Using machine vision to build field maps of forage quality and the need for agriculture-specific machine vision networks

Paul W. Nugent[1], Jasmine Neupane[2]

1. Montana State University, paul.nugent@montana.edu
2. University of Missouri-Columbia, jasmine.neupane@missouri.edu

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

Machine vision systems have truly come of age over the past decade. These networks are relatively simple to implement with systems such as YOLOv5 or the more recent YOLOv8. They are also relatively easy and computationally cheap to retrain to a custom data set, allowing for customization of these networks to new object detection and classification tasks. This has resulted in an explosion of these networks and their application through all aspects of agriculture.

To prove the capabilities and the need for multi-spectral classification, we present a technique where a GPS and a MicaSense red-edge multi-spectral camera are mounted inside the cab of a swather. A machine vision network is then transfer-trained to build a system for image classification, allowing us to construct maps of weed species density in alfalfa or grass hay fields. Two different networks were studied: a classification-only network based on an implementation of Xception and a detection and classification network based on YOLOv5.These networks are fundamentally designed to work with standard RGB imagery. However, our imagery relies on five-channel data. To accommodate the higher number of color channels, we explored multiple techniques for fitting our data into the existing networks. The simplest selected three of our five channels to produce three-channel data, resulting in multiple networks. We also used parallel networks of different three-channel combinations with a concatenation layer before the retrained classification layer. We also explored pre-processing in front of the network to reduce our five channels to three using principal component analysis. Through each of these iterations, some networks were highly capable of detecting particular weed species within the green crop, such as a combination of Red, Red-Edge, and NIR had a MAP of 0.95 for the detection of Canada thistle, outperforming other channel combinations. In contrast, other channel combinations provided superior detection for other species or crop conditions. The result of these different approaches has exposed the need for a more robust five (or greater) network specifically designed to work with agricultural data.

Keywords.

Machine learning, object detection, weeds, weed science, multi-spectral imagery

Introduction

Machine vision systems have significantly advanced over the past decade. These networks, such as YOLOv5 (You Only Look Once) (Jiang et al., 2021) or the more recent YOLOv8 (Vijayakumar & Vairavasundaram, 2024), are relatively easy to implement. They can also be retrained with custom data sets at a low computational cost, enabling customization for new object detection and classification tasks. This has led to an explosion in the applications of these networks to tasks through all aspects of agriculture (Badgujar et al., 2024).

While capable of operating with more than three color images, most pre-trained versions of these networks only use an input layer that accepts images with three color channels. Most of the work we see in the literature uses these three-channel pre-trained models as a starting point (Badgujar et al., 2024). However, this limitation may significantly limit the performance of image classification and object detection in agriculture. Through this work, we have come up against these limitations, as they are fundamentally designed to work with standard RGB imagery, akin to what digital cameras and smartphones capture.

Traditionally, AI systems in various fields, including agriculture, have relied heavily on standard RGB datasets. This imagery forms the backbone of numerous AI models due to the vast availability of RGB data, making it easier to build training datasets, and these RGB data are the starting point for training most pre-trained networks (Deng et al., n.d.; Lin et al., 2014). However, in the agricultural sector, imagery needs to extend beyond the capabilities of standard RGB data. Much of agricultural imagery, if not the majority, relies on more channels. In agriculture, it's not just about creating visually interpretable images by humans. There's a need to capture more than what meets the eye. A plant and its parts can exhibit unique characteristics in different light spectrums. For example, chlorophyll absorption occurs in both the red and blue spectrums, while a higher reflectance produces green in vegetation. These RGB channels miss the high reflectance in the NIR and the transition in the red edge from high absorption to high reflectance.

This necessitates additional color channels, such as near-infrared (NIR) and red-edge, extending into shortwave infrared (SWIR) or thermal imaging. These expanded spectrums can provide four to seven color channels or even greater, offering a more comprehensive view of agricultural landscapes. The imagery of our example relies on five-channel data, RGB, with the addition of red-edge and near-infrared (NIR) channels. These channels are useful in agriculture because of the relationship between plant health, NIR, and red-edge reflectance. The most common imaging systems in agriculture use five color channels covering the typical red, green, and blue (RGB) imaging and have a near-infrared (NIR, N) and/or red edge (E). These channels are used because of the particulars of plant reflectance and absorption with high absorption from photosynthesis in the red and blue. Using RGB-only imagery has shown promise in agriculture; however, using only RGB or just three of these five (or more) channels is throwing away information. Some studies have used information from multispectral images in the form of an NDVI or similar masks to supplement the three-channel detection rather than using all five channels directly (Osorio et al., 2020).

