Dynamic Color Transform for Wheat Head Detection

Chengxin Liu Kewei Wang Hao Lu Zhiguo Cao*

Key Laboratory of Image Processing and Intelligent Control, Ministry of Education
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology

{cx_liu,zgcao}@hust.edu.cn

Abstract

Developing accurate algorithms for wheat head detection is challenging due to the variability of observation circumstances and the uncertainty of wheat head appearances. In this work, we propose a simple but effective idea—dynamic color transform (DCT)—for accurate wheat head detection. This idea is based on an observation that modifying the color channel of an input image can significantly alleviate false negatives and therefore improve detection results. DCT follows a linear color transform and can be easily implemented as a dynamic network. A key property of DCT is that the transform parameters are data-dependent such that illumination variations can be corrected adaptively. The DCT network can be incorporated into any existing object detectors. For example, DCT plays an important role in our solution participating in the Global Wheat Head Detection (GWHD) Challenge 2021, where our solution ranks the first on the initial public leaderboard, with an Average Domain Accuracy (ADA) of 0.821*, and obtains the runner-up reward on the final complete testing set, with an ADA of* 0.695*.*

1. Introduction

With the prevalence of affordable camera platforms (*e.g.*, unmanned aerial vehicles and smartphones), in-field imagery has become a convenient image acquisition choice for monitoring the characteristics of wheat [4, 5]. One of the important tasks is wheat head detection. It enables automatic measurements of wheat traits, such as head population, wheat maturity stage, and wheat size. By deploying wheat head detection, tedious manual measurements can be avoided, and farmers can make fast decisions based on automated observation.

*Corresponding author

While many methods have been developed for generic object detection [1, 2, 12, 13, 14, 15], these methods are not directly applicable to in-filed wheat head detection. In contrast to images captured in the natural context, in-filed wheat head images exhibit visual challenges from two aspects: first, different observation circumstances affect the quality of images and wheat head appearances, *e.g.*, wind can result in blurry images, atmospheric light can lead to unbalanced image contrast, and the observation conditions also affect wheat head orientation; second, the phenotypes of wheat head vary significantly under different growth stages, *e.g.*, the color of the spike is green at the post-flowering stage, but turns yellow at the ripening stage.

Recently, much effort has been made to crop detection [3, 6, 11, 16]. Ghosal *et al*. [6] proposes a weakly-supervised framework for sorghum head detection. Chandra *et al*. [3] tackles crop detection via point supervision based active learning. On the other hand, Zou *et al*. [16] presents a comprehensive study on maize tassels detection. Different from previous studies that aim to reduce labeling burdens, we focus on developing high-performance detectors for wheat head detection.

In this work, we find that appropriate treatment of color cues can greatly benefit wheat head detection, particularly in alleviating false negatives. Specifically, we present an analysis on the impact of the color channel and propose to deal with colors with dynamic color transform (DCT). The DCT is in the same spirit of recent dynamic networks [9, 10] that enable date-dependent inference. For example, the DCT follows a linear color model that dynamically generates 6 parameters to modulate the color of the input image. Our main contributions include the following:

- We investigate the impact of the color channel and observe that modifying the color channel of the input image can improve detection results;
- We introduce a DCT network based on our observation, which improves wheat head detection;

- Our DCT network achieves the runner-up performance on the Global Wheat Head Detection Challenge 2021.

2. Impact of the Color Channel

Color is an important attribute of in-filed wheat head images. However, the color information is often ignored in existing object detectors. Here we empirically investigate the impact of the color channel on wheat head detection, which lays the foundation of our approach. In the following, we first introduce the baseline object detector—Scaled-YOLOv4 [15]. Then, we demonstrate the impact of the color channel on this model.

