



Deep learning-based high-throughput phenotyping can drive future discoveries in plant reproductive biology

Cedar Warman^{1,2} · John E. Fowler¹

Received: 15 November 2020 / Accepted: 15 February 2021
© The Author(s) 2021

Abstract

Key message Advances in deep learning are providing a powerful set of image analysis tools that are readily accessible for high-throughput phenotyping applications in plant reproductive biology.

High-throughput phenotyping systems are becoming critical for answering biological questions on a large scale. These systems have historically relied on traditional computer vision techniques. However, neural networks and specifically deep learning are rapidly becoming more powerful and easier to implement. Here, we examine how deep learning can drive phenotyping systems and be used to answer fundamental questions in reproductive biology. We describe previous applications of deep learning in the plant sciences, provide general recommendations for applying these methods to the study of plant reproduction, and present a case study in maize ear phenotyping. Finally, we highlight several examples where deep learning has enabled research that was previously out of reach and discuss the future outlook of these methods.

Keywords Deep learning · Computer vision · Neural network · Phenotyping · Reproduction

Introduction

Linking genotype to phenotype is a powerful method to understand complex biological systems. However, the increasing availability of genotypic data has not been matched by increasing capabilities for high-throughput phenotyping. This is particularly the case in the field of plant reproduction, where relevant biological processes are dependent on subtle cell–cell interactions, often taking place deep within vegetative and reproductive tissues. Until recently, computer vision-based plant phenotyping methods have been dominated by easily observable structures, such as leaves (Arvidsson et al. 2011; Zhang et al. 2012, 2017; Junker et al. 2014; Choudhury et al. 2016; Awlia et al. 2016) and roots (Yazdanbakhsh and Fisahn 2009; Clark et al. 2013; Slovak et al. 2014; Passot et al. 2018; Jiang et al. 2019).

These methods have relied on traditional computer vision, which leverages mathematical approaches such as thresholding (Otsu 1979), morphological opening and closing (Haralick et al. 1987), and k-means clustering (Lloyd 1982) to separate the object of interest from the background on a pixel level. These approaches are generally computationally efficient, but struggle when faced with complex backgrounds, closely packed objects, and structures with inconsistent features. Such challenging aspects are often encountered when phenotyping reproductive mutants, limiting the utility of traditional image processing approaches.

Over the past several years, neural network-based deep learning methods have been increasingly applied to address a wide range of biological questions (reviewed in (Ching et al. 2018)). Neural network-based computer vision approaches are fundamentally different from the previous generation of image processing techniques because they rely on large sets of training data to function. Neural networks are composed of multiple layers. For image analysis, the first layer, called the input layer, is made up of the image pixels. The final layer, called the output layer, contains the model's predictions. Between the input and output layers are one or more hidden layers. Each layer is composed of a number of nodes (or neurons), which perform simple mathematical operations. Nodes are variously connected to adjacent layers and

Communicated by Dolf Weijers.

✉ Cedar Warman
cedardalewarman@gmail.com

¹ Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

² School of Plant Sciences, University of Arizona, Tucson, AZ, USA

have adjustable parameters that can be trained computationally to accurately predict patterns in the input layer. Data flow from the input layer, through the hidden layers, and finally to the output layer. Neural networks with multiple hidden layers are called deep neural networks.

Neural networks have been used as versatile solutions to many phenotyping challenges. Recent improvements in ease of use and the development of advanced computational resources make deep learning strategies well-situated to drive future discoveries in plant reproductive biology. Here, we examine potential applications of deep learning-based image analysis to overcome phenotyping challenges. We highlight current plant phenotyping systems that take advantage of deep learning, as well as useful deep learning frameworks, guidelines for implementing models, and the future outlook for these methods in plant reproductive biology.

Opportunities for high-throughput phenotyping in plant reproduction

To date, deep learning approaches in plant phenotyping have been focused on similar targets to traditional computer vision approaches, such as leaves (Ubbens and Stavness 2017; Ziamtsov and Navlakha 2019; Hüther et al. 2020) and roots (Douarre et al. 2018; Wang et al. 2019; Yasrab et al. 2019). Additional applications of deep learning in plants include disease identification (Wang et al. 2017; Polder et al. 2019), inflorescence motion tracking (Gibbs et al. 2019), and fruit shape (Feldmann et al. 2020). These implementations have performed well on images that challenge traditional computer vision approaches, particularly those with complex target objects and backgrounds. However, published tools for phenotyping reproductive tissues are notably scarce.

Deep learning approaches first require digital representations of the phenotype(s) of interest, generally in the form of images. Conceptually, there are three general targets for plant reproductive phenotyping (Fig. 1a). For the first, whole plant or individual organs such as inflorescences and fruit can be imaged using existing methods. Whole plant or

