Automated Root Nodule Detection

Priyam Mazumdar
Department of Statistics
University of Illinois Urbana-Champaign
Urbana, IL 61801
priyamm2@illinois.edu

Yang Yue
Department of Statistics
University of Illinois Urbana-Champaign
Urbana, IL 61801
yangyue3@illinois.edu

Simrita Jayasimha
Department of Computer Science
University of Illinois Urbana-Champaign
Urbana, IL 61801
simrita2@illinois.edu

Abstract

Machine Learning algorithms have been pushing massive advancements in automating many repetitive procedures. Applications of this technology would be especially useful in agricultural settings where farmers have to manage thousands of crops over acres of lands. Computer Vision techniques offer tools to make this possible to monitor plant health at large scale and we will be exploring detection of plant root nodules. Root nodules are small lumps found along the length of a root where nitrogen-fixing bacteria like Rhizobium reside inside the membrane [3]. This bacteria forms a symbiotic relationship with the plant by converting atmospheric nitrogen to ammonium that the plants can then use for their development. The number and quality of nodules can be an indicator of plant health and so having automated detection would be a powerful tool. In this research we will explore the use of different learning algorithms to detect root nodules and explore the benefits and pitfalls of the techniques.

1 Introduction

The quality and number of plant root nodules are a core indicator of plant development and health in ammonia-poor soils. Ammonia is the most important source of Nitrogen for plants and nearly 80 percent of the worlds ammonia production goes towards manufacturing fertilizers. Unfortunately fertilizer production and use have caused major increases to greenhouse gas release [6], harms soil fertility over constant exposure, pollutes bodies of water as the soil drains, and plays a significant role in climate change. Root nodules are a fully natural method of bacteria converting Nitrogen in the air to ammonia in soil that plants can leverage. Therefore for farmers and crop scientists, the detection of the number of root nodules are crucial to determine the plants health and soil fertility.

To perform this detection we will be exploring a few different methods. The first is a classification approach by splitting each image of roots into patches and the predicting if each patch contains a root nodule or not. We also tested prediction on features extracted from these patches such as applying a Histogram of Oriented Gradients. Our second method was to test traditional object detection algorithms such as YOLO. Lastly we wanted to perform semantic segmentation to see if we could detect the location of all nodules within an image.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

2 Dataset and Preprocessing

The dataset provided includes 10 images of plant roots as well as the coordinate bounds of the detection boxes around each nodule. From the supplied preprocessing, each image was resized to a 500 by 500 pixels size and ground-truth masks of the location of each bounding box was aligned.

This dataset allowed us to take a few different approaches in preprocessing. Our first goal was to attempt a classification based approach. We first calculated that the average size of a bounding box was around 5 by 7 pixels in our images, so we used a buffer of two pixels to generate 3 by 5 sized patches from the image, such that each patch was smaller than most of our nodules. After grabbing these patches from the image, each patch was then labeled with a binary classification for if there was a nodule located within the patch. We had significantly more patches without nodes than with, so we sampled and balanced our dataset for training. Once a model was trained, for inference, we could separate an image into patches and predict if a nodule is located in each patch. Depending on the prediction, the patch was then relabeled to 0 or 1 and all the patches were restructured back into the original image shape.

The second method was to explore semantic segmentation to find if we can utilize a UNet model [2] for predicting whether each pixel belongs to a nodule. A few changes were made to prepare data for this model. First, the original images provided were about 12 megapixels in size, so we could downsample to an image size of 2000 by 2000 without any interpolation effects. We don't mind having multiple nodes as we hope our segmentation methods can detect them, so we grabbed 64 by 64 image patches from the image and stored their corresponding segmentation masks generated from the bounding boxes. Also, most of the nodes were circular rather than the rectangular bounding boxes provided which would hurt segmentation performance. We instead converted these rectangular segmentation to circles with a diameter of the average of the length and width of the original bounding box.

3 Classification Based Approaches

Our first approach was to predict whether small patches of our image was a root nodule. Each patch was of size 3 by 5 pixels by 3 channels. This effectively meant there was 45 flattened input features to our classifier models which prompted a PCA based dimensionality reduction before prediction. It was hard to intuitively decide which classifier model may give the best performance, so we opted to test Random Forest, Support Vector Machines + HOG, and Ridge Regression Classifier.