Baseline Object Detector. We adopt a state-of-the-art object detector—Scaled-YOLOv4 [15] as our baseline. Scaled-YOLOv4 proposes a network scaling method that can modify the depth, width, resolution, and structure of the detection network, thus maintaining the balance between speed and accuracy. The reasons why we choose Scaled-YOLOv4 include:

1) it reports strong performance on generic object detection;

2) it is clean to enable flexible modifications.

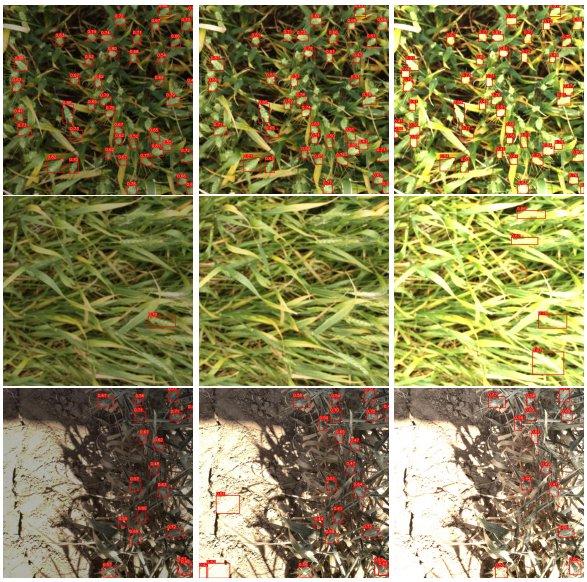
Color Channel Modification. Specifically, given an object detector trained on the GWHD [5] dataset (*e.g.*, we adopt Scaled-YOLOv4 [15]), we manually modify the value of each color channel and compare the detection results under different color conditions. In particular, we use a linear color transform to adjust the color channel as follows:

$$\begin{cases} \boldsymbol{R}' = \alpha \boldsymbol{R} + \beta \\ \boldsymbol{G}' = \alpha \boldsymbol{G} + \beta \\ \boldsymbol{B}' = \alpha \boldsymbol{B} + \beta \end{cases}, \tag{1}$$

where $\boldsymbol{R}$, $\boldsymbol{G}$, and $\boldsymbol{B}$ denote the red, green, and blue color channels of an image, respectively. $\boldsymbol{R}'$, $\boldsymbol{G}'$, and $\boldsymbol{B}'$ are transformed color channels, and α and β are tunable parameters.

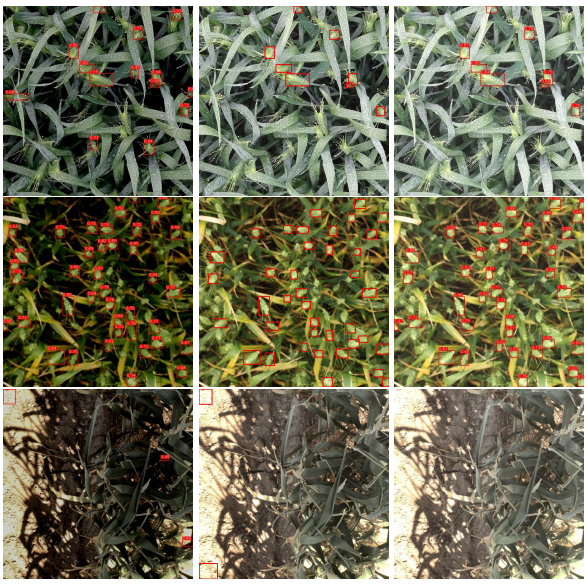
To investigate the impact of the color channel, we separately modify the values of α and β. Specifically, we first fix $\beta = 0$ and vary α ($\alpha \in \{0.7, 1.0, 1.5\}$). The qualitative results are shown in Fig. 1. Note that when $\alpha = 1.0$ and $\beta = 0$, the transformed image is the same as the original image. Interestingly, we observe that modifying α can improve the detection results. For instance, false negatives are alleviated in the first and second rows. In addition, false positives are suppressed in the third row. Next, we fix $\alpha = 1.0$ and vary β ($\beta \in \{-50, 0, 20\}$). Fig. 2 shows the qualitative results. Similarly, modifying the value of β can also improve detection.