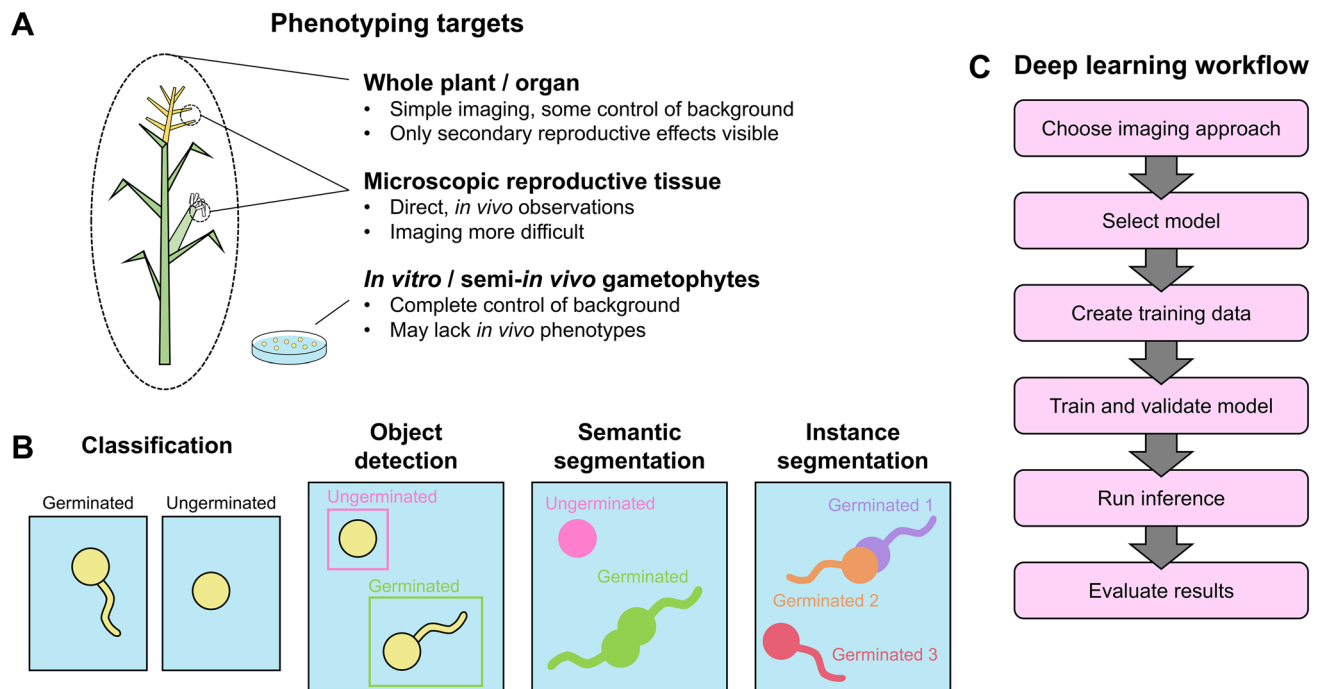


Fig. 1 Leveraging deep learning for high-throughput phenotyping in plant reproduction. **a** Potential imaging targets for high-throughput phenotyping of reproductive systems can range from the whole plant to individual organs to microscopic structures, both *in vivo* and *in vitro*. **b** General deep learning strategies for image analysis. In classification, the class of each image is described, here germinated versus ungerminated pollen. When a single image contains more than one object, object detection can be used, a method that identi-

fies objects and classes by bounding boxes. Semantic segmentation identifies the class of objects in an image on a pixel level, allowing for the identification of object attributes like shape and area. As with semantic segmentation, instance segmentation identifies pixel classes. In addition, instance segmentation differentiates multiple instances of the same object class that are touching or overlapping. **c** Conceptual steps for implementing deep learning models

organ imaging is generally simpler than other approaches, but observations are largely limited to secondary effects of reproductive processes, such as abnormalities in fruit shape due to aberrant seed set. Such imaging has been used to track mutant inheritance through kernel phenotypes (Warman et al. 2021), estimate seed yield (Uzal et al. 2018; Khaki et al. 2020), measure fruit characteristics (Hamidinekoo et al. 2020), and classify kernel abortion (Chipindu et al. 2020). Key considerations for whole plant or organ imaging include imaging systems, background selection, and development of methods for discerning the secondary reproductive effects of interest.

A second target for reproductive phenotyping is small reproductive tissues, specifically the male and female gametophytes, the zygote and embryo, and surrounding sporophytic tissues. Phenotyping these tissues are critical because they offer the ability to directly observe reproductive processes *in vivo*. However, obtaining microscopic images of these tissues on a large scale can be challenging. Often the gametophytes are located deep within other tissues and require special staining or fluorescent markers to visualize structures of interest. In addition, complicated backgrounds can make computer vision approaches challenging, even with the use of deep learning approaches. Preliminary applications of deep learning in such instances have focused on identifying cells from images of plant tissues using segmentation (Wolny et al. 2020). These methods have been used to describe ovule development (Vijayan et al. 2021).

A final type of phenotyping target is gametophytes or zygotes imaged *in vitro* or semi-*in vivo* (Higashiyama et al. 2001; Palanivelu and Preuss 2006). These approaches have the potential to allow high-throughput imaging of gametophytes in a controlled environment, making them amenable to deep learning approaches, but may fail to capture some phenotypes that are only observed *in vivo*. Pollen has been the primary target of these types of approaches, likely because it is relatively easy to isolate compared to the female gametophyte. Previous studies have used deep learning to identify pollen from several different species (Dunker et al. 2020) and differentiate pollen developmental stages (García-Forkea et al. 2020).

