LEVERAGING DEEP LEARNING FOR THE RECONSTRUCTION OF PLANT HYPERSPECTRAL DATA FROM RGB IMAGES

Serge Sarkis[1], **Ibrahim Issa**[23], **Dany Abou Jaoude**[13], **and Salma Talhouk**[4]
[1]Department of Mechanical Engineering
[2]Department of Electrical and Computer Engineering
[3]Center for Advanced Mathematical Sciences
[4]Department of Landscape Design and Ecosystem Management
American University of Beirut Beirut, Lebanon
{ss296,ii19,da107,ntsalma}@aub.edu.lb

ABSTRACT

Hyperspectral imaging is an important tool used in plant health assessment. It opens the door for early detection of plant stress which allows for timely intervention and thus improved conservation efforts. However, the high cost and complexity of hyperspectral cameras has limited their usage. To mitigate this issue, the problem of reconstructing plant hyperspectral data from RGB images is investigated. The proposed model reconstructs the visual and near-infrared range (400 - 1000 nm) while being trained solely on images of vegetation, in contrast with existing "generic" models. It is hypothesized that training a less complex model on a specific material will achieve good accuracy even with a relatively small training dataset. The HSCNN-D model is adopted with a simplified architecture. Despite being much smaller than the original model, it achieves comparable performance to state-of-the-art models on images of vegetation.

1 INTRODUCTION

Climate change and human expansion have continuously strained forest and agriculture lands over the years. The result is an increase in plant diseases and pest infestations Elad & Pertot (2014). Plant health monitoring is an essential task for forest management teams and agriculturalists to limit and possibly avoid a decrease in biodiversity and crop yield. Traditional plant health assessment methods are based on visual inspection by experts Steinman (2000) which is laborious and time consuming. It also suffers from grading bias, as multiple experts can provide different grading for certain health parameters Ferretti (1998).

To reduce this ambiguity and obtain more precise and accurate health assessment methods, studies have explored the usage of remote sensing tools for plant health assessment Lausch et al. (2016). In particular, hyperspectral data has proved to be highly useful for health assessment Kureel et al. (2022) and species identification Liu et al. (2022). This type of data can be obtained from satellite images (such as AVIRIS or sentinel-2) Ahlswede et al. (2022), or using hyperspectral cameras (on-ground or by UAV) Fraser & Congalton (2021). The former, while being freely available in some cases, lacks the required spatial resolution for a proper assessment Liu et al. (2022). Hyperspectral cameras provide the required data with good spatial and spectral resolution. They come, however, at a high cost and require complex data processing. This significantly limits their use. Multispectral cameras are cheaper but lack the spectral resolution required for an in-depth analysis of plant health.

Recently, researchers have been investigating the reconstruction of hyperspectral images from RGB images Zhang et al. (2022). Multiple models have been proposed and tested on common datasets such as those of the NTIRE 2018 Arad et al. (2018), NTIRE 2020 Arad et al. (2020), and NTIRE 2022 Arad et al. (2022) competitions. They focus on the reconstruction of hyperspectral data in the 400-700 nm range at 31 bands of resolution, neglecting the near-infrared (NIR) band which is crucial for plant health analysis. In particular, the HSCNN-D model proposed in Shi et al. (2018)

and used in the NTIRE 2018 competition has been used extensively for plant health Hamzah et al. (2022); Fu et al. (2022). Some studies have attempted to predict the NIR range from RGB images such as Shukla et al. (2022); Aslahishahri et al. (2021) but have been limited to 1-2 bands.

In this paper, we develop a model that reconstructs both the visual and near-infrared range (400 - 1000 nm) from RGB images. Using RGB images to obtain such valuable information greatly improves conservation programs and opens the door for citizen scientist to contribute to large scale surveying of forest and agricultural lands. Indeed, we envision developing a simple tool (e.g., a phone application) that can be used for plant health assessment. As such, this requires relatively small models so that they can run locally on mobile phones. Currently available datasets are limited in size and the models proposed are too complex to use for on-edge computations. However, considering that the reflectance of an object is material dependent, we hypothesise that training on a dataset of vegetation only should reduce the required model complexity, data required, and training/inference time. To that end, we use a simplified version the HSCNN-D model Arad et al. (2018) (i.e., reduced layers) and train it on a dataset of vegetation only to reconstruct the full scale VNIR range. We investigate the effect of the number of bands predicted and different data preprocessing techniques on prediction accuracy.