Moreover, we also compare the detection performance of Scaled-YOLOv4 under different α's and β's on the GWHD 2021 test set. Table 1 illustrates the detailed results, where $\alpha = 1.0$ and $\beta = 0$ denote the baseline, and ADA is the



(a) $\alpha = 0.7$ (b) $\alpha = 1.0$ (c) $\alpha = 1.5$

Figure 1: Qualitative results of Scaled-YOLOv4 [15] under different α's ($\alpha = 1.0$ denotes original image), where β is fixed to 0. The numbers above the red detection boxes are the confidence scores. Best viewed by zooming in.



(a) $\beta = -50$ (b) $\beta = 0.0$ (c) $\beta = 20$

Figure 2: Qualitative results of Scaled-YOLOv4 [15] under different β's ($\beta = 0$ denotes original image), where α is fixed to 1.0. The numbers above the red detection boxes are the confidence scores. Best viewed by zooming in.

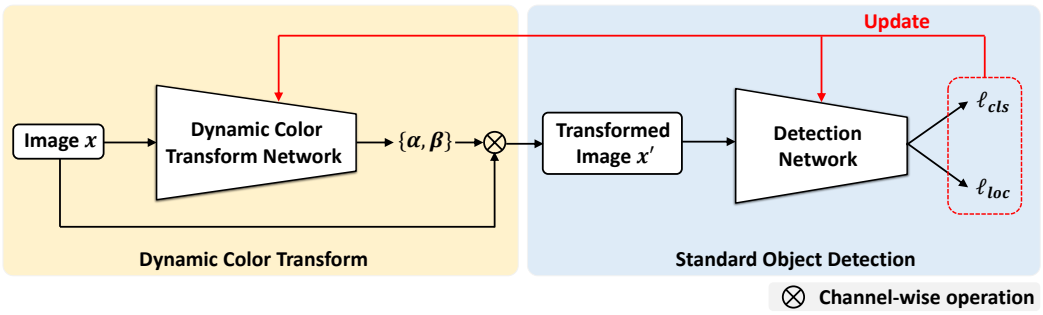


Figure 3: An overview of our method. The input image $\boldsymbol{x}$ is first transformed to $\boldsymbol{x'}$ by DCT network, then $\boldsymbol{x'}$ is sent to detection network for calculating loss ℓ_{cls} and ℓ_{loc}, which are used to update the DCT and the detection network.

Table 1: Quantitative results of Scaled-YOLOv4 under different α's and β's on the GWHD 2021 test set. The evaluation metric is ADA.

Type	α	β	Test ADA
Varying α	0.7	0	0.652
	1.0	0	0.641
	1.5	0	0.614
Varying β	1.0	-50	0.618
	1.0	0	0.641
	1.0	20	0.644

evaluation metric (see Sec. 4.1 for details). We notice that an appropriate choice of α and β can indeed improve detection. For example, setting $\alpha = 0.7$ improves the ADA from 0.641 to 0.652. Our results in Table 1 are consistent with the observation in Fig. 1 and Fig. 2.

To summarize, our results indicate that color is an important clue in wheat head detection, which motivates us to leverage the color information. However, we remark that, despite color is useful, it is not sufficient to tackle object detection based on colors solely. The reasons are two-fold:

1) since wheat heads vary significantly in different domains, color information is not shared among different areas;

2) color is sensitive to observation/illumination conditions, thus color distortions may occur when perturbation appears.

Therefore, we relieve the role of the color and incorporate color information into existing object detectors to improve detection.

3. Dynamic Color Transform Network

We first give an overview of our method. Then, we explain the details of color transform modeling before we introduce the loss functions and implementation details.