A variety of general purpose deep learning frameworks can be applied to reproductive phenotyping

Deep learning for image analysis can be broadly divided into four categories based on how objects are identified in the image: classification, object detection, semantic segmentation, and instance segmentation (Fig. 1b). In image recognition problems, classification answers what class, or type of object, describes a single image. Classification is

possible when individual images contain only one object to be identified, such as single germinated or ungerminated pollen grains (Fig. 1b). If more than one object is present in an image or if the location of the object needs to be determined, an object detection approach may be more suitable. Object detection uses deep learning (often convolutional neural networks or CNNs) to identify both the classes and locations of multiple objects in an image, such as germinated and ungerminated pollen grains, and signifies them by labeled bounding boxes. Some types of experiments, such as shape or area measurements, may require semantic segmentation or instance segmentation. These approaches identify the class of every pixel in an image, outputting a mask over objects in the image that shows not only their location, but their shape. Instance segmentation goes a step farther than semantic segmentation by resolving touching or overlapping objects, and thus also enables the identification of individual objects. Instance segmentation has been used for crop seed phenotyping (Toda et al. 2020), and both methods show potential for measuring other reproductive phenotypes, such as the path of growing pollen tubes or the area of developing embryos. The choice of deep learning strategy will vary by phenotyping task and is important to consider before choosing a model, as models are optimized for specific strategies. An additional consideration is the ease of generating training data, a challenge that is described in more detail in the following section. While all deep learning approaches require training data, creating these data for classification tasks are faster than for object detection and semantic segmentation.

Several programming frameworks are available for implementing deep learning models. Most require some knowledge of the Python or R programming languages, as well as a basic familiarity with command line interfaces and Linux/Unix operating systems. Many machine learning libraries exist, but here we highlight two freely available and open source frameworks that could serve as a useful starting point. Currently, the most popular and versatile library is TensorFlow (Abadi et al. 2016). TensorFlow has a well-developed set of machine learning tools, as well as thorough documentation and an active community of users. Many of the newest deep learning models are implemented in TensorFlow, and it remains at the forefront of machine learning. Taking full advantage of TensorFlow requires a thorough understanding of neural networks and a high degree of coding ability. However, the Keras application programming interface (API) simplifies many of TensorFlow's functions (Chollet and Others 2015) and increases accessibility to researchers who are not primarily focused on machine learning. Several well-established and generalizable TensorFlow methods exist for computer vision tasks, including for object detection and semantic segmentation.

PyTorch (Paszke et al. 2019) is a major alternative to TensorFlow. While the user base and available resources

are not as large as with TensorFlow, PyTorch provides some of the same versatile building blocks for neural network construction. PyTorch is tightly integrated with the Python programming language, so researchers with experience in Python may find PyTorch more intuitive to learn than TensorFlow. Similar network architectures can be created with both TensorFlow and PyTorch, so the best library for biological image analysis will depend both on the researcher's preferences and the availability of established methods in each library.

Implementing deep learning methods

A general workflow for implementing deep learning models can be broken down into a number of conceptual steps (Fig. 1c). Before pursuing deep learning, it is important to ensure the research question is best answered by this approach. Generating large amounts of high-quality training data require a substantial initial time investment. Traditional computer vision methods may be better suited to certain challenges, such as the segmentation of a green plant from a solid background. In addition, hand annotation might be a more efficient method with small sets of images. Deep learning approaches only start to see returns in truly high-throughput cases, with the analysis of hundreds or thousands of images.

If deep learning is well suited to the research question at hand, the first step is to choose which tissues or structures to image, and to develop a process to generate a large quantity of images. Many commercial and academic phenotyping platforms exist for image generation, but the specific challenges of imaging reproductive structures may require the development of new technologies. Alongside image generation, the researcher should choose the category of image analysis (classification, object detection, semantic segmentation, or instance segmentation) and deep learning model that can most effectively answer the research question. For example, counting fertilized ovules may only require object detection, whereas measuring the size of developing seeds could benefit from semantic segmentation.

Next, labeled training images must be generated, and subsequently split into training, validation, and test sets. Training and validation sets allow the model to learn and evaluate itself during the training process, whereas test sets are reserved for unbiased final evaluations of the trained model. A key consideration when training any deep learning model is to avoid overfitting the data. Overfitting results when the model predicts the training data too closely and is consequently not generalizable to new data. Reserving an exclusive test dataset aside from training and validation datasets provides a final check that the model is generalizable beyond the datasets used for its creation.

The number of training images required for an accurate model will vary by task, but expect to generate between 100 and 1000 training images, at a minimum. Measuring highly variable phenotypes, such as identifying cell types in mutant ovules, will likely require more training data than measuring relatively straightforward phenotypes, such as counting seeds on a solid background. One commonly used strategy to leverage small amounts of training data is to employ transfer learning (Douarre et al. 2018; Hüther et al. 2020). Transfer learning begins with a well-trained model from a previous deep learning task, often a generalized task such as the identification of common objects, then retraining the top layers of the model with a new training set for the task at hand. Another strategy is to use semi-automated methods to increase the size of training sets, a particularly useful method for semantic segmentation training data (Adams et al. 2020).