Current methods of expanding current networks

With the prominence of multi-spectral data in agriculture, integrating multispectral data into artificial intelligence algorithms is a crucially needed development in modern agriculture. The challenge arises in adapting AI models, particularly convolutional neural networks (CNNs), to handle these multispectral inputs. Standard CNNs, like YOLO (Jiang et al., 2021; Vijayakumar & Vairavasundaram, 2024), RCNN (Region-based CNN) (Kumar et al., n.d.), Inception (Szegedy et al., n.d.)or Xception (Chollet, n.d.) and ResNet (Visin et al., 2015), are predominantly designed for three-channel (RGB) inputs.

In this work, we explore and demonstrate the limitation of three methods to accommodate the higher number of color channels. We explored multiple techniques for fitting our data into the existing networks, showed where each works well, and demonstrated the need for true multi-spectral networks.

1. The simplest selected three of our five channels to produce three-channel data. These networks tended to perform best for particular weeds, resulting in multiple networks.
2. The second approach was to use parallel networks of different three-channel combinations with a concatenation layer before the retrained classification layer. This is similar to a Siamese network approach.
3. The third method was the use of a pre-processing network in front of the network to reduce our five channels to three through methods such as principal component analysis (PCA). Here, we are exploring both PCA reduction and trained network reduction techniques, but only the PCA reduction is robust enough to present in this publication.

Using these methods to adapt existing networks to multispectral data has proven effective to an extent; each method inherently discards valuable data from unused channels or cross-channel analysis, which is not possible if all channels are not present in the data. The underutilization of the full spectrum of data available in multispectral imagery is present in the above techniques. When AI models are constrained to only a subset of available channels, their ability to distinguish between objects, such as different weed species from crops, can be significantly limited.

Thus, these different approaches have exposed the need for a more robust five (or greater) network specifically designed to work with agricultural data. Through each of these iterations, some networks were highly capable of detecting particular weeds or other problems within the green crop, but they underperformed others. For example, a combination of Red, Red-Edge, and NIR had a MAP of 0.95 for detecting Canada thistle, outperforming other channel combinations. In contrast, other channel combinations provided superior detection for other species or crop conditions. There's an ongoing need to develop AI algorithms that can fully exploit the richness of multispectral data, thereby enhancing decision-making and accuracy in agricultural practices. This advancement would represent a significant leap in precision agriculture, enabling more nuanced and effective farming strategies. To enable this development, a robust multi-channel labeled dataset is first needed.

Example Application

This study investigates the need for multi-spectral CNN image classification through a specific and potentially unique application, green-on-green weed detection forage quality assessment. The ability to detect weeds in green-on-green systems such as corn or soybean crops or alfalfa and grass forage for site-specific management of weeds has seen a surge of interest in recent years. These studies have focused on weed detection for herbicide or other management methods using remote sensing from Uncrewed Aerial Vehicles (UAVs), spray nozzle control systems and satellite platforms. However, the use of green-on-green weed detection and mapping of weeds past the point of management, and rather to understand forage quality, is relatively novel. To achieve this, the spatial variability of weed density across a field is detected and mapped, potentially relating this back to the contents of individual bales. To enable these applications, agriculture-specific machine learning networks need to be developed that don't require the utilization of only a portion of the collected data.

We placed a MicaSense red-edge multi-spectral camera with built-in GPS into various farm vehicles such as swathers, trucks, and combine harvesters to achieve our mapping system. In the data we present here, the MicaSense camera and GPS were mounted inside the cab of a swather, and alfalfa and grass hay were cut in preparation for baling. Post-cutting, a machine vision network was transfer-trained to build a system for image classification. This allowed us to construct maps of weed species density across the field, including alfalfa and grass hay fields. These maps are produced using image classification convolutional neural networks (CNN) to identify and classify the weeds within the images.

