1000 randomly sub-sampled images were extracted. Since the random subsampling could lead to cases where the image has only a single class, filtering was performed to drop images containing a single class. No sub-sampling was performed for the independent test image as the model was directly tested on the complete image.

2.3. Semantic segmentation with U-net

Semantic segmentation is the task of assigning each pixel of an image to its correct class. U-net is an advanced semantic segmentation algorithm first proposed for the segmentation of biomedical images (Ronneberger et al., 2015). U-net is an end-to-end fully connected neural network (FCN) that has two main parts. The first is the encoding part where a series of convolutional and max pooling layers capture the context in the image scene. The second is the decoding part where a symmetrical expansion allows precise localisation using up-sampling and transposed convolutions. The main benefit of U-net is that being an FCN it allows handling of images of any size. This means a model trained on a small image can later be used for any other image size input. This property is of particular interest for this study where the model training is performed with a patch size of 128 × 128 while the final model test must be performed on the independent test data of size 1024 × 2796. A summary of the different layers of the U-net used in this study is shown in Fig. 1. The U-net was implemented using the Python (3.6) language and Keras/Tensorflow (2.1.0) running on a workstation equipped with a NVIDIA GPU (GeForce RTX 2080 Ti), an Intel® Core™ i7-4770 k @3.5 GHz and 64 GB RAM, running Microsoft Windows 10. The model weights were initialized with the ‘he normal’ initializer in Keras. Furthermore, an adaptive moment (ADAM) optimizer was used to minimize the categorical cross-entropy loss function which can be estimated as in Eq (1):

\[
loss = - \sum_{i=1}^{n_{c}} y_{i} \log \hat{y}_{i} \tag{1}
\]

where \( \hat{y}_{i} \) is the i-th scalar value in the model output, \( y_{i} \) is the corresponding target value, and output size is the number of scalar values in the model output.

A batch size of 8 was set to meet the memory limitations of the computer. Automatic learning rate adaption based on watching the
Intersection of union (IoU) score along the training process was implemented using the 'ReduceLROnPlateau' function from Keras, IoU being the area of overlap between the predicted segmentation and the ground truth. An early stop was implemented using the 'EarlyStopping' function from Keras, where the training process was automatically stopped if no further significant improvement was noted in the validation loss, during the training process. Each model was trained for 1000 epochs but due to the 'EarlyStopping' function, all models were ended within ~100 epochs. Checkpointer was used to save the best model weights during the training process.

Once the model was trained, the model weights were saved locally to test the model on the independent test set. The model was reinitialized with the input shape of the test set image i.e., $1024 \times 2796 \times n$ and the weights saved after model training. The input shape was changed as the model was trained with the $128 \times 128 \times n$ image. The model was then fitted on the independent test set image using the 'model.fit' function from Keras, which resulted in soft segmentation maps for which the 'argmax' was estimated to binarize the segmentation of plants from the background. The performance of the model on the independent test set image was reported in terms of IoU score, i.e., the area of overlap between the predicted segmentation and the ground truth. The IoU score ranges from 0 to 1, where 0 shows no overlap and poor segmentations while 1 means a perfect segmentation with respect to the ground truth. In this study, the IoU scores were used to compare the performance of different pre-processing methods applied prior to the DL modelling.

3. Results

3.1. Failure of NDVI-based threshold segmentation

A common approach to plant segmentation is based on the threshold of NDVI images. The NDVI values ranges from $-1$ to 1, where a higher value is related to the presence of healthy living green plant while negative and relatively low positive values are related to non-plant materials. Usually, a threshold is used to segment the green plant from the background (Asaari et al., 2019; Asaari et al., 2018). However, the main problem with the NDVI threshold is when the background of the plant contains objects that attain NDVI values similar to those of the plant. As can be seen in Fig. 2A, there are many pixels in the background, corresponding to algae or moss, that also have high NDVI values due to their spectral properties resembling those of the target plant. In Fig. 2B, the histogram of the pixels in Fig. 2A shows that there were two main distributions, corresponding to the plant and the background, but that the distributions strongly overlap making it difficult to find a suitable threshold. The segmentation performance based on varying the NDVI threshold is shown in Fig. 3. Based on the pixel distribution shown in Fig. 2B, a threshold value of 0.7 seems to be practical to segment the plant from the background. However, this threshold still results in many background pixels being identified as the plant (Fig. 3D). Increasing the threshold from 0.7 to 0.8 removed useful plant pixels while reducing the threshold from 0.7 to 0.6 increased the number of background pixels included in the segmentation. Such a challenge shows the need for more advanced methods that can learn from both the spatial context within the image as well as the spectral properties of the objects in the scene.