After sufficient training data are generated, the model is trained. In deep learning models, training is an iterative process in which parameters of the network's nodes are adjusted to optimize predictions. After each iteration, or epoch, the model adjusts to better predict outputs. Adjustments are typically made to minimize a "loss function," which is a measurement of how accurately the model predicted validation data. Model training time depends on the size of the training dataset and model complexity, and can vary widely, from minutes to days or even longer. The process can be vastly sped up with high performance GPUs, which are available through local or cloud-based computational resources. Many cloud services offer computational resources optimized for deep learning frameworks.

Once a model is trained, predictions, or inference, can be made on experimental datasets. Inference is typically less computationally intensive than model training and can often be run on a consumer grade machine. It is important that the researcher defines how the performance of the model is evaluated throughout the deep learning process. For image classification tasks, performance metrics typically aim to quantify the number of true positive, false positive, false negative, and true negative predictions (often termed the "confusion matrix"). Specific combinations of these quantities include accuracy, precision, recall, and specificity. "Area under the receiver operating characteristic curve" (AUC) is a more thorough formulation of the previous measures. For object detection and semantic segmentation, the accuracy of object location and size predictions can be evaluated using intersection over union (IoU) and mean average precision (mAP). More accessible metrics can also be used. For example, with regression tasks, coefficients of determination comparing model predictions to ground truth are useful metrics, as well as mean absolute deviations.

Maize ear phenotyping: a case study

Our recent work to create a high-throughput phenotyping system for maize ears can serve as a practical example of one implementation of deep learning methods for plant reproductive phenotyping (Warman et al. 2021). Our goal was to track changes in Mendelian inheritance caused by mutations in genes highly expressed in the male gametophyte. To do so, we used a collection of 56 GFP-tagged transposable element insertion mutants (Li et al. 2013; Warman et al. 2020). Kernels carrying mutant alleles express the Green Fluorescent Protein (GFP), allowing quantification of mutant transmission by counting fluorescent kernels. To facilitate high-throughput phenotyping, our method targeted the maize ear as a proxy, as functional defects in the male gametophyte will result in altered transmission frequencies. We developed a custom rotational scanning system to capture images of the ear for downstream analysis (Fig. 2a). We chose to use a deep learning approach to quantify kernels because traditional computer vision methods failed to overcome variations in our images, such as kernel shape and fluorescence intensity. For our deep learning model, we used the TensorFlow

Object Detection API (Huang et al. 2016) implementation of Faster R-CNN with Inception Resnet v2 (Ren et al. 2015; Szegedy et al. 2016). This approach combined a powerful model with a simple API. We chose object detection because our primary goal was to count kernels to track mutant transmission. Segmentation-based approaches would have provided more detailed descriptions of each kernel, but bounding boxes gave us the information we needed, while minimizing the time required to produce training data. For training the model, we generated 300 images annotated with bounding boxes (Fig. 2b), evenly distributed across ears generated in two field seasons. Each image took approximately 20 min to label, for a total of 100 h of labor to create the entire dataset.

After image annotation, we split the dataset into training and validation sets. We used a transfer learning approach, taking advantage of a network pre-trained on the Collection of Common Objects (COCO) dataset, which consists of ~200,000 labeled images of everyday objects. Transfer learning reduced both the amount of training data that was required, as well as the time required to train our model. The model was trained on an Nvidia V100 GPU for approximately one hour. After the model was trained, we performed inference on the testing set, a set of 320 manually counted