3.1 Dimensionality Reduction

To better understand the high dimensional space we also provide some plots on PCA and T-SNE. Due to the large feature size, it was important to explore if we can compress our input data to the model and throw out redundant information. After performing PCA 1 we can see that we can easily explain the majority of the data variance with only 20 components so we can readjust our data by performing this transformation.

To also visualize the high dimensional space we can look at the T-SNE emebddings 2. Although we do see two clear groups between patches that include nodes and those that don't, there is also a large overlap between them. We would hypothesize that this overlap most likely comes from the roots themselves that have similar colors to the nodules. Because we flattened the vectors, we have thrown out all spatial information and we hope that the different models we experiment with can resolve the fine color differences between nodules and roots.

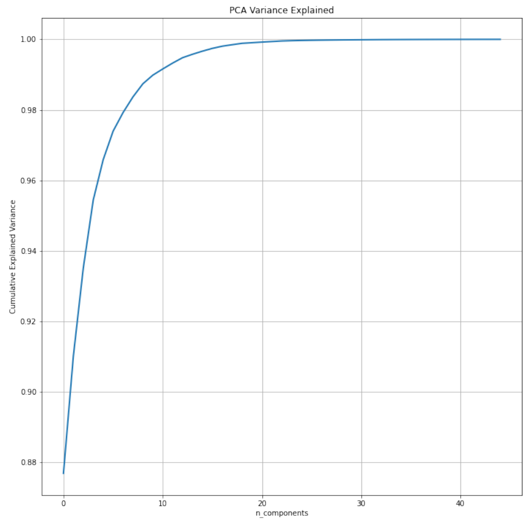

Figure 1: Variance Explained on 3 by 5 Patches from Images

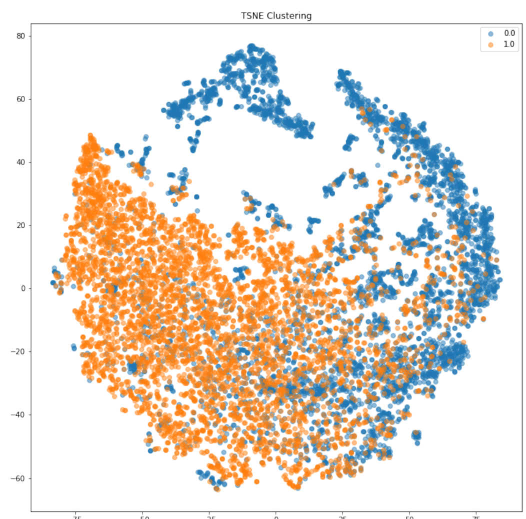

Figure 2: TSNE Embeddings on 3 by 5 Patches

3.2 Random Forest

The first model we tried was Random Forest as we believed that because we are looking for very subtle differences between nodules and other parts of the plant, it may be beneficial to utilize an ensemble method of many predictors. Through some additional model testing we saw that the Random Forest was very quick to overfit to the data and was sensitive to the number of estimators permitted.

After hyperparameter tuning the model settled on 16 estimators for both the raw data and the PCA compressed data. The results were about an 84.5% accuracy when predicting with all 45 features and 87.3% accuracy when using the first 20 components of the PCA.

When looking at the inference performance of the patches 3 we can start to see where this methodology works and problems that arise. First, we can see that we are picking up where most of the patches are, but it also seems to be selecting the roots themselves. The model also classified the border around the images as a part of the segmentation which is incorrect.