3.2. Principal components analysis for differently pre-processed data

Prior to DL, spectral pre-processing and PCA analysis are needed to remove the illumination effects and to reduce the dimensionality of the images so that they can be easily handled during DL model training. A summary of PCA analysis on the pre-processed images is show in Figs. 4, 5 and 6. In Fig. 5, the variance is plotted as a function of the number of PCs extracted. In the legend of Fig. 5, the number of PCs selected for each pre-processing are also shown. For raw data only 2 PCs were selected as no gain in modelled variance was seen after the 2nd PC. A reason for just 2 PCs explaining such a high part of the variance could be related to the additive and multiplicative effects present in the image due to the illumination effects. Usually in the presence of additive and multiplicative effects, the underlying variation is dominated by the global differences in spectral intensities which also make the PCA inefficient to extract the useful variations in the imaged scene.

To have an insight into the transformation performed by the PCA, a plot of the first two principal components is shown in Fig. 6. The blue dots represent the plant pixels, the red dots represent the background pixels, and the green dots represent the algae or moss pixels. As can be seen in Fig. 6, the plant pixels are clustered together while the background pixels are spread out. This suggests that the PCA is able to capture the main variations in the data, which are related to the spectral properties of the objects in the scene.
analysis on differently pre-processed images, the 1st PC images for the raw reflectance and the differently pre-processed reflectance images are shown in Fig. 5. For raw reflectance (Fig. 5A), it can be noted that similar scores were distributed evenly in both the plant and the background pixels. Hence, the 1st PC of raw reflectance did not efficiently separate the variations related to the plant and to the background. Furthermore, the same plant leaf has a wide variation in its scores. Such a wide variation of scores on the same leaf is indicative of the presence of...
illumination effects due to its local curvature (Mishra et al., 2020b). This is also clear from the loadings corresponding to the 1st PC (Fig. 6A), where only the global shape of the plant spectra was modeled as the main source of variation. Modelling of such a global shape shows the presence of a high variability due to additive and multiplicative effects. After 1st derivative pre-processing, the scores for the 1st PCs created a contrast between the plant and its background. This contrast is relevant as the corresponding loadings showed two major positive peaks at ~500 nm and 700 nm, related to plant photosynthetic pigments and the red-edge region (Fig. 6A). However, there were several leaf regions where the illumination effects were dominant: for example, due to one leaf being below another (Fig. 5B). The effect of the 2nd derivative was similar to that of the 1st derivative (Fig. 5C). Furthermore, the peaks in the loadings correspond to the leaf pigments and to plant photosynthetic activity (Fig. 6A). The SNV pre-processing showed higher negative scores for the plant and scores close to 0 for background objects (Fig. 5D). The higher negative scores on plants were related to the higher negative loading weights in the spectral range 750–1000 nm, which could be indicative of the presence of high moisture in the plants (Fig. 6B). The SNV followed by 1st derivative pre-processing brought a sharp contrast between the plant and the background pixels (Fig. 5E). This was possible as the 1st derivative on the SNV spectra extracted the key underlying peaks related to plant pigments and the red-edge region (Fig. 6B). However, the 1st PC of the SNV followed by the 2nd derivative pre-processing did not capture any distinct information in the image (Fig. 5F). The corresponding loading also reached zero weight in the complete spectral range except for high weights around ~ 450 nm. Please note that in Fig. 5F only the 1st PC image of SNV followed by 2nd derivative pre-processing is shown, the following PCs captured information related to the plant (results not presented). The SNV pre-processing also produced a contrast between the plant and the background (Fig. 6G). This contrast was related to the high loading weights in the spectral range >800 nm (Fig. 6C). The 1st and 2nd derivative pre-processing on the SNV pre-processed data showed enhanced contrast between the plant and the background (Fig. 5H, I). The reason for such a contrast is the extraction of the underlying peaks related to the plant pigments and the red-edge region characterizing the photosynthetic activity of the plant (Fig. 6C).

Both the normalization and derivative operations enhanced the contrast between the plant and the background pixels. Spectral normalization techniques such as SNV and VSN were found to be effective in reducing the illumination effects since scores in the same leaf were homogenized. The derivatives were less effective in terms of reducing the illumination effects as several regions still showed shadow effects.