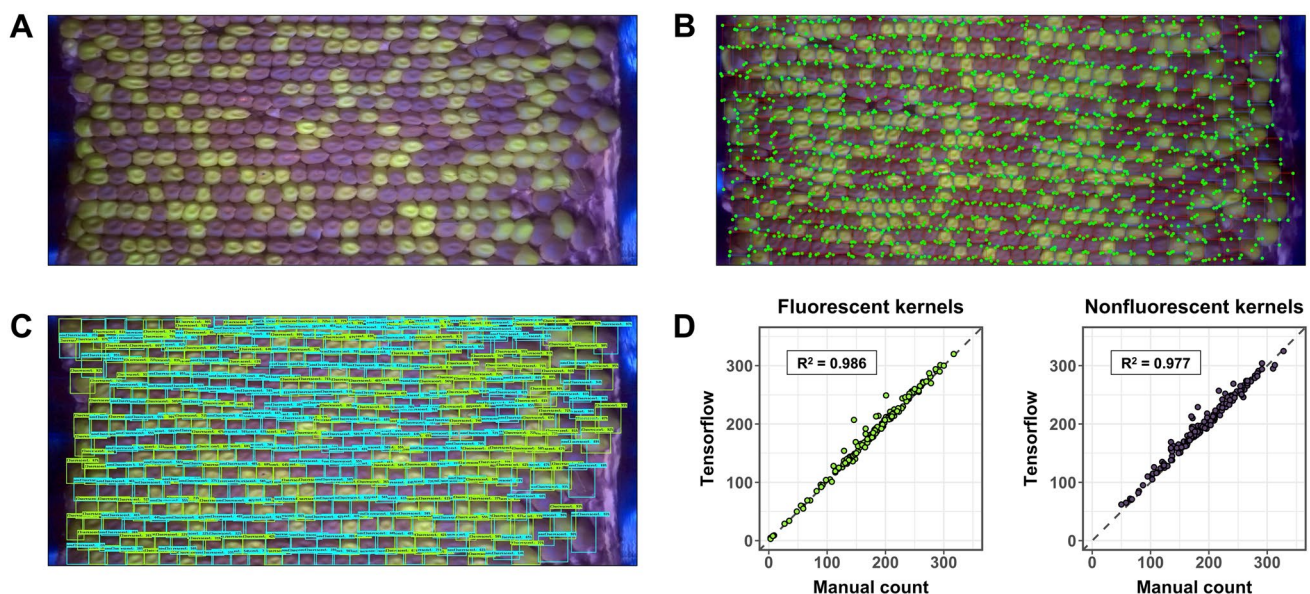


Fig. 2 Maize ear phenotyping as an example deep learning workflow. **a** First, a rotational scanning system creates a flat projection of the surface of an ear. Fluorescent kernel markers are visible in this projection, signifying the presence of a genetically engineered transposable element insertion in a gene of interest. The ratio of fluorescent (mutant) kernels to non-fluorescent (wild-type) kernels can be tracked to screen for non-Mendelian inheritance of the mutant alleles. **b** Next, a training set of 300 projections with manually assigned bounding boxes labeling each kernel (corners marked by green circles) was

generated. A transfer learning approach and the TensorFlow Object Detection API was then used to create a model based on the training dataset. **c** Model inference on the independent test set generates bounding boxes predicting the locations of the objects of interest in the image. Blue boxes signify non-fluorescent kernels and green boxes signify fluorescent kernels. **d** A comparison between model predictions and manual counts for fluorescent and non-fluorescent kernels (160 ear projections) was used to validate the model

images that were not used in the training process (Fig. 2c). We measured mAP, as well as calculated linear regressions and mean absolute deviations comparing hand counts to model predictions. The outcome was a highly accurate model, with adjusted R^2 for kernel counts between 0.96 and 0.98, depending on field season and phenotypic marker (Fig. 2d).

The main challenges to implementing deep learning methods were imaging and generating the training data. We found that using consistent imaging conditions (e.g., the same camera) was important, as networks trained on data from one camera were not easily transferable to data collected on another camera. In addition, the large number of kernels present in each image reached the limits of our GPU's memory capacity, requiring a strategy of image subdivision and recombination. Ultimately, deep learning methods enabled us to both (1) survey a larger number of ears than would have been possible by manual counting, leading to the discovery of new mutants with quantitative phenotypic effects; and (2) generate larger datasets for each allele, enabling a more accurate estimation of transmission rate and increased statistical power to identify significant effects.

The future of deep learning in reproductive phenotyping

Deep learning is a rapidly advancing field. Many deep learning models are general purpose, allowing advancements in unrelated fields to be rapidly applied to specific cases like plant reproduction phenotyping. A thorough understanding of general purpose machine learning libraries like TensorFlow and PyTorch will enable biologists to take advantage of new research in image analysis. While classification and object detection models are relatively mature, semantic segmentation models are an area of active research. These models will gradually improve, allowing for more reliable identification of complex objects in noisy images, such as pollen tubes growing down the pistil or ovules at fertilization.

Large, high-quality training datasets improve the accuracy and versatility of models. Generating training datasets remains a major barrier to implementing these models. However, various methods of data augmentation can help researchers make the most out of available data. Traditional image augmentations include transformations like random cropping, horizontal and vertical flips, hue adjustments, and the addition of random noise. Recent work has developed new methods for increasing the size of training sets. One method, called "domain randomization," has successfully been used in wheat to generate artificial training data by semi-randomly layering cropped images of single wheat seeds to simulate a large number of seeds on a flat surface

(Toda et al. 2020). Domain randomization has also been used in Arabidopsis to generate artificial training data for leaf segmentation (Ward et al. 2018). Another method takes advantage of generative adversarial networks, or GANs (Zhu et al. 2018). GANs are a type of machine learning that uses two neural networks to compete against each other to generate realistic data. One network generates data (in this case images), while the other network evaluates the data to determine which examples are real and which are artificial. The network learns from these comparisons and can use this knowledge to create novel images that have the characteristics of the input set. These types of networks have been used to improve accuracy by increasing the size of training datasets.

Another potential future direction is an expansion of what constitutes a phenotype. Historically, phenotypes have included simple concepts like malformed or undeveloped structures, growth defects, or color changes, as well as quantitative effects, such as natural variation in fertilization success in populations across a species' geographic distribution. Neural networks allow for the measurement of more subtle, multidimensional phenotypes, such as abstractions for how images of plants change in response to treatments like drought or heat. An application of this concept, called latent space phenotyping (LSP), shows strong potential at characterizing changes in images of plants unrecognizable by the human eye (Gage et al. 2019; Ubbens et al. 2020). This approach compares sets of images from treated and control samples by using a neural network to create compressed representations of visual characteristics of the images, termed latent space. This process does not require labeled data beyond experimental and control sets, relying on the neural network to determine multidimensional phenotypes characteristic of the experimental set. These LSPs can then be interpreted using additional networks, ultimately uncovering complex phenotypes that, for example, can be used for genome-wide association studies (GWAS). LSP circumvents the need for hand-labeled training sets and also has the potential to incorporate temporal changes into such phenotypes (Taghavi Namin et al. 2018). Additional methods can take more explicit approaches, such as multitask learning, which optimizes networks based on several predetermined phenotypes simultaneously (Dobrescu et al. 2020).