2 Method

2.1 Dataset

Our training is conducted on the Virginia Tech RGB-VNIR dataset Brown & Moser (2021). It consists of indoor/outdoor RGB images of everyday objects, cars, and vegetation, alongside the corresponding hyperspectral images. This open source dataset is similar to the datasets used in the NTIRE competitions Arad et al. (2018; 2020; 2022) with the advantage of covering the near-infrared region and having a vegetation specific section. The latter contains 62 images of 9 different species of trees and plants (each image has a 400 x 800 resolution). The hyperspectral images are generated using the HS-CL-V10E SPECIM camera. It has a range of 400-1000 nm with a 1.4nm resolution and a total of 420 bands. The dataset contains multiple images of the same object in different conditions (atmospheric: cloudy, sunny; distance: near, far; different exposure levels), allowing us to train a more robust model better suited for field applications.

Our comparisons with state-of-the-art models utilizes the NITRE competitions' datasets. The competition focuses only on the visible range: 400-700 nm, with a 10 nm resolution for a total of 31 bands. The datasets are significantly larger than the one we are using for training with the 2022 dataset containing around 1000 images Arad et al. (2022). As such, achieving a comparable accuracy with a smaller dataset is challenging.

2.2 Preprocessing

As in the NTIRE competition, our images are divided into 50 x 50 patches. We mainly focused on segmentation: since we would like to train our model for vegetation only, we segmented the images to exclude all other parts. The RGB images were first segmented using the Segment Anything model (SAM) developed by Meta Kirillov et al. (2023). This allowed us to isolate the plant or tree from the background. However, some parts of the background were still visible between leaves and branches. A second segmentation step was added based on index thresholds. Multiple indices were considered for segmentation, such as NDVI (normalized difference vegetation index) Rouse et al. (1974), ExG (excess green index) Woebbecke et al. (1995), and the mean intensity ratio of red, blue and green Kawashima & Nakatani (1998). Another method considered was converting all RGB images to HSV color spectrum and isolating the pixels in the green color region only. After testing, SAM followed by NDVI segmentation with a 0.5 threshold gave the best results and was used to prepare the RGB masks.

2.3 Model

The model used in this study is the HSCNN-D model that was proposed in Shi et al. (2018) and gave the best results in the NTIRE 2018 competition (shown in Figure 1).

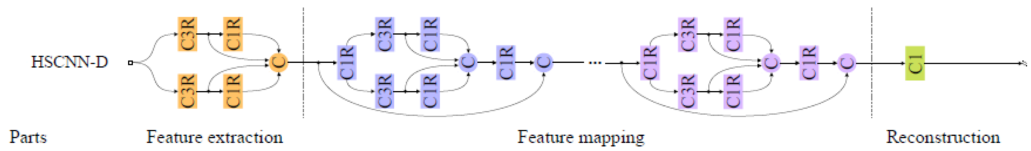

Figure 1: HSCNN-D model architecture (figure taken from Shi et al. (2018))

It is mainly composed of three blocks: a feature extraction block, a feature mapping block, and a reconstruction block. A schematic of the model used can be found in in Figure.2 in the supplementary material.Each block contains convolutional layers with 16 or 8 filters, and a ReLU activation function.

The complexity of the architecture is mainly controlled by the feature mapping block. This block adds 16 channels to the input tensor and is repeated b times, in series, across the pipeline, with the default value being 60. The feature extraction and feature mapping pipeline is executed 3 times (with different initialization), then the results are concatenated and passed to the reconstruction layer. We introduced some slight modifications to prevent overfitting. Dropout layers were added in the feature mapping blocks after each convolutional layer with a 0.25 dropout probability. Also, to reduce computational cost, the value of b (number of feature mapping blocks) was varied between 5 and 10 as opposed to $b = 60$ used in the original model.

