A deep learning-based approach for segmenting and counting reproductive organs from digitized herbarium specimen images using refined Mask Scoring R-CNN

Abdelaziz Triki[1], Bassem Bouaziz[1], Jitendra Gaikwad[2] and Walid Mahdi[1]

[1]*MIRACL/CRNS, University of Sfax, Tunisia*
[2]*Friedrich Schiller University Jena, Germany*

Abstract

The accurate segmentation and counting of the reproductive organs within the herbarium specimen play an important role in studying the impact of climate change on plant development over time. Recently, the researchers have gained a lot of knowledge about plant phenology owing to herbaria's digitization efforts, which may help accelerate plant phenology research by making large digitized specimen collections publicly available. Nevertheless, the automatic segmentation and counting of the reproductive organs is a challenging problem. This is because of the high variability of reproductive organs, which vary in size, shape, orientation, and color. The use of machine learning techniques, including deep learning, has recently been shown to be helpful in this endeavor. We proposed in this paper a deep learning method based on the refined Mask Scoring R-CNN approach to segment and count reproductive organs, including buds, flowers, and fruits from specimen images. Our proposed method achieved a precision rate of 94.5% and a recall rate of 93%.

Keywords

Reproductive Organs, Mask Scoring R-CNN, Digitized Herbarium Specimen Images,

Nowadays, the impact of climate change on organisms has received significant attention from scientists. Variations in phenological events have been important where species are failing to react phenologically to climate change, causing serious consequences for long-term diversity preservation [1].

Information about plant phenology found on herbarium specimens, including flowers, buds, and fruits, has recently been proven to be of considerable importance. Currently, herbaria across the globe have collected millions of physical specimen sheets, reducing their use and accessibility by scientists. The gathered specimen sheets include relevant information for addressing research questions such as the flowering period [2].

unisian Algerian Conference on Applied Computing (TACC 2021), December 18–20, 2021, Tabarka, Tunisia
abdelaziz.triki@yahoo.fr (A. Triki); bassem.bouaziz@isims.usf.tn (B. Bouaziz); jitendra.gaikwad@uni-jena.de (J. Gaikwad); walid.mahdi@isims.usf.tn (W. Mahdi)
https://www.researchgate.net/profile/Abdelaziz-Triki (A. Triki);
https://www.researchgate.net/profile/Bassem-Bouaziz-2 (B. Bouaziz);
https://www.researchgate.net/profile/Jitendra-Gaikwad-2 (J. Gaikwad);
https://www.researchgate.net/profile/Walid-Mahdi (W. Mahdi)
0000-0001-5818-2941 (A. Triki); 0000-0002-3692-9482 (B. Bouaziz); 0000-0002-7565-4389 (J. Gaikwad); 0000-0003-3465-0397 (W. Mahdi)

To make millions of stored specimen images more easily accessible, herbaria worldwide have started a massive digitization process [3] of their collections, providing not only access to meta-data, but also to the content of the specimen. Consequently, scientists employed the digitized herbarium collections to track the effects of phenological events on climate change across locations and time [4].

The remainder of this paper is structured as follows. Section 2 presents the related work for segmentation and counting of reproductive organs, while the third section shows our proposed approach. Section 4 discusses the findings of the assessment and experiments we conducted to determine how well our method performed. Section 5 presents the conclusion and future directions.

1. Related Work

Numerous studies using digitized herbarium specimen (DHS) images have focused on a single phenological event such as the flowering period to investigate climate change. Only a few studies have investigated several reproductive organs at the same time (such as measuring the fruit area) and determined how different phenological events are linked [5].

To estimate flowering time, Park et al. [6] have proposed the first initiative, which manipulates the phenological organs, such as whether or not flowers or fruits are present, together with the day of the year of gathering. Nevertheless, the researchers completely overlooked other reproductive organs such as buds, seeds, and roots.