Implementing deep learning methods for phenotyping can be intimidating, as the time and effort necessary to develop a successful model is significant, and therefore not appropriate for all experiments. However, the power and versatility of these methods make them particularly useful for high-throughput phenotyping, allowing for increased throughput, precision, and the measurement of complex phenotypes that may be difficult to describe using more traditional methods. As their use spreads, deep learning methods will become more accessible, making now an ideal time for applications

in high-throughput phenotyping in plant reproductive biology.

Definitions sidebar: deep learning terminology

Neural network A biologically inspired computing system that transforms input data (in this case images) into outputs (in this case predictions). Neural networks contain nodes, inspired by biological neurons, that are arranged in multiple layers, or collections of nodes. In its most basic form, each node can aggregate data from previous nodes, transform the data using trainable weights, and send the data to nodes in the following layer. Deep learning is a subset of machine learning that combines many-layered neural networks with large amounts of training data to enable predictions from heterogeneous datasets.

Classification The prediction of a label (also known as a class) from input data. For a single image input, a single class would be predicted, such as germinated or ungerminated pollen.

Object detection The detection and localization of objects of one or more classes in an image. Unlike classification, object detection enables multiple objects to be identified in a single image.

Semantic segmentation Classification at the pixel level. Semantic segmentation enables the identification of objects in greater detail than object detection.

Instance segmentation Classification at the pixel level, with the addition of differentiation of touching or overlapping instances of one or more classes.

Convolutional neural network (CNN) A neural network containing convolutional layers. Convolutional layers aggregate image data by moving a filter across the image and summarizing pixel values using mathematical operations. Convolutional layers allow the network to learn patterns in pixel values across the entire image.

Training, validation, and test sets Data used to train a neural network model are typically divided into training, validation, and test sets. The training set, usually the largest portion of the input data, is used to train the model. During each round of training, the training set is used to adjust the weights of individual nodes to improve the model's accuracy. The validation set is used in the training process to evaluate the

model's performance after each training round. Separate training and validation sets are used to prevent overfitting, a situation that can arise from the model learning the patterns of the training dataset too specifically and lacking generalizability. The test set is used after training to evaluate the model. Because the model is not exposed to the test set until after training, evaluating the model with the test set provides an unbiased method to test the model's performance.

Loss function During the training process, neural networks are evaluated using a loss function. Loss functions quantify the accuracy of the network's predictions. Neural networks iteratively optimize the loss function during training using a process called stochastic gradient descent.

Definitions sidebar: evaluating model performance

Confusion matrix Table used to evaluate binary classification models, based on class predictions and class ground truth.

		Actual class	
		Positive	Negative
Predicted class	Positive	True positive (TP)	False positive (FP)
	Negative	False negative (FN)	True negative (TN)

Accuracy $\frac{TP+TN}{TP+TN+FP+FN}$ A measure of how often the model predicted the correct class. Accuracy is a poor measure of model performance because models with higher accuracy do not always have more predictive power than models with lower accuracy.

Precision $\frac{TP}{TP+FP}$ Precision measures the fraction of correct positives over the total number of positive predictions. This metric is useful for measuring a model's performance when true positives are uncommon, and the cost of a false positive is high.

Recall $\frac{TP}{TP+FN}$ Recall measures the proportion of positives that were correctly identified by the model. Precision and recall are often reported together, as an increase in recall is typically associated with a decrease in precision.

Specificity $\frac{TN}{TN+FP}$ Specificity measures the proportion of negatives that were correctly identified by the model.

Area under the receiver operating characteristic curve (AUC) AUC is a measure of a binary classification model's performance that summarizes true positive and false positive rates. An AUC of 1 represents perfect separation of two classes, whereas a model with an AUC of 0.5 performs no better than random.

Intersection over union (IoU) IoU is a method to evaluate predictions in object detection and semantic segmentation models. The method compares predicted bounding boxes or masks to ground truth annotations. Intersection describes the overlapping area between the prediction and the ground truth. Union describes the total area covered by both the prediction and the ground truth. If the prediction and ground truth perfectly overlap, IoU is 1, whereas incomplete overlap generates values less than 1.

Mean average precision (mAP) mAP is a performance metric often used to evaluate object detection or semantic segmentation models. This method summarizes a precision recall curve (similar to AUC) across all classes and at various IoU thresholds.

Author contribution statements CW conceptualized the article and performed literature review. CW and JF wrote and revised the manuscript. CW and JF read and approved the manuscript.

Acknowledgements The authors sincerely thank Dr. Ravishankar Palanivelu for reading the manuscript and providing excellent suggestions. In addition, we thank Justin Preece and Christopher M. Sullivan for guidance in deep learning techniques and computational infrastructure.

Funding This work was funded in part by National Science Foundation grant no. IOS-1832186 and the Department of Botany & Plant Pathology at Oregon State University.