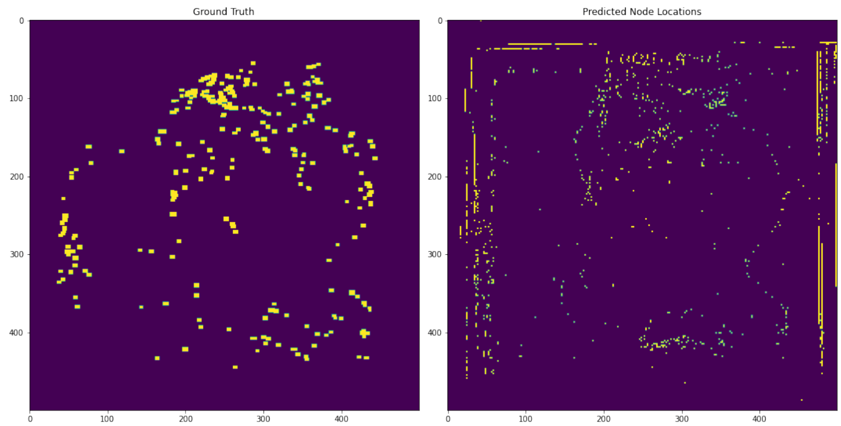

Figure 3: Random Forest Inference vs. Ground Truth

We can have a few majors takeaways from this that we will explore in the coming paper. Being able to distinguish between the nodes and roots based on RGB values mainly is not sufficient and we will have to see how we can improve the model via some gradient methods. Secondly, by using patch sizes smaller than the given bounding boxes, it is tougher to get an idea of the actual sizes of nodes based on the smaller predictions. Lastly, these nodes are very clearly defined by their shape and introducing spatial features will be key to build a successful classifier.

3.3 Ridge Regression Classifier

With the data labels converted into $\{-1, 1\}$, we can apply ridge regression to classification problems. As an extension of linear regression, ridge regression penalizes model complexity through regularization. By reducing the concern of model overfitting, the results tend to generalize better to new data.

Formally, the regularization is achieved by adding an additional term to the standard sum of squares cost function

$$\sum_{i=1}^{n}(y_i - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$$

In practice, we can tune the value of λ to adjust how much we want to penalize high values of β coefficients.

Similar to the our previous classification models, we opt to use a buffer size of 2 and end up with image patches of size 3 by 5 in the data preprocessing pipeline. We experiment with different λ values and achieved a testing accuracy of 80.35% without applying PCA to the data and 82.5% with PCA 4.

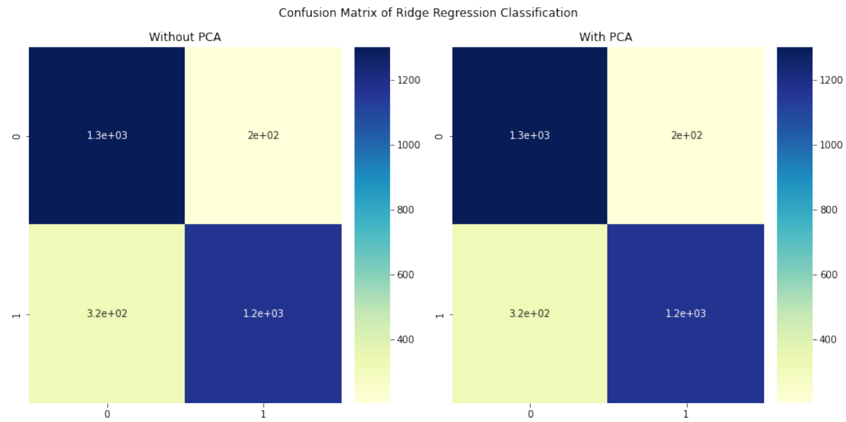

Figure 4: Confusion Matrix of Ridge Regression Classification

3.4 Support Vector Machine

The next binary classifier we explored was Support Vector Machine. For this classifier, we experimented with a buffer size of 3 and image patches of size 2 by 4 in the data preprocessing pipeline. SVM draws a decision boundary which is a hyperplane between any two classes in order to separate them, and uses regularization to prevent the model from overfitting. We applied this to the PCA compressed data using hinge loss. After hyper-parameter tuning, the best results were achieved using RBF kernel with regularization parameter 100 and gamma value 0.75. With this we achieved an accuracy score of 85.9% on the testing data. 5

4 Object Detection: HOG

The HOG (Histogram of Oriented Gradients) feature descriptor is used to count the occurrences of gradient orientation in localized portions of an image. The model works by calculating gradients for every pixel in each of the input images. Using the gradients, we determine the magnitude and direction for each pixel value. Finally, histograms are created using these gradients and orientations. We applied this HOG model to our patched images with 9 orientations, 2x2 pixels per cell and L2 block normalization to extract features from the root images. This was then fed into an SVM classifier which resulted in 96% accuracy.