Recently, with the advent of deep learning techniques, several methods based on Convolutional Neural Networks (CNN) have been developed to assess the reactiveness of different points in a given phenophase. The study developed by Lorieul et al. [7] showed that a deep learning method based on CNN could detect the presence/absence of fruits or flowers within DHS images and overlook the buds. However, the detection accuracy of flowers and fruits was low ($\approx 85\%$ and $\approx 80\%$ accuracy, respectively). Besides, Ellwood et al. [8] established a new method for classifying reproductive organs into two major groups: fertile and sterile specimens. However, the proposed approach is unable to segment and count the reproductive organs accurately within the DHS images.

On the other hand, Goëau et al. [9] developed a deep learning-based approach using Mask R-CNN [10] to segment and count the reproductive organs of a small dataset containing one species (*Streptanthus tortuosus*). The testing findings showed that flowers within DHS images are more consistently recognized and counted than fruits. To overcome the limitations of Goëau's approach, Davis et al. [11] have built a deep learning model to cope with a higher number of species. The developed solution locates and counts the reproductive organs of buds, flowers, and fruits within the specimen sheets. Also, they used 3000 specimens of five common wildflower species from the eastern United States of America to train their model. According to the experimental results, the counting was accurate, but it differed depending on the reproductive organs. Here, the counting of flower organs is less accurate than counting buds and fruits due to the morphological variability of flowers within the scanned specimens.

In this paper, we developed a deep learning-based approach for segmenting and counting reproductive organs, including buds, fruits, and flowers, from DHS images using refined Mask

Scoring R-CNN [12]. Our method generates a boundary mask for each segmented object and classifies it based on that mask. Furthermore, our suggested solution was trained and evaluated using a dataset of DHS images collected from the Herbarium Haussknecht of Jena (https://www.herbarium.uni-jena.de/).

2. Specimen Dataset Collection

We used in this paper several species gathered from the herbarium Haussknecht of Jena, Germany, to train and test our proposed approach. This dataset comprises 4000 DHS images containing different families of species with obvious phenotypic changes during the reproductive cycle to prevent singularity.

The collected specimens were resized by changing their sizes from 6400×3400 pixels to 2048×1024 pixels for height and width, respectively. It includes objects of different sizes and shapes. The resizing process minimizes unnecessary memory usage and maintains the aspect ratio of the scanned herbarium sheets. Moreover, the VGG Image Annotator (VIA) annotation tool [13] was used to annotate each specimen by hand, and the resulting JSON file represents the bounding mask for each reproductive organ inside the scanned specimen images. Each class has a bounding polygon to indicate its bounding mask due to its uneven form.

On the other hand, this study splits the dataset into two halves at random: 70% of the dataset was used to train our approach, and 30% of the dataset was used to perform the test phase.

3. Proposed Approach

The quantitative analysis of reproductive organs within DHS images, including buds, flowers, and fruits, will substantially enhance the use of specimen collections in climate change research. Given the massive number of DHS images stored in the Haussknecht herbarium in Jena and the diversity of reproductive organs introduced within the herbarium scans (i.e., scales, orientations, shapes, and colors), we proposed a deep learning-based approach for counting these reproductive organs and improving their segmentation accuracy. The developed approach is a refined version of the state-of-the-art instance segmentation technique, Mask Scoring RCNN [12], in which the backbone model was altered by proposing a new Feature Pyramid Networks (FPN) network.

3.1. Mask Scoring RCNN

Mask Scoring RCNN [12] is an improved variant of Mask-RCNN [10] that improves instance segmentation accuracy. In comparison to Faster R-CNN [14], Mask R-CNN is more efficient in terms of time and accuracy since it supports pixel-level instance segmentation. Additionally, Mask-RCNN utilizes classification confidence as the mask's consistency metric. Nevertheless, mask accuracy is often uncorrelated with classification confidence, resulting in sub-optimal precision and durability of predicted masks. Mask Scoring RCNN learns and predicts mask consistency using a specific network block to overcome the cited problems. This block contains the instance-specific features and the expected mask. Additionally, it measures the intersection-over-union (IoU) between the expected mask and the ground truth. The IoU is employed to determine the predicted mask's shape. This enables Mask Scoring RCNN to get much better instance segmentation outcomes than Mask-RCNN.