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Informed consent All the authors gave their consent to participation and for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abadi M, Agarwal A, Barham P, et al (2016) Tensor flow: large-scale machine learning on heterogeneous distributed systems. arXiv [cs.DC]
- Adams J, Qiu Y, Xu Y, Schnable JC (2020) Plant segmentation by supervised machine learning methods. *Plant Phenom J* 3:6980. <https://doi.org/10.1002/ppj2.20001>
- Arvidsson S, Pérez-Rodríguez P, Mueller-Roeber B (2011) A growth phenotyping pipeline for *Arabidopsis thaliana* integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytol* 191:895–907. <https://doi.org/10.1111/j.1469-8137.2011.03756.x>
- Awlia M, Nigro A, Fajkus J et al (2016) High-throughput non-destructive phenotyping of traits that contribute to salinity tolerance in *Arabidopsis thaliana*. *Front Plant Sci* 7:1414. <https://doi.org/10.3389/fpls.2016.01414>
- Ching T, Himmelstein DS, Beaulieu-Jones BK et al (2018) Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*. <https://doi.org/10.1098/rsif.2017.0387>
- Chipindu L, Mupangwa W, Mtsilizah J et al (2020) Maize kernel abortion recognition and classification using binary classification machine learning algorithms and deep convolutional neural networks. *AI* 1:361–375. <https://doi.org/10.3390/ai1030024>
- Chollet F, Others (2015) Keras. <https://keras.io>
- Choudhury SD, Stoerger V, Samal A, et al (2016) Automated vegetative stage phenotyping analysis of maize plants using visible light images. In: KDD workshop on data science for food energy and water, San Francisco: California USA. researchgate.net
- Clark RT, Famoso AN, Zhao K et al (2013) High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. *Plant Cell Environ* 36:454–466. <https://doi.org/10.1111/j.1365-3040.2012.02587.x>
- Dobrescu A, Giuffrida MV, Tsafaris SA (2020) Doing more with less: a multitask deep learning approach in plant phenotyping. *Front Plant Sci* 11:141. <https://doi.org/10.3389/fpls.2020.00141>
- Douarre C, Schielein R, Frindel C et al (2018) Transfer learning from synthetic data applied to soil-root segmentation in X-ray tomography images. *J Imaging* 4:65. <https://doi.org/10.3390/jimaging4050065>
- Dunker S, Motivans E, Rakosy D et al (2020) Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytol*. <https://doi.org/10.1111/nph.16882>
- Feldmann MJ, Hardigan MA, Famula RA et al (2020) Multi-dimensional machine learning approaches for fruit shape phenotyping in strawberry. *Gigascience*. <https://doi.org/10.1093/gigascience/giaa030>
- Gage JL, Richards E, Lepak N et al (2019) In-Field Whole-plant maize architecture characterized by subcanopy rovers and latent space phenotyping. *Plant Phenom J* 2:1–11. <https://doi.org/10.2135/tppj2019.07.0011>
- García-Fortea E, García-Pérez A, Gimeno-Páez E et al (2020) a deep learning-based system (microscan) for the identification of pollen development stages and its application to obtaining doubled haploid lines in *Eggplant*. *Biology*. <https://doi.org/10.3390/biology9090272>
- Gibbs JA, Burgess AJ, Pound MP et al (2019) Recovering wind-induced plant motion in dense field environments via deep learning and multiple object tracking. *Plant Physiol* 181:28–42. <https://doi.org/10.1104/pp.19.00141>