Two different networks were studied: a classification-only network based on an implementation of Xception (Chollet, n.d.) and a detection and classification network based on YOLOv5. We adapted the Xception network to use multispectral data through approaches 1 and 2 mentioned earlier. We also adapt YOLOv5 and the more advanced YOLOv8 to perform detection and classification tasks on multi-spectral data using methods 1 and 3. These networks were selected because they are straightforward to implement and can be efficiently retrained on custom datasets. This allowed for their adaptation to our various object detection and classification tasks.

Through these various methodologies, we have identified the pressing need for a more specialized network that can efficiently handle five or more channels, particularly for agricultural data. This need is underscored by our findings, where certain channel combinations, such as Red, Red-Edge, and NIR, achieved high accuracy in detecting specific weed species within crops. Meanwhile, other combinations proved more effective for different species or crop conditions.

Table 1. Color Channel Names and Abbreviations. For the sake of clarity, the following abbreviations are used for each color channel, and example combinations are shown.

Channel	**Abbreviation**
Red	R
Green	G
Blue	B
RedEdge	E
Near Infrared (NIR)	N
Example Combinations	**Abbreviation**
Red + Green + Blue	RGB
Red + RedEdge + NIR	REN
Blue + Green + NIR	BGR

Building an initial training set

The initial labeled dataset was based on pre-trained Xception for feature extraction, UMAP for dimensional reduction (McInnes et al., 2018), and hdbscan for clustering (McInnes et al., 2017). Multispectral analysis was conducted through a concatenated network approach where the top layer was removed from 2 or more Xception networks, and different color combinations were inserted into each network. The output of each Xception layer was an average pooling layer, resulting in a 1x2048 vector. These were then concatenated together to produce a 1x4096 vector or greater, depending on the number of color channels. To accommodate our higher-resolution imagery and the 299x299 input, each image was broken into a series of 299x299 tiles, and each tile was run through the network independently.

To build a training set, this concatenated network was used to run a large number of tile images randomly selected across our imagery, and the 1x4096 feature vector was calculated and stored for each tile image. These 1x4096 vectors for all images were then run through UMAP and reduced to a 2-dimensional projection of each tile image's features. Within this 2-dimensional dataset, clusters represented tile images with similar features. To identify these features hdbscan used cluster the 2-D data. This produced datasets of similar images, these sets were observed by human experts, cleaned of incorrect images, and provided appropriate labels. Some image sets were then iteratively reprocessed to produce further refined image sets. Figure 1 shows an example of this process, with original clusters detected, then cluster 1 from the left side was further processed.

During this process, different color channel combinations were explored, involving combinations of RGB, RGE, RGN, RBN, REN, GBE, GBN, GEN, and BEN color groups. For this network architecture, it was that a concatenation of RGB + REN color groups produced the best results, and then resulting labeled tile images were used to build and train a full classifier.

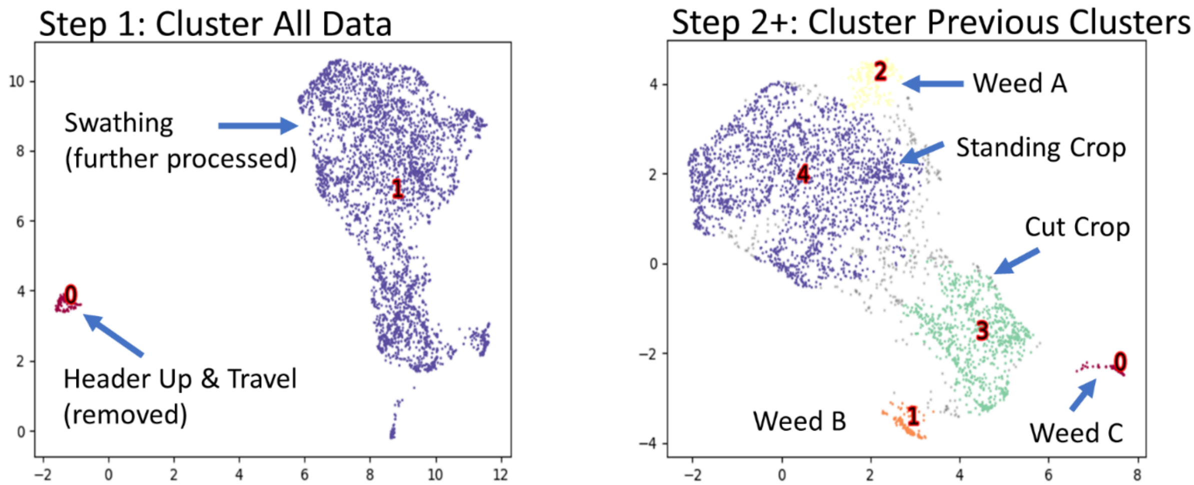

Figure 1. Clustering is produced by UMAP and hdbscan, as well as the manually assigned cluster labels.