2.4 Experiments

Most studies in literature imitate the NTIRE competition format by predicting 31 bands in the 400-700 nm range. Other reconstructed the NIR at a multispectral level only Aslahishahri et al. (2021); Fu et al. (2022); Shukla et al. (2022). Based on this, several experiments were conducted to investigate the following: 1) predicting 31 bands across the VNIR range, with and without segmentation; 2) predicting 42 bands across the VNIR range, with and without segmentation; and 3) predicting 61 bands across the VNIR range (equivalent to the same spectral resolution as the NTIRE competitions). Five-folds cross validation was used during training. Different values of L_2 and L_1 regularization were also investigated. The loss function used is the MRAE (mean relative absolute error) as in the NTIRE competitions in addition to the RMSE (root mean square error). The loss function was adjusted in order to account only for the losses of the region of interest (i.e., ignoring the masked areas). The models were trained on a high performance computing cluster (HPC), with 128 GB RAM allocated and one NVIDIA V100 tensor core GPU.

3 Results and Discussion

The MRAE and RMSE validation results for multiple trials are shown in Table 1. All results were obtained using L_2 regularization with an alpha value equal to 0.01, and training for 500 epochs.

n	b	Segmentation	MRAE	RMSE
31	8	None	0.4591	0.2537
31	8	SAM	0.3150	0.1194
42	10	SAM	0.297	0.1267
42	10	SAM + NDVI	0.268	0.105
61	10	SAM + HSV	0.49	0.165

Table 1: Experimental results for different trained models. n is the number of reconstructed bands. b is the number of duplicated feature mapping blocks

The effect of segmentation can be clearly seen in the results in Table 1. For $n = 31$ bands and $b = 8$ duplicate feature mapping blocks, the MRAE drops significantly (from 0.46 to 0.32) after applying SAM segmentation. Even refining the segmentation from SAM to SAM+NDVI improves performance as can be seen from the third and fourth row of the table.

3.1 Error analysis

Our best result is highlighted in red, wherein we trained a model with $b = 10$ duplicate feature mapping blocks to learn $n = 42$ bands. Henceforth , we will refer to this model as *HSCNN-Veg*. It achieves an MRAE smaller than 0.27. To put this number into context, an MRAE of 0.27 would rank among the top 10 performing models in the NTIRE 2022 competition Cai et al. (2022a). In other words, for the hyperspectral image reconstruction problem, this is a reasonably small error.

To better understand the behavior of the model, we computed average errors for the visible and near-infrared range separately. The average MRAE in the visible range (400-700nm) is 0.232, while it is 0.314 in the near-infrared range (700-1000 nm) and 0.229 in the 400-800 nm range. As such, we notice that a large contribution of the error is due to the 800-1000 nm range. Some bands had a more sizeable contribution to this error, especially at 400 nm, 1000 nm, and 987 nm. This might be caused by the camera used. Recomputing the error without these three wavelengths resulted in an average MRAE of 0.195,0.202, and 0.283 for the 400-700 range, 400-800nm range, and 700-1000 nm range respectively.

To get a better understanding of the effects of different image properties (such as weather, distance from object, plant species) on our task, we conducted an analysis of variance (ANOVA) on each wavelength with respect to these groups.

Results revealed that the mean reflectance in an image is affected by the species of the tree and/or plant for all 420 wavelengths. Weather conditions showed to affect all wavelengths except in the 500-600 nm range where they achieved a p-value> 0.1. The distance from the object was found to be statistically significant when studying the variation in the 532-565 nm and 700-929 nm. Note that the findings regarding the distance are less reliable due to the change of camera angle with the change of distance (which may affect the measured reflectance) .

3.2 Comparison with orginal HSCNN-D

Finally, we compare the performance of our HSCNN-Veg model with the performance of the HSCNN-D model trained on the NTIRE 2022 dataset (weights obtained from Cai et al. (2022b)). To that end, we considered 14 images of trees and plants from the NTIRE 2022 dataset. Note that this comparison is done in the visible region only, limited by the available data from the NTIRE 2022 competition. Results are shown in Table 2. Remarkably, our model produces a smaller MRAE

	HSCNN-Veg	HSCNN-D
Average MRAE	1.5238	2.9263
Median MRAE	0.6874	1.4949

Table 2: Comparison between HSCNN-D 2018 model and our HSCNN-Veg model using 14 plant and tree images from the NTIRE 2022 competition.

on average (and in fact for most cases) despite being trained on a much smaller dataset and having smaller size overall. However, both models perform very badly on some samples (hence the large difference between the average and the median values).