- Hamidinekoo A, Garzón-Martínez GA, Ghahremani M et al (2020) DeepPod: a convolutional neural network based quantification of fruit number in Arabidopsis. *Gigascience*. <https://doi.org/10.1093/gigascience/giaa012>
- Haralick RM, Sternberg SR, Zhuang X (1987) Image analysis using mathematical morphology. *IEEE Trans Pattern Anal Mach Intell* 9:532–550. <https://doi.org/10.1109/tpami.1987.4767941>
- Higashiyama T, Yabe S, Sasaki N et al (2001) Pollen tube attraction by the synergid cell. *Science* 293:1480–1483. <https://doi.org/10.1126/science.1062429>
- Huang J, Rathod V, Sun C, et al (2016) Speed/accuracy trade-offs for modern convolutional object detectors. *arXiv [cs.CV]*
- Hüther P, Schandry N, Jandrasits K, et al (2020) araDEEPopsis: from images to phenotypic traits using deep transfer learning. *bioRxiv* 2020.04.01.018192
- Jiang N, Floro E, Bray AL et al (2019) Three-Dimensional time-lapse analysis reveals multiscale relationships in maize root systems with contrasting architectures. *Plant Cell* 31:1708–1722. <https://doi.org/10.1105/tpc.19.00015>
- Junker A, Muraya MM, Weigelt-Fischer K et al (2014) Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front Plant Sci* 5:770. <https://doi.org/10.3389/fpls.2014.00770>
- Khaki S, Pham H, Han Y et al (2020) Convolutional neural networks for image-based corn kernel detection and counting. *Sensors*. <https://doi.org/10.3390/s20092721>
- Li Y, Segal G, Wang Q, Dooner HK (2013) Gene tagging with engineered ds elements in maize. In: Peterson T (ed) *Plant transposable elements: methods and protocols*. Humana Press, Totowa, NJ, pp 83–99
- Lloyd S (1982) Least squares quantization in PCM. *IEEE Trans Inf Theor* 28:129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9:62–66
- Palanivelu R, Preuss D (2006) Distinct short-range ovule signals attract or repel Arabidopsis thaliana pollen tubes in vitro. *BMC Plant Biol* 6:7. <https://doi.org/10.1186/1471-2229-6-7>
- Passot S, Moreno-Ortega B, Moukouanga D et al (2018) A new phenotyping pipeline reveals three types of lateral roots and a random branching pattern in two cereals. *Plant Physiol* 177:896–910. <https://doi.org/10.1104/pp.17.01648>
- Paszke A, Gross S, Massa F et al (2019) PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A et al (eds) *Advances in neural information processing systems*, vol 32. Curran Associates Inc, Red Hook, pp 8026–8037
- Polder G, Blok PM, de Villiers HAC et al (2019) Potato virus Y detection in seed potatoes using deep learning on hyperspectral images. *Front Plant Sci* 10:209. <https://doi.org/10.3389/fpls.2019.00209>
- Ren S, He K, Girshick R, Sun J (2015) Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes C, Lawrence ND, Lee DD et al (eds) *Advances in neural information processing systems*, vol 28. Curran Associates Inc., Red Hook, pp 91–99
- Slovak R, Göschl C, Su X et al (2014) A scalable open-source pipeline for large-scale root phenotyping of Arabidopsis. *Plant Cell* 26:2390–2403. <https://doi.org/10.1105/tpc.114.124032>
- Szegedy C, Ioffe S, Vanhoucke V, Alemi A (2016) Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv [cs.CV]*
- Taghavi Namin S, Esmaeilzadeh M, Najafi M et al (2018) Deep phenotyping: deep learning for temporal phenotype/genotype classification. *Plant Methods* 14:66. <https://doi.org/10.1186/s13007-018-0333-4>
- Toda Y, Okura F, Ito J et al (2020) Training instance segmentation neural network with synthetic datasets for crop seed phenotyping. *Commun Biol* 3:173. <https://doi.org/10.1038/s42003-020-0905-5>
- Ubbens J, Cieslak M, Prusinkiewicz P et al (2020) Latent space phenotyping: automatic image-based phenotyping for treatment studies. *Plant Phenoms* 2020:5801869. <https://doi.org/10.34133/2020/5801869>
- Ubbens JR, Stavness I (2017) Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front Plant Sci* 8:1190. <https://doi.org/10.3389/fpls.2017.01190>
- Uzal LC, Grinblat GL, Namías R et al (2018) Seed-per-pod estimation for plant breeding using deep learning. *Comput Electron Agric* 150:196–204. <https://doi.org/10.1016/j.compag.2018.04.024>
- Vijayan A, Tofanelli R, Strauss S et al (2021) A digital 3D reference atlas reveals cellular growth patterns shaping the Arabidopsis ovule. *Elife*. <https://doi.org/10.7554/eLife.63262>
- Wang G, Sun Y, Wang J (2017) Automatic image-based plant disease severity estimation using deep learning. *Comput Intel Neurosci* 2017:2917536. <https://doi.org/10.1155/2017/2917536>
- Wang T, Rostamza M, Song Z et al (2019) SegRoot: a high throughput segmentation method for root image analysis. *Comput Electron Agric* 162:845–854. <https://doi.org/10.1016/j.compag.2019.05.017>
- Ward D, Moghadam P, Hudson N (2018) Deep Leaf Segmentation Using Synthetic Data. *arXiv [cs.CV]*
- Warman C, Panda K, Vejlupekova Z et al (2020) High expression in maize pollen correlates with genetic contributions to pollen fitness as well as with coordinated transcription from neighboring transposable elements. *PLoS Genet* 16:e1008462. <https://doi.org/10.1371/journal.pgen.1008462>
- Warman C, Sullivan CM, Preece J et al (2021) A cost-effective maize ear phenotyping platform enables rapid categorization and quantification of kernels. *The Plant Journal*. <https://doi.org/10.1111/tbj.15166>
- Wolny A, Cerrone L, Vijayan A et al (2020) Accurate and versatile 3D segmentation of plant tissues at cellular resolution. *Elife*. <https://doi.org/10.7554/eLife.57613>
- Yasrab R, Atkinson JA, Wells DM et al (2019) RootNav 2: deep learning for automatic navigation of complex plant root architectures. *Gigascience*. <https://doi.org/10.1093/gigascience/giz123>
- Yazdanbakhsh N, Fisahn J (2009) High throughput phenotyping of root growth dynamics, lateral root formation, root architecture and root hair development enabled by PlaRoM. *Funct Plant Biol* 36:938–946. <https://doi.org/10.1071/FP09167>
- Zhang X, Hause RJ Jr, Borevitz JO (2012) Natural genetic variation for growth and development revealed by high-throughput phenotyping in Arabidopsis Thaliana. *G3* 2:29–34. <https://doi.org/10.1534/g3.111.001487>
- Zhang X, Huang C, Wu D et al (2017) High-throughput phenotyping and QTL mapping reveals the genetic architecture of maize plant growth. *Plant Physiol* 173:1554–1564. <https://doi.org/10.1104/pp.16.01516>
- Zhu Y, Aoun M, Krijn M, Vanschoren J (2018) Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants. *BMVC*
- Ziamtsov I, Navlakha S (2019) Machine learning approaches to improve three basic plant phenotyping tasks using three-dimensional point clouds. *Plant Physiol* 181:1425–1440. <https://doi.org/10.1104/pp.19.00524>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.