3.1. Overview

Motivated by the observation that modifying color channels can improve detection results (Sec. 2), we propose a DCT network for accurate wheat head detection. An overview of our method is depicted in Fig. 3. Specifically, we first pass the input image $\boldsymbol{x}$ through the DCT network to obtain the transformed image $\boldsymbol{x'}$. Then, we perform standard object detection to compute the loss, which is used to update the DCT and the detection network.

3.2. Color Transform Modeling

Given an input RGB image $\boldsymbol{x}$, we adopt a linear function to model color transform

$$\begin{cases} \boldsymbol{R'} = \alpha_R \boldsymbol{R} + \beta_R \\ \boldsymbol{G'} = \alpha_G \boldsymbol{G} + \beta_G \\ \boldsymbol{B'} = \alpha_B \boldsymbol{B} + \beta_B \end{cases}, \tag{2}$$

where $\boldsymbol{R}$, $\boldsymbol{G}$, and $\boldsymbol{B}$ denote the red, green, and blue color channels of the input image $\boldsymbol{x}$, respectively. $\boldsymbol{R'}$, $\boldsymbol{G'}$, and $\boldsymbol{B'}$ are transformed color channels. α_R, α_G, α_B, β_R, β_G, and β_B are predicted color transform parameters. Note that we model the transform of each color channel independently.

Formally, a DCT network ϕ parameterized by θ is applied to the input image $\boldsymbol{x}$, predicting color transform parameters $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ by

$$\{\boldsymbol{\alpha}, \boldsymbol{\beta}\} = \phi_\theta(\boldsymbol{x}), \tag{3}$$

where $\boldsymbol{\alpha} = [\alpha_R, \alpha_G, \alpha_B]$ and $\boldsymbol{\beta} = [\beta_R, \beta_G, \beta_B]$. The transformed input image $\boldsymbol{x'}$ can be written as

$$\boldsymbol{x'} = \boldsymbol{\alpha} \cdot \boldsymbol{x} + \boldsymbol{\beta}, \tag{4}$$

where $\cdot$ denotes channel-wise multiplication.

Practically, DCT can be easily implemented as a dynamic network [9, 10]. In particular, any off-the-shelf networks can be adopted as the DCT network ϕ, *e.g.*, ResNet [7]. Note that, the structure of the DCT network is not limited to existing networks, and a few convolution layers may also work.

Table 2: Final and partial leaderboard of the Global Wheat Head Detection Challenge 2021.

Final Leaderboard			Partial Leaderboard		
Rank	Participants	ADA	Rank	Participants	ADA
1	randomTeamName	0.700	1	**SMART**	0.821
2	**SMART**	0.695	2	kosung	0.812
2	david_jeon	0.695	3	wheat_hunters	0.811
4	keyhan_najafian	0.692	4	randomTeamName	0.807
5	hitsz	0.689	4	david_jeon	0.807
6	maxim	0.682	6	hitsz	0.805
7	kosung	0.676	7	augly_wheat	0.792
8	augly_wheat	0.671	8	Wu_Chun_Huan_	0.790
9	Wu_Chun_Huan_	0.669	9	UoL	0.787
10	Ural	0.666	10	vlad_barbu	0.786

3.3. Loss Function

Given an object detector f parameterized by ω and the transformed input image $\boldsymbol{x'}$, the training loss is formulated as:

$$\min_{\theta,\omega} \mathcal{L}(f_\omega(\boldsymbol{x'}), \{y_i, \boldsymbol{b}_i\}) \,, \tag{5}$$

where $\{y_i, \boldsymbol{b}_i\}$ is the ground-truth label (y_i is the class label and $\boldsymbol{b}_i$ is the bounding box). In practice, $\mathcal{L}$ is composed of classification loss and localization loss [1, 15]. Thus, Eq. (5) can be rewritten as follows:

$$\min_{\theta,\omega} \ell_{cls}(f_\omega(\boldsymbol{x'}), \{y_i, \boldsymbol{b}_i\}) + \ell_{loc}(f_\omega(\boldsymbol{x'}), \{y_i, \boldsymbol{b}_i\}) \,. \tag{6}$$

where ℓ_{cls} and ℓ_{loc} are classification loss and localization loss, respectively.