On the other hand, Mask Scoring RCNN model is divided into three parts. The first part of the approach shows a standard convolutional layer for extracting fine features. As a result, the feature maps produced will be sent into the Region Proposal Network (RPN), which will create and correct the region proposals. The last part of the model is used to fine-tune the bounding boxes and the resulting masks.

3.2. Feature Extraction

For image feature extractions, deep convolutional networks with different weight layers are frequently employed. With more convolution layers, the training error may rise, and classification accuracy may fall. ResNet effectively tackled this issue by figuring out how to represent residuals between inputs and outputs. By distinguishing between various types of objects, ResNet speeds up training time and improves the accuracy of predictions.

The conventional Mask Scoring RCNN model uses ResNet [15] in conjunction with FPN [16] as the backbone model to extract the image features. However, extracting the reproductive organs from the DHS image using this backbone may produce irrelevant predictions with lower performance. This is due to several scale variations that may affect how well reproductive organs can be extracted from DHS images.

When it comes to object instance segmentation using Mask Scoring R-CNN, one of its weaknesses is scale variation. Due to the wide variety of objects' sizes and shapes within the DHS images, detectors have difficulty segmenting the most miniature objects such as buds. To remedy the existing method problems, we present in this paper a refined FPN network to enhance the backbone structure of the Mask Scoring R-CNN network model.

The conventional FPN of Mask Scoring R-CNN comprises bottom-up and top-down paths to form a two-channel feature extraction path model. As a result, the high-level features may be increased, and all features in FPN with a decent classification ability will be improved. Nevertheless, the conventional two-channel path does not provide sufficient coverage to collect semantic feature data from the pyramid network's bottom layer and the upper layer of the network.

To resolve the issues involved with the conventional two-channel pathways of FPN, we developed a three-channel FPN pathway. A more significant number of paths are available to acquire the features, as illustrated in Figure 1, resulting in a greater improvement in segmentation accuracy and reducing the training time. By adding the third channel path, the model will be updated as follows:

- A 3×3 convolution kernel with a step size of 3 was conducted by the layer a to reduce the spatial size.
- To create the merged feature layer, we performed a lateral join on the second channel feature layer x.
- As with the second step, add layers one at a time until achieving the layer d and then stop. The final a, b, c layers represent the third channel feature extraction layers added from the bottom up path.

Figure 1 represents the refined FPN structure, which facilitates continuous information propagation and improves the segmentation performance.