Mapping with a Classification Network

To demonstrate the capabilities of green-on-green mapping of specific species, an RGB + REN Xception-based CNN was used. We calculated an object presence fraction based on the presence of specific objects within the total number of tiles. For example, in the images of Figure 2 and Figure 3, we identify the presence of Creeping Thistle and Sainfoin. A value of 50% means that 50% of the processed tiles detected the class or 50% of the area contained the class. To ensure accurate classification, the camera was positioned to capture the field exclusively. Some images featured the header of the swather, which meant it was raised, and the crop was not being swathed. Thus, these images were disregarded during the mapping process.

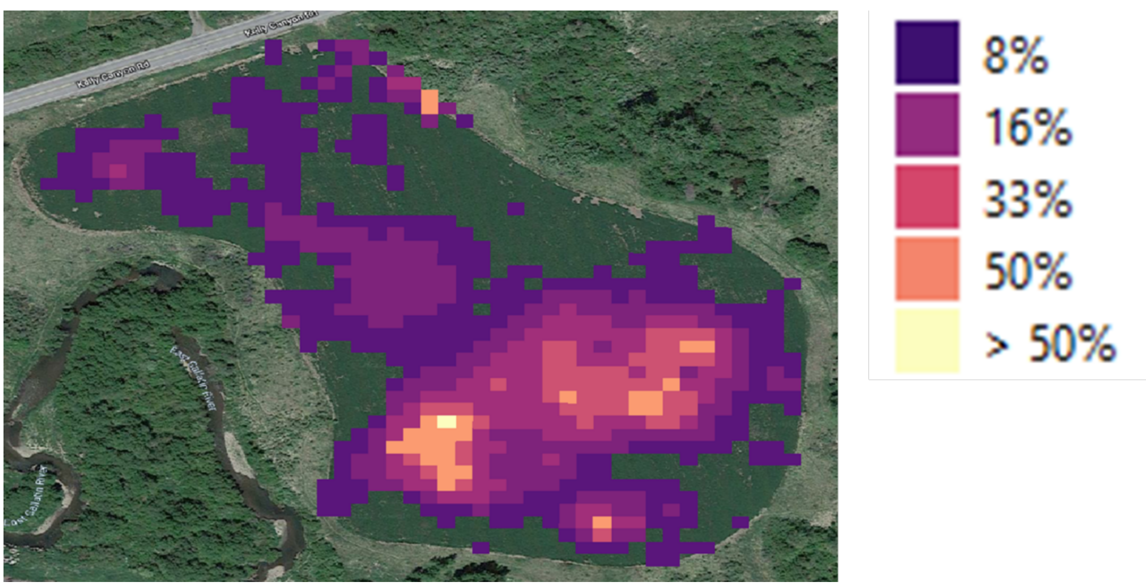

Figure 2. Sainfoin density across the study field in 2022.