It is worth mentioning that our DCT network is not limited to specific object detectors. Here we only instantiate an application of the DCT network on Scaled-YOLOv4 [15].

3.4. Implementation Details

The Structure of the DCT Network. We adopt ResNet18 as the DCT Network, where the output channels of ResNet18 are set to 6 (*i.e.*, α_R, α_G, α_B, β_R, β_G, and β_B). The parameters of the DCT network are $11.2M$. We remark that the architecture of the DCT network is not limited to existing networks. One may design their own DCT network.

Training Details. Following [15], the model is trained for 300 epochs. The learning rate of the DCT network and Scaled-YOLOv4 are set to 0.01 and 0.1, respectively. We adopt Stochastic Gradient Descent (SGD) as the optimizer.

Testing. To further improve the detection performance, we adopt model ensemble and pseudo labeling [8] during testing. We apply model ensemble on a set of predictions, where predictions are obtained by test time augmentation (*i.e.*, up-down flip, left-right flip, and rotation). About pseudo labeling [8], we retrain the model with a fusion of the training and testing data, where the predictions of our model are treated as pseudo labels on the test set.

4. Results and Discussion

4.1. Experiment Setup

Dataset. The Global Wheat Head Dataset 2021 [5] is used by the Global Wheat Head Detection Challenge 2021[1]. It contains $3.6K$ training images, $1.4K$ validation images, and $1.3K$ test images. Note that the validation set and the test set correspond to the partial leaderboard and the final leaderboard, respectively.

Evaluation Metric. We use Average Domain Accuracy (ADA) as the evaluation metric. The accuracy of each image is calculated by:

$$\texttt{Accuracy}_{\texttt{image}} = \frac{\texttt{TP}}{\texttt{TP} + \texttt{FN} + \texttt{FP}} \,, \tag{7}$$

where TP, FN, and FP are true positive, false negative, and false positive, respectively. A ground-truth box is considered to match with one predicted box if their Intersection over Union (IoU) is higher than a threshold of 0.5. The accuracy of all images from the same domain is averaged to obtain the domain accuracy. The ADA is the average of all domain accuracy.

4.2. Quantitative Results on the GWHD Dataset

Table 2 shows the competition results, the username of our team is SMART. We ranks second in the final leaderboard of the Global Wheat Challenge 2021, with an ADA of 0.695. In addition, we rank first in the partial leaderboard