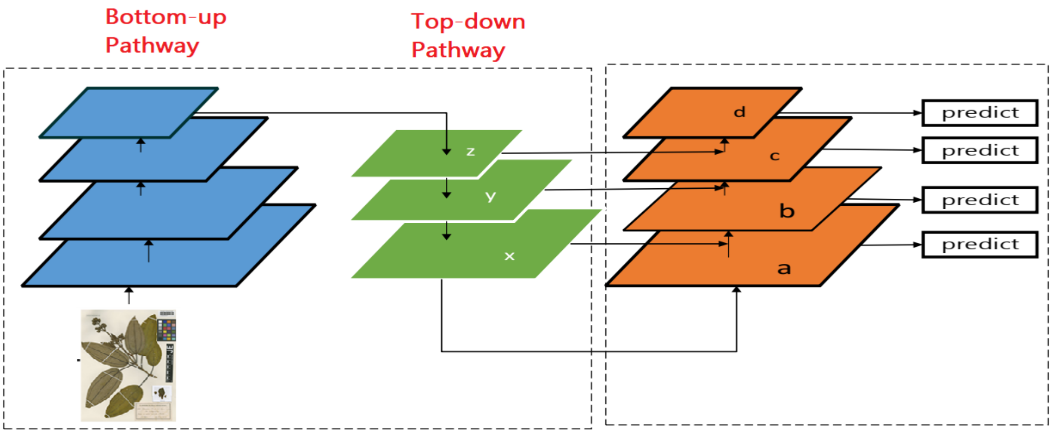

Figure 1: Refined FPN structure

4. Experimental Results

In this paper, we implemented our proposed approach using Facebook's Open Source Maskrcnn-Benchmark project (https://github.com/facebookresearch/maskrcnn-benchmark). We also employed Google Collaboratory [17] to perform the training and testing phases. Besides, our approach was trained for 100 epochs with 300 steps for each epoch, while the stochastic gradient descent (SGD) was initialized with a learning rate of 0.001, a momentum of 0.9, and a weight decay of 0.0005.

During the test phase, we selected randomly 300 DHS images of five common species (*Aplopappus stoloniferous*, DC., var; *Eupatorium scordonioides* A. Gray; *Penstemon pulchellus Lindl*; *Senecio Chapalensis*, Watson and *Russelia trachypleura*, Rob). The reproductive organs on each herbarium sheet vary in size and position. We identified in this paper 1450 reproductive organs to be used as ground truth data for this research, comprising 262 buds, 483 fruits, and 705 flowers.

To examine the model's consistency and measure the segmentation performance, we measured the Precision (Eq. 1), Recall (Eq. 2) and Average Precision (AP) metrics for our method, the original Mask Scoring-RCNN and the Mask-RCNN, using various backbone networks of the ResNet-50/101 architectures.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{1}$$

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives} \tag{2}$$

As shown in Table 1, our proposed method had an overall precision and recall of 88.9% and

Table 1
The precision/recall of our approach

Metrics	Buds	Fruits	Flowers	Overall
Precision	86%	89.2%	91.6%	88.9%
Recall	88.5%	91.7%	87.1%	89.1%

89.1%, respectively. A total of 19 buds, for example, were mistakenly labeled as flowers owing to the overlapping of various organs within the DHS images.

According to the segmentation findings in Table 1, our method can correctly identify the presence or absence of reproductive organs within the DHS images. Segmenting smaller objects (such as buds) with varying occlusion levels is also feasible. For flower segmentation, our architecture performs better than other reproductive organs. It achieved a precision of 91.6% and a recall rate of 87.1%.

4.1. The impact of data augmentation approaches on segmentation results

Over the last few years, object segmentation has shown impressive results using deep learning methods. However, the scarcity of data is seen as a significant drawback since large quantities of labeled datasets are required to improve model accuracy and prevent overfitting. As a workaround for the absence of large annotated datasets, we performed several data augmentation strategies, including rotation, translation, scale, and brightness [4]. Additionally, the goal of this study is to determine the effect of different image augmentation techniques on the segmentation of reproductive organs.

To examine our model's stability while counting the reproductive organs, we created the confusion matrix (Table 2). This matrix illustrates our approach's reliability by comparing the actual count and the predicted count of different reproductive organs within the DHS images.

As shown in Tables 2 and 3, when raising the amount of training data using data augmentation methods, our model's performance improved significantly. As a result, the overall precision of our approach was increased by 5.6%, while the overall recall was increased by 3.8%, respectively. This demonstrates that the proposed method with data augmentation may enhance the model's capacity to conduct instance segmentation.

4.2. The counting evaluation

To evaluate the counting capability of our proposed approach, we measured the *Mean Absolute Error (MAE)* for each reproductive organ within the DHS images. The MAE is a common metric used to evaluate the model's accuracy. It measures the average absolute value of the counting error of each reproductive organ as follows (Eq. 3):

$$MAE = \frac{1}{n}\sum_{n}^{i=1} |x_i - x| \tag{3}$$

Where *n* is the number of errors.