A main advantage of the HSCNN-Veg model is its practicality due to its relatively small size. In particular, it uses around 683,000 parameters (i.e., around 5.2 MB to store model parameters) and requires less than 900 MB of RAM when evaluating an image. Therefore, such a model is small enough to run locally on mobile phones. By contrast, the HSCNN-D model has over 9,200,000 parameters (i.e., more than 70 MB to store model parameters) and requires a minimum of 64 GB of RAM required for inference ngchc (2019).

The testing results underline the need for more data in order to improve the generalizability of the models for real world application. A similar conclusion was reached in the NTIRE 2022 competition Arad et al. (2022). Nonetheless, the model proposed in this study was able to outperform the original model while having a simpler architecture and being trained on a smaller dataset. This indicates that training on a specific type of material and understanding the varying factor that affects model performance significantly improves the learning process.

4 Conclusion

In this study, the reconstruction of the reflectance spectrum for vegetation in the range of 400-1000 nm was investigated. The HSCNN-D model with reduced number of layers was trained on a dataset containing 62 images of plants and trees. It achieved an average MRAE of 0.27 which is comparable to state-of-the-art models showcased in the NTIRE 2022 competition. The modified model also outperformed the original HSCNN-D on the NTIRE 2022 dataset, while having fewer parameters and requiring significantly less memory for inference.

Achieving such results on a much smaller dataset and architecture underlines the efficiency of training a model for specific applications and on targeted material types. The demonstrated practicality of the model enables its usage on smartphones and facilitates field applications. Future works will concentrate on building a larger hyperspectral dataset for vegetation only and investigating the applicability of this method for early stress detection in plants and trees.

References

Steve Ahlswede, Christian Schulz, Christiano Gava, Patrick Helber, Benjamin Bischke, Michael Förster, Florencia Arias, Jörn Hees, Begüm Demir, and Birgit Kleinschmit. Treesatai benchmark archive: A multi-sensor, multi-label dataset for tree species classification in remote sensing. *Earth System Science Data Discussions*, 2022:1–22, 2022.

Boaz Arad, Ohad Ben-Shahar, Radu Timofte, Luc Van Gool, Lei Zhang, Ming-Hsuan Yang, Zhiwei Xiong, Chang Chen, Zhan Shi, Dong Liu, Feng Wu, Charis Lanaras, Silvano Galliani, Konrad Schindler, Tarek Stiebel, Simon Koppers, Philipp Seltsam, Ruofan Zhou, Majed El Helou, Fayez Lahoud, Marjan Shahpaski, Ke Zheng, Lianru Gao, Bing Zhang, Ximin Cui, Haoyang Yu, Yigit Baran Can, Aitor Alvarez-Gila, Joost van de Weijer, Estibaliz Garrote, Adrian Galdran, Manoj Sharma, Sriharsha Koundinya, Avinash Upadhyay, Raunak Manekar, Rudrabha Mukhopadhyay, Himanshu Sharma, Santanu Chaudhury, Koushik Nagasubramanian, Sambuddha Ghosal, Asheesh K. Singh, Arti Singh, Baskar Ganapathysubramanian, and Soumik Sarkar. Ntire 2018 challenge on spectral reconstruction from rgb images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1042–104209, 2018. doi: 10.1109/CVPRW.2018.00138.

Boaz Arad, Radu Timofte, Ohad Ben-Shahar, Yi-Tun Lin, and Graham D Finlayson. Ntire 2020 challenge on spectral reconstruction from an rgb image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 446–447, 2020.

Boaz Arad, Radu Timofte, Rony Yahel, Nimrod Morag, Amir Bernat, Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, et al. Ntire 2022 spectral recovery challenge and data set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 863–881, 2022.