[1] https://www.aicrowd.com/challenges/global-wheat-challenge-2021

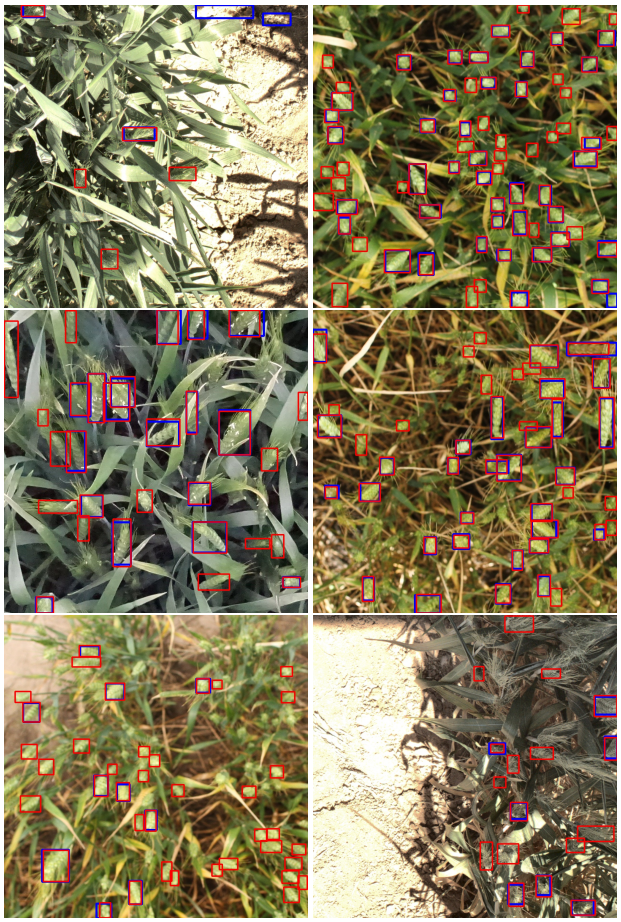


Figure 4: Visualization of detection results. Red boxes are the results our DCT Scaled-YOLOv4, while blue boxes are the results of the standard Scaled-YOLOv4 (without DCT).

Table 3: Ablation Study of our DCT on Scaled-YOLOv4.

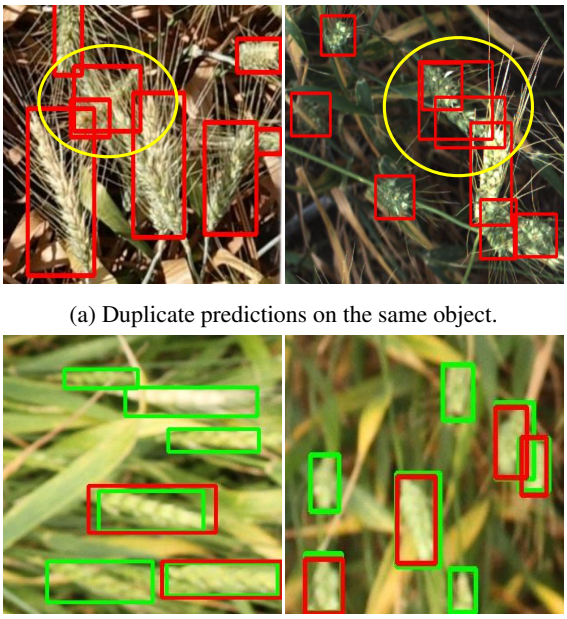
Method	Test ADA
Baseline	0.641
Baseline + DCT	**0.661**

(*i.e.*, initial public leaderboard), with an ADA of 0.821. Note that we only show the results of the top 10 teams, we refer readers to the leaderboard page[2] for full results.

4.3. Ablation Study

Table 3 shows the comparison results of standard Scaled-YOLOv4 and DCT Scaled-YOLOv4. Our DCT boosts the baseline from 0.641 to 0.661, which validates the effectiveness of our method. To understand the impact of DCT, we further visualize the detection results in Fig. 4. Our DCT model is robust to various illumination conditions and performs consistently better than standard Scaled-YOLOv4. For instance, we significantly reduce the number of false negatives. In addition, our model is capable of suppressing false positives in the bright area.

[2]https://www.aicrowd.com/challenges/global-wheat-challenge-2021/leaderboards



(a) Duplicate predictions on the same object.

(b) Missing detections on blurred images.

Figure 5: Failure cases of our DCT on the GWHD 2021 test set. The predictions are in red, while the ground-truth boxes are in green.

4.4. Failure Case Analysis

Although our method achieves promising results on the GWHD 2021 dataset, some limitations exist. Fig. 5 illustrates the failure cases of our DCT. First, our model tends to predict duplicate boxes on the same object, which leads to false positives. Second, the blurred image may render detection failure, *i.e.*, missing detections. We may incorporate the attributes of the wheat head objects (*e.g.*, the relationship between objects) to address the above problems.

5. Conclusion

In this work, we introduce a simple but effective idea — dynamic color transform — for wheat head detection. By incorporating our DCT network into an existing object detector, we observe a notable improvement in detection performance. The DCT network exhibits robustness to various illumination conditions and indicates that a simple idea can make a difference if it is done right. Experiments and comparisons on the Global Wheat Head Detection Challenge 2021 dataset validate the effectiveness of our method.