Table 2
The confusion matrix of our approach when applying the data augmentation approaches using different species

Species Names	No. of DHS	No. of Buds		No. of Fruits		No. of Flowers	
Aplopappus stoloniferous	54 ,	**Ground Truth**	53	**Ground Truth**	97	**Ground Truth**	142
		Detected	49	**Detected**	94	**Detected**	139
Eupatorium scordonioides	46	**Ground Truth**	49	**Ground Truth**	88	**Ground Truth**	134
		Detected	44	**Detected**	82	**Detected**	130
Penstemon pulchellus	65	**Ground Truth**	46	**Ground Truth**	89	**Ground Truth**	127
		Detected	39	**Detected**	80	**Detected**	120
Senecio Chapalensis	72	**Ground Truth**	51	**Ground Truth**	101	**Ground Truth**	140
		Detected	43	**Detected**	92	**Detected**	136
Russelia trachypleura	63	**Ground Truth**	63	**Ground Truth**	108	**Ground Truth**	162
		Detected	56	**Detected**	103	**Detected**	158
Total	300	**Ground Truth**	262	**Ground Truth**	483	**Ground Truth**	705
		Detected	228	**Detected**	451	**Detected**	683

Table 3
The precision/recall of our approach when using the data augmentations

Metrics	Buds	Fruits	Flowers	Overall
Precision	93.8%	96.1%	93.6%	94.5%
Recall	87%	93.3%	98.2%	93%

Table 4
Predicted and true counts of reproductive organs, including buds, flowers, and fruits

	Buds	Fruits	Flowers	All
True Number of reproductive organs	262	483	705	1450
Predicted Number of reproductive organs	228	451	683	1362
Mean Absolute Error (MAE)	0.34	0.32	0.22	0.8

As shown in Table 4, the MAE value of our approach was very low for all kinds of reproductive organs, where the lowest value was obtained by the flower organs.

4.3. Experimental results and evaluation of multiple segmentation approaches

To assess the efficacy of our proposed approach in segmenting the reproductive organs from the herbarium sheets, we compared it to other state-of-the-art approaches with various backbone models. Table 5 presents the findings of the assessment. By employing ResNet101 as our backbone network, our method outperformed Mask Scoring RCNN and Mask-RCNN when identifying reproductive organs.

Using the original Mask Scoring RCNN approach, many reproductive organs were incorrectly segmented, resulting in lower overall results (the achieved overall precision and recall were 84% , and 86% , respectively). Furthermore, as shown in Figure 2, our proposed model's reliability

Table 5
The comparison results of our approach using different ResNet50/101 backbones

Models	**Precison**	**Recall**
Original Mask Scoring RCNN with ResNet50	86.3%	88%
Original Mask Scoring RCNN with ResNet101	84%	86%
Mask-RCNN with ResNet50	84%	82%
Mask-RCNN with ResNet101	85%	83%
Our approach with ResNet50	90.6%	91.3%
Our approach with ResNet101	**94.5%**	**93%**

may be confirmed. This curve gives additional insight into the proposed approach. Also, Figure 3 shows that the loss remains constant after 100 epochs, whereas the model training takes around 21 hours. As a result, when ResNet101 is used as a backbone network, the total loss was minimized while accuracy increased during the training phase.

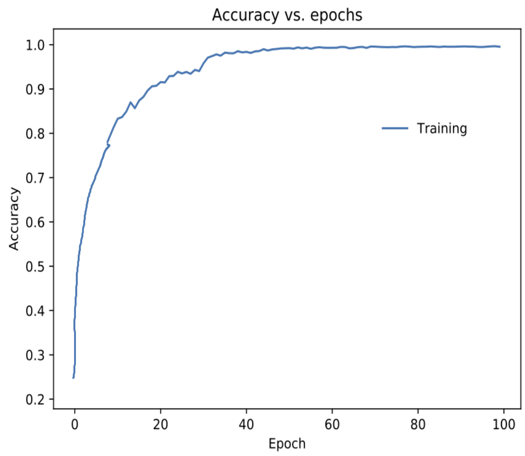

Figure 2: Accuracy curve of our proposed approach

5. Discussion

We present in this paper a new deep learning based-approach for segmenting and counting the reproductive organs such as buds, flowers, and fruits from the DHS images. For the sake of this research, 4000 different herbarium images containing reproductive organs were manually annotated. However, several factors impact the performance of our proposed approach, including the quantity of the training dataset and the high variability of reproductive organs within the scanned herbarium, which vary in size, shape, orientation, and color.