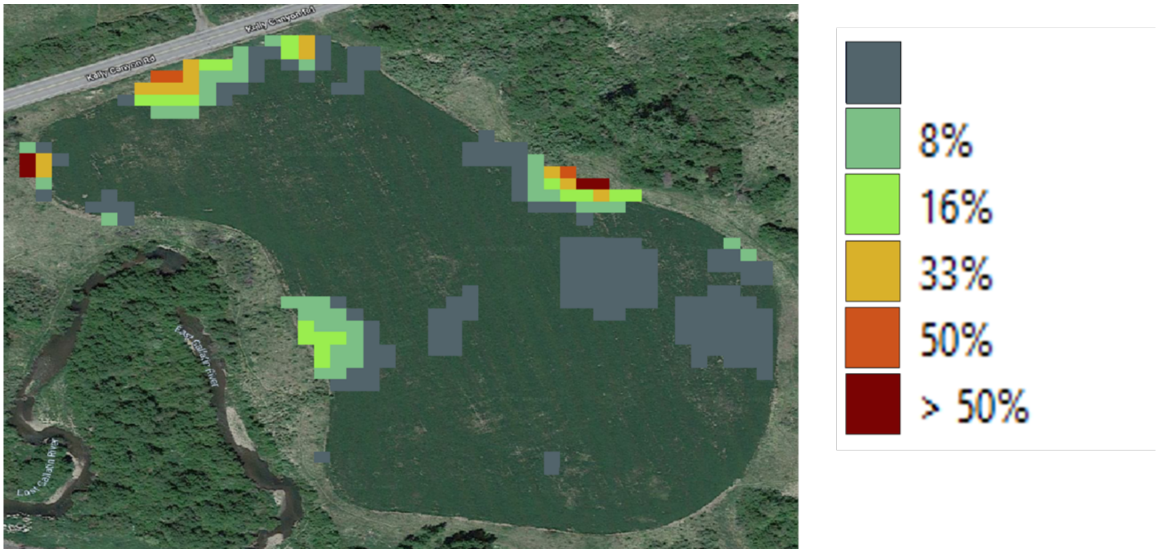

Figure 3. Creeping Thistle density across the study field in 2022.

Processing with a Detection and Classification Network

To study the importance of specific spectral channels with a detection and classification network, a YOLOv5 network was transfer-trained for various color combinations. This included a principle component analysis (PCA) projection of the five-channel to a three-channel dataset using a random selection of imagery across all object classes. Training each network started using a 3-channel input YOLOv5 and the COCO weights. Then, each combination was trained independently using the same set of object labels.

The object labels were built using the images containing objects of interest identified using Xception-based labeling. These identified images were converted to fully segmented images using "segment anything." The bounding boxes for these segments were then used to train a series of YOLOv5 networks for different channel combinations. Table 2 provided the MAP that was experienced for the training of each of the color channel combinations. Some combinations provided superior performance for particular weeds such as REN for Creeping Thistle. At the same time, the PCA projection produced broadly good results but had a poor performance for particular weeds such as Creeping Thistle. This might be because........

Table 2. YOLOv5 classification results for specific object classes (often weed species) and different combinations of 3 channels and a principle components projection to select 3 channels.

	REN	**BGN**	**PCA**
Cut Alfalfa	0.91	0.87	**0.98**
Oxeye Daisy	0.56	0.74	**0.83**
Wild Mustard	**0.76**	0.62	0.63
Pennycress	0.84	0.56	**0.90**
Sainfoin (volunteer)	0.72	0.75	0.71
Creeping (Canada) Thistle	**0.95**	0.90	0.81

Work toward a YOLOv8 detection and segmentation model is underway. This work will include a June 2024 alfalfa harvest leading to three years or repeat measurements on the same fields, allowing us to observe changes in weed species throughout the field and the potential for changes in overall forage quality. Since the 2024 harvest has not yet occurred at the time of writing this paper, initial results will be presented during the oral presentation at the ICPA 16 conference, with an in-depth analysis in a subsequent refereed journal publication.

Toward a multi-channel data set

Building upon the challenges of integrating multispectral data into AI models, the first step is the creation of a comprehensive labeled image database. The size of these networks matters when we start to determine the number of images required to build a true multi-spectral image dataset. In comparison, models like Xception and Inception have 22.8 million and 23.6 million parameters, respectively. YOLOv8, the successor to YOLOv5, also shows a range in parameter count, with YOLOv8s having about 11.2 million and YOLOv8x around 68.2 million parameters. The relationship between the model size and its performance is significant. Larger models like YOLOv5x possess more parameters, which generally translates to better performance but at the cost of increased computational demands, slower processing times, and larger training data sets. Scaling to multi-spectral networks should create a scaling proportional to the number of channels. To ensure robust training and accuracy, a general guideline is that the number of unique training examples in the dataset should be substantially larger than the number of variables in the network. Some metrics suggest a ratio of at least 10:1, meaning that for every variable in the network, there should be ten times as many training examples.