References

[1] Alexey Bochkovskiy, Chien-Yao Wang, and H. Liao. Yolov4: Optimal speed and accuracy of object detection. *CoRR*, abs/2004.10934, 2020. 1, 4

[2] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *CVPR*, pages 6154–6162, 2018. 1

[3] Akshay L Chandra, Sai Vikas Desai, V. Balasubramanian, S. Ninomiya, and W. Guo. Active learning with point supervision for cost-effective panicle detection in cereal crops. *Plant Methods*, 16, 2020. 1

[4] Etienne David, Simon Madec, Pouria Sadeghi-Tehran, Helge Aasen, Bangyou Zheng, Shouyang Liu, Norbert Kirchgessner, Goro Ishikawa, Koichi Nagasawa, Minhajul A. Badhon, Curtis Pozniak, Benoit de Solan, Andreas Hund, Scott C. Chapman, Frédéric Baret, Ian Stavness, and Wei Guo. Global wheat head detection (gwhd) dataset: A large and diverse dataset of high-resolution rgb-labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics*, 2020:3521852, Aug 2020. 1

[5] E. David, Mario Serouart, Daniel Smith, S. Madec, Kaaviya Velumani, Shouyang Liu, Xu Wang, F. P. Espinosa, Shahameh Shafiee, I. Tahir, H. Tsujimoto, S. Nasuda, B. Zheng, Norbert Kichgessner, H. Aasen, A. Hund, Pouria Sadhegi-Tehran, K. Nagasawa, G. Ishikawa, S. Dandrifosse, A. Carlier, B. Mercatoris, Kentaro Kuroki, Haozhou Wang, Masanori Ishii, M. A. Badhon, C. Pozniak, D. LeBauer, Morten Lilimo, J. Poland, S. Chapman, B. D. Solan, F. Baret, I. Stavness, and Wei Guo. Global wheat head dataset 2021: more diversity to improve the benchmarking of wheat head localization methods. In *CoRR*, 2021. 1, 2, 4

[6] Sambuddha Ghosal, B. Zheng, S. Chapman, A. Potgieter, D. Jordan, Xuemin Wang, Ashutosh Kumar Singh, Arti Singh, M. Hirafuji, S. Ninomiya, B. Ganapathysubramanian, Soumik Sarkar, and Wei Guo. A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics*, 2019, 2019. 1

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 3

[8] Dong-Hyun Lee. Pseudo-label : The simple and efficient semi-supervised learning method for deep neural networks. In *ICML 2013 Workshop*, 07 2013. 4

[9] Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. Indices matter: Learning to index for deep image matting. In *ICCV*, pages 3265–3274, 2019. 1, 3

[10] Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. Index networks. *IEEE TPAMI*, pages 1–1, 2020. 1, 3

[11] S. Madec, Xiuliang Jin, Hao Lu, B. D. Solan, Shouyang Liu, F. Duyme, Emmanuelle Heritier, and F. Baret. Ear density estimation from high resolution rgb imagery using deep learning technique. *Agricultural and Forest Meteorology*, 264:225–234, 2019. 1

[12] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE TPAMI*, 39(6):1137–1149, 2017. 1

[13] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: Fully convolutional one-stage object detection. In *ICCV*, pages 9626–9635, 2019. 1

[14] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos: A simple and strong anchor-free object detector. *IEEE TPAMI*, pages 1–1, 2020. 1

[15] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Scaled-yolov4: Scaling cross stage partial network. In *CVPR*, pages 13029–13038, 2021. 1, 2, 4

[16] Hongwei Zou, Hao Lu, Yanan Li, Liang Liu, and Z. Cao. Maize tassels detection: a benchmark of the state of the art. *Plant Methods*, 16, 2020. 1