In this paper, we investigate the effect of organ size on the segmentation of reproductive organs within the herbarium sheets, where smaller objects, such as buds, perform worse than

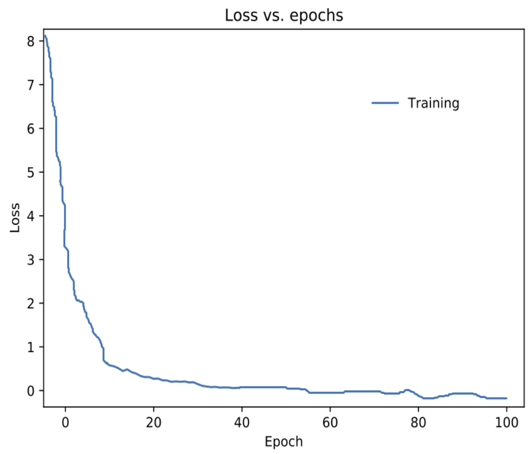

Figure 3: Loss curve of our proposed approach

larger ones. Buds are performing worse for a variety of reasons. One example is the small number of bud organs found in the DHS images, where about 70% of the samples were devoid of them. Flowers and fruits have higher visual distinctiveness, whereas buds are smaller and have lower visual distinctiveness. As a result, flowers seem to be the most segmented organ, followed by fruits and buds, because they are more significant than buds and they are not as widely distributed as fruits and buds (Figure 4).

Further analysis of counting results presented in the confusion matrix (Table 2) and Table 4 reveal that the overall number of counted reproductive organs using our approach was quite close to the actual value (ground truth). For example, the lowest MEA was for the flower counts due to the morphological uniformity and abundance of flowers within the herbarium sheets. As demonstrated in Table 2, the total number of counted flower organs was 683, while their true value was 705 organs (MAE = 0.22). As a result of their irregular shape and scarcity within the DHS images, MAE for counting buds was considerably poorer than for flowers or fruits, as previously mentioned (MAE = 0.34).

By comparing our approach with other state-of-the-art techniques, we found that our method performed well for segmenting and counting the reproductive organs within the DHS images (Table 5). The precision and recall reached 94.5% and 93%, respectively. However, additional research is needed to determine if variations in the appearance of reproductive organs among dried and pressed specimens complicate the automated segmentation of phenological features.

6. Conclusion and Future Directions

For the sake of segmenting and counting the reproductive organs within the DHS images, we presented in this paper a deep learning-based approach, where the FPN network was refined by creating three-channel pathways FPN. Our method is focused on precisely segmenting different-sized and oriented buds, fruits, and flowers within the DHS images. Then, we predicted the number of reproductive organs by counting their occurrence within the DHS images. Based on

Figure 4: Our proposed approach Outputs

the findings, our approach can segment reproductive organs with high precision and recall.

It is important to note that the success of our method is also dependent on a relatively training dataset. Enriching the dataset with specimens from various spatio-temporal scales may thus significantly improve performance.

Acknowledgments

This work was part of the MAMUDS project (Management Multimedia Data for Science), supported by BMBF, Germany (Project No. 01D16009) and MHESR, Tunisia.

References

[1] S. Piao, Q. Liu, A. Chen, I. Janssens, Y. H. Fu, J. Dai, L. Liu, X. Lian, M. Shen, X. Zhu, Plant phenology and global climate change: Current progresses and challenges., Global change biology 25 6 (2019) 1922–1940.