### 3.3. Performance of U-net for differently pre-processed images

The next step after the PCA analysis was to train the U-net models. In total, 9 models were trained. It was hypothesized that an effective pre-processing approach would complement the U-net model during training as well as in testing. A summary of the training history of the U-net model for all 9 models is shown in Fig. 7. For each model, loss (blue) and validation loss (orange) are shown as a function of epochs. For raw reflectance (Fig. 7A), 1st (Fig. 7B) and 2nd (Fig. 7C) derivative pre-processed reflectance, the validation loss diverged from the loss after
the initial epoch. Such a divergence is indicative of model overfitting. For derivatives estimated after normalization (Fig. 7E, F, H, I), high overfitting was noted. However, for the normalization with SNV (Fig. 7D) and VSN (Fig. 7G), the model overfitting was lowest compared to raw reflectance or derivatives. Such a low difference between the loss and validation loss during the training shows a good model training (Fig. 7D, G). For both the SNV (Fig. 7D) and VSN (Fig. 7G) models, the training required the fewest epochs and ended in ~80 epochs.

After model training, the performances of all 9 DL models were tested on the independent test image and the IoU scores are shown in Table 1. The model calculated on the raw reflectance data performed the poorest with an IoU score of 0.84. The model calculated on the SNV and VSN pre-processed data performed the best with a similar IoU score of 0.96. The model for the derivative pre-processed data showed intermediate performance. Such high IoU scores reached with the model made using SNV and VSN normalized data proves the importance of chemometric pre-processing and particularly spectral normalization prior to DL modelling.

The model based on SNV and VSN pre-processed data reached a similar IoU = 0.96 on the independent test set image. To have a visual interpretation of the results, the segmentation mask predicted by the models along with the ground truth image and mask are shown in Fig. 8. Comparing the segmentation mask predicted by the SNV model (Fig. 8C) with the ground truth (Fig. 8B) mask, some misclassification related to the location of the checker box can be noted (marked by red circles in Fig. 8C). However, such misclassifications were absent from the segmentation mask predicted by the VSN model (Fig. 8D). Furthermore, a key thing to note is that the manually segmented ground truth mask (Fig. 8B) missed some key plant parts (yellow circles in Fig. 8C, D) which were also accurately predicted by the SNV and VSN models. In other words, the performance of the DL model was better than the manual segmentation of the ground truth image. Furthermore, although the SNV and VSN models reached a similar IoU = 0.93, the VSN model performed better as it miss-classified fewer pixels.

4. Conclusions

This study presents the first application of DL based semantic segmentation of spectral images of plants in combination with chemometrics pre-processing. Two chemometrics concepts, i.e., spectral pre-processing and PCA, were combined with a U-net based semantic segmentation model. The suggested approach gave better results than the methods commonly used in the field of spectral imaging of vegetation, in particular combining spectral normalization using SNV and VSN prior to DL modelling led to efficient training of the U-net segmentation model. Such efficiently trained models outperformed those built using the raw reflectance data by reducing non-relevant variability in the data due to illumination effects. Although both the SNV and VSN attained a similar metric of IoU = 0.96, based on the quality of the predicted segmentation map, the VSN pre-processing performed better. Furthermore, thanks to

![Fig. 7. A summary of model training and validation loss for differently pre-processed data. (A) Raw reflectance, (B) 1st derivative, (C) 2nd derivative, (D) SNV, (E) SNV + 1st derivative, (F) SNV + 2nd derivative, (G) VSN, (H) VSN + 1st derivative, and (I) VSN + 2nd derivative.](image-url)
Fig. 8. A summary of the segmentation performance on the independent test image. (A) 1st PC image of the test image, (B) manually labelled segmentation, (C) semantic segmentation with U-net based on SNV pre-processed data, and (D) semantic segmentation with U-net based on VSN pre-processed data.

the PCA, the DL models could be efficiently trained without an out-of-memory errors occurring. Although the work presented here deals with the segmentation of spectral images of plants, the method could be used for segmenting any type of spectral images.

CRediT authorship contribution statement

Puneet Mishra: Conceptualization, Methodology, Software, Writing - original draft, Data curation. Roy Sadeh: Conceptualization, Methodology, Software, Writing - original draft, Data curation. Ehud Bino: Conceptualization, Methodology, Software, Writing - original draft, Data curation. Gerrit Polder: Conceptualization, Methodology, Writing - review & editing. Martin Boer: Conceptualization, Methodology, Writing - review & editing. Douglas N. Rutledge: Conceptualization, Methodology, Writing - review & editing. Ittai Herrmann: Conceptualization, Methodology, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgements

This research was supported by the Wageningen University & Research knowledge base program ‘Data Driven & High Tech’. Authors like to acknowledge Dr. Idan Efroni from The Hebrew University of Jerusalem for supplying the plant seeds used in this study.

References