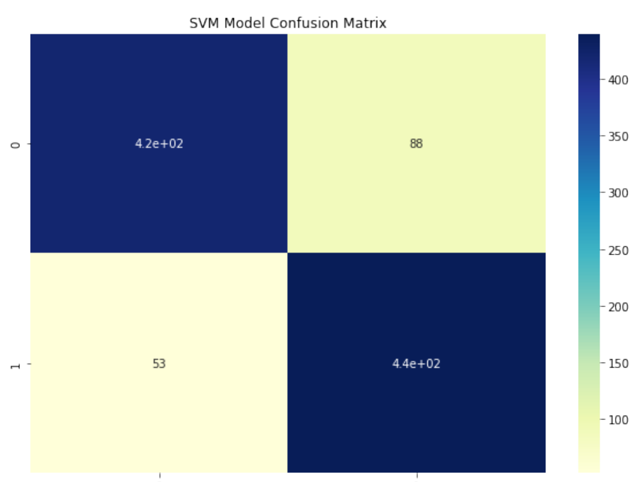

Figure 5: Confusion Matrix of Support Vector Classifier

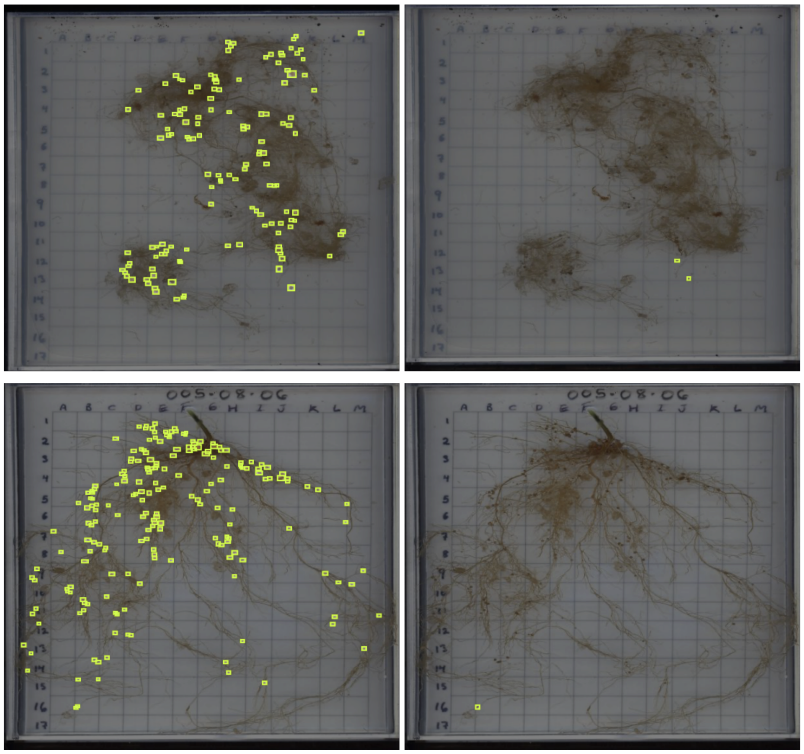

Figure 6: YOLO Detection of Roots

5 Object Detection: YOLO

Unlike other object detection systems that utilize localization in the classification process, YOLO model [1] applies a single neural network to the full image (hence "you only look once"). The network breaks the full image into different regions and predicts bounding boxes and the associated probabilities for each region. One obvious advantage is that YOLO can take the global context from the full image into account. And we believe this advantage is ideal for the task at hand because the images contain large area of null information. That is, the existence of our detection objectives, roots, is relatively scarce. And the ability to look at each image as a whole by the network could be valuable.

Unfortunately, as shown in 6, our YOLO model fails to generalize its learning to the validation and testing data set. The images in the first column show the ground truth bounding boxes for the roots and the second column contains the YOLO detection outcomes. While the model is able to identify some root nodes correctly, it fails to recognize the majority of the nodes in the images. On the one hand, it could be due the limited amount of training images that are available to us. On the other hand, we could increase the training epochs or look for suitable pre-trained models for future investigations.

6 Semantic Segmentation: UNet

Semantic segmentation will help solve many of the problems we encountered before where we weren't taking spatial features into consideration. To perform segmentation, we will rebuild our dataset to be 64 by 64 pixel patches and their corresponding ground truths. The ground truth segmentations have also been updated such that we are using circular masks for each nodule rather than rectangles given by the original dataset. An example image and mask pair is shown 7.