[2] T. Borsch, A. Stevens, E. Häffner, A. Güntsch, W. Berendsohn, M. Appelhans, C. Barilaro, B. Beszteri, F. Blattner, O. Bossdorf, H. Dalitz, S. Dressler, R. Duque-Thüs, H. Esser, A. Franzke, D. Goetze, M. Grein, U. Grünert, F. Hellwig, J. Hentschel, E. Hörandl, T. Janssen, N. Jürgens, G. Kadereit, T. Karisch, M. Koch, F. Müller, J. Müller, D. Ober, S. Porembski, P. Poschlod, C. Printzen, M. Röser, P. Sack, P. Schlüter, M. Schmidt, M. Schnittler, M. Scholler, M. Schultz, E. Seeber, J. Simmel, M. Stiller, M. Thiv, H. Thüs, N. Tkach, D. Triebel, U. Warnke, T. Weibulat, K. Wesche, A. M. Yurkov, G. Zizka, A complete digitization of german herbaria is possible, sensible and should be started now, Research Ideas and Outcomes 6 (2020).

[3] P. W. Sweeney, B. Starly, P. J. Morris, Y. Xu, A. Jones, S. Radhakrishnan, C. Grassa, C. Davis, Large-scale digitization of herbarium specimens: Development and usage of an automated, high-throughput conveyor system, Taxon (2018).

[4] D. F. Lima, J. H. F. Mello, I. T. Lopes, R. Forzza, R. Goldenberg, L. Freitas, Phenological responses to climate change based on a hundred years of herbarium collections of tropical melastomataceae, PLoS ONE 16 (2021).

[5] C. G. Willis, E. R. Ellwood, R. B. Primack, C. C. Davis, K. D. Pearson, A. S. Gallinat, J. M. Yost, G. Nelson, S. J. Mazer, N. L. Rossington, T. H. Sparks, P. S. Soltis, Old plants, new tricks: Phenological research using herbarium specimens, Trends in Ecology and Evolution 32 (2017) 531–546. URL: https://www.sciencedirect.com/science/article/pii/S0169534717300939. doi:https://doi.org/10.1016/j.tree.2017.03.015.

[6] D. Park, A. Williams, E. Law, A. Ellison, C. Davis, Assessing plant phenological patterns in the eastern united states over the last 120 years, Environmental Data Initiative (2018).

[7] T. Lorieul, K. D. Pearson, E. R. Ellwood, H. Goëau, J. Molino, P. W. Sweeney, J. Yost, J. Sachs, E. Mata-Montero, G. Nelson, P. Soltis, P. Bonnet, A. Joly, Toward a large-scale and deep phenological stage annotation of herbarium specimens: Case studies from temperate, tropical, and equatorial floras, Applications in Plant Sciences 7 (2019).

[8] E. R. Ellwood, K. D. Pearson, G. Nelson, Emerging frontiers in phenological research, Applications in Plant Sciences 7 (2019).

[9] H. Goëau, A. Mora-Fallas, J. Champ, N. Love, S. Mazer, E. Mata-Montero, A. Joly, P. Bonnet, A new fine-grained method for automated visual analysis of herbarium specimens: A case study for phenological data extraction, Applications in Plant Sciences 8 (2020).

[10] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask r-cnn, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988. doi:10.1109/ICCV.2017.322.

[11] C. Davis, J. Champ, D. S. Park, I. K. Breckheimer, G. M. Lyra, J. Xie, A. Joly, D. Tarapore, A. Ellison, P. Bonnet, A new method for counting reproductive structures in digitized herbarium specimens using mask r-cnn, Frontiers in Plant Science 11 (2020).

[12] Z. Huang, L. Huang, Y. Gong, C. Huang, X. Wang, Mask scoring r-cnn, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 6402–6411.

[13] A. Dutta, A. Zisserman, The via annotation software for images, audio and video, Proceedings of the 27th ACM International Conference on Multimedia (2019).

[14] S. Ren, K. He, R. B. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2015) 1137–1149.

[15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 770–778.

[16] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, S. J. Belongie, Feature pyramid networks for object detection, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 936–944.

[17] C. Shorten, T. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data 6 (2019) 1–48.