Masoomeh Aslahishahri, Kevin G Stanley, Hema Duddu, Steve Shirtliffe, Sally Vail, Kirstin Bett, Curtis Pozniak, and Ian Stavness. 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, 2021.

Ryan Brown and Josh Moser. Labeled Hyperspectral and RGB Images of Several Tree Species. 10 2021. doi: 10.7294/14113193.v2. URL `https://data.lib.vt.edu/articles/dataset/Labeled_Hyperspectral_and_RGB_Images_of_Several_Tree_Species/14113193`.

Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction, 2022a.

Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *CVPRW*, 2022b.

Yigal Elad and Ilaria Pertot. Climate change impacts on plant pathogens and plant diseases. *Journal of Crop Improvement*, 28(1):99–139, 2014.

M Ferretti. Potential and limitation of visual indices of tree condition. *Chemosphere*, 36(4-5): 1031–1036, 1998.

Benjamin T Fraser and Russell G Congalton. Monitoring fine-scale forest health using unmanned aerial systems (uas) multispectral models. *Remote Sensing*, 13(23):4873, 2021.

Jun Fu, Jindai Liu, Rongqiang Zhao, Zhi Chen, Yongliang Qiao, and Dan Li. Maize disease detection based on spectral recovery from rgb images. *Frontiers in Plant Science*, 13:1056842, 2022.

Raseeda Hamzah, Khyrina Airin Fariza Abu Samah, Muhammad Faiz Abdullah, and Sharifalillah Nordin. Investigation of rgb to hsi conversion methods for early plant disease detection using hierarchical synthesis convolutional neural networks. *JOIV: International Journal on Informatics Visualization*, 6(1):1–5, 2022.

Shigeto Kawashima and Makoto Nakatani. An algorithm for estimating chlorophyll content in leaves using a video camera. *Annals of Botany*, 81:49–54, 1998. URL `https://api.semanticscholar.org/CorpusID:85890980`.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

Neha Kureel, Jyoti Sarup, Shafique Matin, Suresh Goswami, and Kapil Kureel. Modelling vegetation health and stress using hypersepctral remote sensing data. *Modeling Earth Systems and Environment*, pp. 1–16, 2022.

Angela Lausch, Stefan Erasmi, Douglas J King, Paul Magdon, and Marco Heurich. Understanding forest health with remote sensing-part i—a review of spectral traits, processes and remote-sensing characteristics. *Remote Sensing*, 8(12):1029, 2016.

Keng-Hao Liu, Meng-Hsien Yang, Sheng-Ting Huang, and Chinsu Lin. Plant species classification based on hyperspectral imaging via a lightweight convolutional neural network model. *Frontiers in Plant Science*, 13:855660, 2022.

ngchc. HSCNN-Plus. `https://github.com/ngchc/HSCNN-Plus`, 2019.

John Wilson Rouse, Rüdiger H Haas, John A Schell, Donald W Deering, et al. Monitoring vegetation systems in the great plains with erts. *NASA Spec. Publ*, 351(1):309, 1974.

Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1052–10528, 2018. doi: 10.1109/CVPRW.2018.00139.

Ankit Shukla, Avinash Upadhyay, Manoj Sharma, Viswanathan Chinnusamy, and Sudhir Kumar. High-resolution nir prediction from rgb images: Application to plant phenotyping. In *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 4058–4062. IEEE, 2022.

Jim Steinman. Tracking the health of trees over time on forest health monitoring plots. *UNITED STATES DEPARTMENT OF AGRICULTURE FOREST SERVICE GENERAL TECHNICAL REPORT NC*, pp. 334–339, 2000.

David M Woebbecke, George E Meyer, Kenneth Von Bargen, and David A Mortensen. Color indices for weed identification under various soil, residue, and lighting conditions. *Transactions of the ASAE*, 38(1):259–269, 1995.

Jingang Zhang, Runmu Su, Qiang Fu, Wenqi Ren, Felix Heide, and Yunfeng Nie. A survey on computational spectral reconstruction methods from rgb to hyperspectral imaging. *Scientific reports*, 12(1):11905, 2022.