To perform this task we will be using a standard UNet segmentation model and attempt to minimize pixelwise Binary Cross Entropy. Due to evidence of overfitting it was also trained with the AdamW

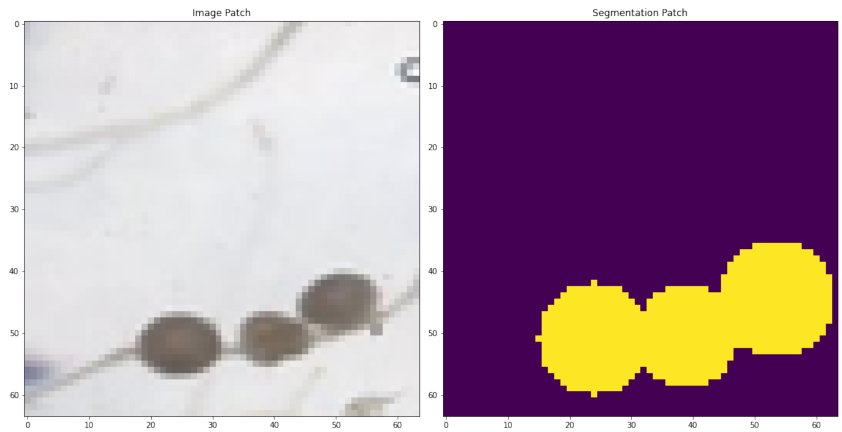

Figure 7: Example Segmentation Data Sample

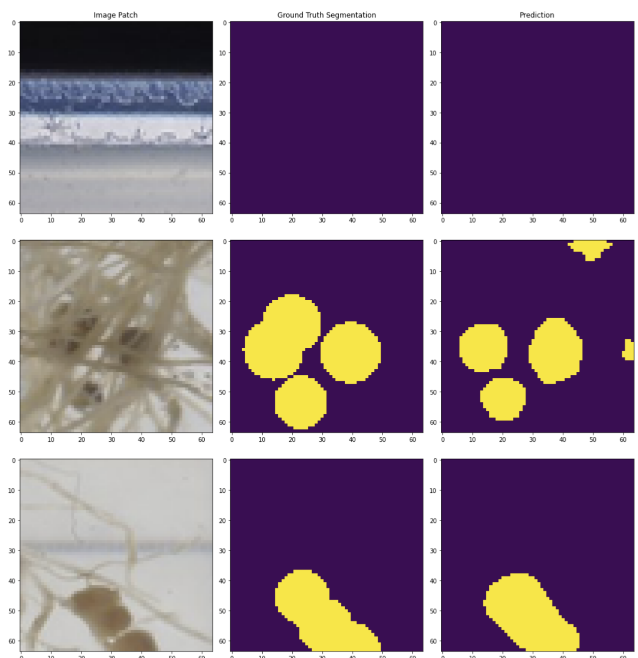

Figure 8: Predicted Segmentations with UNet

optimizer for weight penalties. Additionally, due to the relatively small dataset size (after patching we had about 5800 example images that included some segmentation information), we also leveraged random image transformations to promote generalization.

After our model was trained we were able to get some extremely promising results. We wanted to test for three separate cases, when nodes are easily visible, node that are occluded by roots and when no node is available. We can see in our segmentation 8 that we were able to detect nodes in heavy occlusion (although there were some slight mispredictions as well), when the nodes are easily seen we are able to find them, and finally when no nodes are available, nothing is predicted. This model returned about a 97% pixelwise accuracy and was our strongest performer. We also saw that our Jaccard Index (mIoU) was around a relatively strong 68% and we believe given some additional data we could have brought this up even higher.