These models' apparent large number of parameters is somewhat relaxed as the inputs are images and can be viewed as many individual inputs per image. This means that a relatively lower number of images, such as 1000 images per class, can be sufficient to achieve a balanced dataset. However, these small datasets must be carefully designed to achieve a generalized network. Therefore, given the complexity and the high number of variables in networks, the required dataset size may need millions of input images to achieve a level of training that ensures reliability and accuracy. This requirement poses a significant challenge, especially when dealing with multispectral data, which inherently has more dimensions than standard RGB data.

Conclusions

To effectively use multispectral data in AI models for agriculture, we need comprehensive labeled image databases and advanced neural networks that can handle multiple channels. While existing models like YOLOv5 are robust, they are mainly designed for standard RGB inputs. Our research into adapting these networks for multispectral data shows both the potential and limitations of current methods. Although techniques such as selecting specific channels or using parallel networks offer some solutions, they often lead to the underutilization of valuable data. This highlights the urgent need for specialized networks that can fully leverage the richness of multispectral imagery.

The example application of detecting weed species density in alfalfa and grass hay fields using a MicaSense red-edge multispectral camera mounted on farm vehicles demonstrates the practical implications of these advancements. By adopting machine vision networks to handle five-channel data, we created detailed maps illustrating weed species density, which is crucial for precision agriculture. This case study showcases the potential of integrating multispectral data into AI models and highlights the need for specialized networks to effectively process this data. As we continue to develop and refine these networks, the goal is to enable more nuanced and effective farming strategies, ultimately contributing to the broader objectives of sustainable and efficient agricultural practices.

At Montana State University, we are creating a multi-spectral dataset containing labeled and validated multi-spectral data. Since our initial experiments in 2022, we have installed 2 RedEdge cameras in various vehicles across the MSU farms, including swathers, tractors, combine harvesters, and trucks. Each image is given a unique name based on data, time, camera serial number, and original filename. This naming convention ensures that each file has a distinct and non-repeating identifier, allowing for the easy placement of identifying information, labels, and detections in a database. MicaSense data has the channel number specification in their filenames, which is preserved in the original filename portion of the new name. Looking ahead, we aim to establish a GitHub repository for labeling, camera designs, and assistance software to encourage collaboration with other groups worldwide. The dataset used to produce the example maps in the publication comprised 9554 unique 5-channel images. Subsequently, we have added nearly 300,000 unique multispectral images. We anticipate that these processes will continue to evolve, and we are enthusiastic about the prospect of working with other organizations to further this endeavor.

References

Badgujar, C. M., Poulose, A., & Gan, H. (2024). Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, *223*, 109090. https://doi.org/10.1016/j.compag.2024.109090

Chollet, F. (n.d.). *Xception: Deep Learning with Depthwise Separable Convolutions*.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (n.d.). *ImageNet: A Large-Scale Hierarchical Image Database*. http://www.image-net.org.

Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2021). A Review of Yolo Algorithm Developments. *Procedia Computer Science*, *199*, 1066–1073. https://doi.org/10.1016/j.procs.2022.01.135

Kumar, A., Janmenjoy, D., Bighnaraj, N., Soumen, N., Pati, K., & Pelusi, D. (n.d.). *Advances in Intelligent Systems and Computing 999 Computational Intelligence in Pattern Recognition*. http://www.springer.com/series/11156

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). *LNCS 8693 - Microsoft COCO: Common Objects in Context*.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. http://arxiv.org/abs/1802.03426

Osorio, K., Puerto, A., Pedraza, C., Jamaica, D., & Rodríguez, L. (2020). A Deep Learning Approach for Weed Detection in Lettuce Crops Using Multispectral Images. *AgriEngineering 2020, Vol. 2, Pages 471-488*, *2*(3), 471–488. https://doi.org/10.3390/AGRIENGINEERING2030032

Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (n.d.). *Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning*. www.aaai.org

Vijayakumar, A., & Vairavasundaram, S. (2024). YOLO-based Object Detection Models: A Review and its Applications. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-18872-y

Visin, F., Kastner, K., Cho, K., Matteucci, M., Courville, A., & Bengio, Y. (2015). *ReNet: A Recurrent Neural Network Based Alternative to Convolutional Networks*. http://arxiv.org/abs/1505.00393