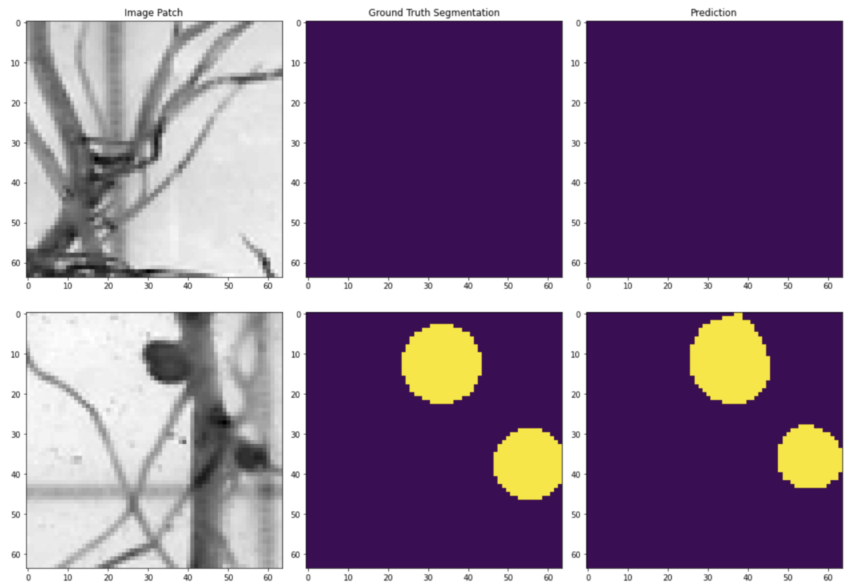

Figure 9: Predicted Segmentations with UNet with Grayscale Images

6.1 Importance of Spatial Features

We wanted to further test this hypothesis of the importance of spatial features. In standard ML algorithms where we flattened the image patches we were throwing away all spatial features and trying to solve the model mainly based on RGB color values. We wanted to reverse this idea to see if we throw away color but keep the spatial features how if may effect our model performance. To do this, we retrained the same UNet from before but passed in grayscale images of our patches and the results were very surprising 9. We were able to note a slight increase in performance to about a 69.7% mIoU which indicates to us that the structural features of how these nodules look are significantly more important than their colors. As a future exercise it would be interesting to explore how we can use transformations such as color jittering and solarization in the processing pipeline to increase model generalization.

7 Conclusion

The automation of root nodule detection would be an important tool for crop scientists to monitor the health of their plants. Our exploration looked at different patching techniques, object detection and segmentation. By patching, and performing additional preprocessing steps such as HOG, we were able to greatly boost our model classification to detect patches that included nodules. Unfortunately in the inference step, we see that the model isn't able to strongly separate between nodules and roots and it seems to miss many parts of the segmentation. We may be able to overcome this given more training data, as the number of patches that actually included nodules were very sparse. What we did find though is performing dimensionality reduction through PCA gave a small performance increase regardless of the model tested. Object detection through YOLO was our next attempt to see if we could directly predict bounding boxes based on the crops. Unfortunately the model missed the majority of nodes potentially due to only having ten images available. Our most successful method was to convert this task to a semantic segmentation problem and use UNet to indicate where nodules may exist. The powerful result of this method was we were able to detect nodules that were occluded by roots which would be necessary on any real world settings.

There are a few interesting next steps that can be taken for this project. First, additionally exploring using SVM as a method to directly predict the bounding boxes may return much better detection performance than seen with our initial attempts. Second, we could take an unsupervised approach, to pretrain a model like DINO [4] or the Masked Autoencoder [5] on many unlabeled images of roots. This may give us the ability to perform even more downstream tasks outside of just nodule detection with limited data. Lastly, a limitation of the current UNet approach is that we performed semantic segmentation but it would have been more appropriate to do Instance or Panoptic segmentation to be able to detect additionally how many nodes are in each image rather than just where the nodes are. Overall, we believe this was a strong exploration of a very small dataset that acts as a proof of concept that such detection is possible.

References

[1] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *CoRR* abs/1506.02640 (2015). arXiv: `1506.02640`. URL: `http://arxiv.org/abs/1506.02640`.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *CoRR* abs/1505.04597 (2015). arXiv: `1505.04597`. URL: `http://arxiv.org/abs/1505.04597`.

[3] Qi Wang, Jinge Liu, and Hongyan Zhu. "Genetic and Molecular Mechanisms Underlying Symbiotic Specificity in Legume-Rhizobium Interactions". In: *Frontiers in Plant Science* 9 (2018). ISSN: 1664-462X. DOI: `10.3389/fpls.2018.00313`. URL: `https://www.frontiersin.org/articles/10.3389/fpls.2018.00313`.

[4] Mathilde Caron et al. "Emerging Properties in Self-Supervised Vision Transformers". In: *CoRR* abs/2104.14294 (2021). arXiv: `2104.14294`. URL: `https://arxiv.org/abs/2104.14294`.

[5] Kaiming He et al. "Masked Autoencoders Are Scalable Vision Learners". In: *CoRR* abs/2111.06377 (2021). arXiv: `2111.06377`. URL: `https://arxiv.org/abs/2111.06377`.

[6] Stefano Menegat, Alicia Ledo, and Reyes Tirado. "Greenhouse gas emissions from global production and use of nitrogen synthetic fertilisers in agriculture". In: *Scientific Reports* 12.1 (2022), p. 14490. DOI: `10.1038/s41598-022-18773-w`. URL: `https://doi.org/10.1038/s41598-022-18773-w`.