Estimation of Crop Production by Fusing Images and Crop Features

Ángela Casado-García, Jónathan Heras
Department of Mathematics and Computer Science
University of La Rioja, E-26004, Logroño, La Rioja, Spain
{angela.casado,jonathan.heras}@unirioja.es

Xabier Simon Martínez-Goñi, Jon Miranda-Apodaca, and Usue Pérez-López
Department of Plant Biology and Ecology, Faculty of Science and Technology
University of the Basque Country, E-48080 Bilbao, Bizkaia, Spain
{xabier.simon, jon.miranda, usue.perez}@ehu.eus

Abstract

The increasing global population and the growing frequency of droughts shows the necessity to enhance global food production and meet future food demands. However, achieving long-term food security and effectively mitigating the impact of climate change require a critical emphasis on sustainable systems to increase food production. Hence, automatic estimation of crop production can enable breeders and farmers to make data-driven decisions to optimise resources and maximise efficiency and sustainability. In this work, we have tackled this estimation task by applying deep learning methods to images taken from a digital RGB camera. Moreover, we have improved the results of those models by feeding the models with not only images but also crop features, such as the amount of fertilisers or the amount of water. The proposed data fusion approach can be applied to convolutional- and transformer-based models obtaining good results in both cases. As a result of our work, we have produced a model that estimates crop production of wheat and spelt with an MAE of 0.666*, and is a first step towards optimising resources and food production.*

Environmental CO_2 has been increasing exponentially over the last years, and it is expected to reach 700 ppm by 2070 [25]. Consequently, global temperatures are increasing, which in turn causes what is known as climate change [25]. Global warming will lead to greater water evaporation and increased aridity throughout the world, increasing extreme drought events in various regions of the planet. Drought is one of the most detrimental and limiting abiotic stresses for crops, causing decreases in photosynthesis, vegetative growth, number of flowers, and pollen germination [5,13]. Therefore, decreases in crop yields will be expected [1]. In addition, since the 1950s, the world population has been growing and is expected to reach 10.88 billion inhabitants by the end of the century [8]. The increase in global demographics along with the increase in drought episodes shows a clear need to increase world food production in order to meet future food demand. However, to mitigate the impact of climate change and ensure long-term food security, it is crucial that increase in food production is achieved through sustainable systems. Currently, most agricultural production is based on the conventional production method, which relies on inorganic nitrogen (N) based fertilizers to improve crop yields [23]. However, a large amount of the added N is lost to the environment, increasing the emission of greenhouse gases [7]. Thus, in the last decades the number of organic farming systems has increased, as it is believed to have the potential to mitigate the impact on climate change [19].

In this context, the high-throughput plant phenotyping (HTPP) has gained significant attention, since enables plant breeders, farmers, and researchers to acquire a vast amount of data on crop physiological status in an automated and efficient manner [6]. Furthermore, the availability of commercial RGB cameras with rigorous factory colour calibration has made it possible to easily acquire various vegetation indices through RGB image processing [20]. This low-cost procedure makes HTPP accessible to a wide range of users. Correlating vegetation indices with yield production would allow the identification and selection of high-yielding and stress-tolerant crop varieties [3, 32]. Therefore, developing methods to predict crop production and yield across various farming systems and growth conditions in a low-cost automatised manner could enable breeders, farmers, and researchers to make data-driven decisions and optimise resource allocation for maximum efficiency and sustainability, revolutionising agriculture.

Methods to estimate crop production can be split into

two groups: machine learning methods based on different physiological and ecological factors; and computer vision methods. In the former solution, crop production can be correlated with a range of vegetation or growth traits [2], and it can be also estimated from ecological factors such as water availability, temperature or other natural resources [10, 36]. In the computer vision approach, the last years have been dominated by deep learning models capable of predicting yield for cotton [30], soybean [24] or rice [34] among others [31]. In this paper, we aim to combine both approaches by means of data fusion methods. Data fusion, which integrates data from multiple modalities using Machine Learning and Deep Learning techniques, has been of growing interest in its application to precision agriculture, since it provides reliable, precise and valuable information [4]. In particular, the contributions of this work are:

- We develop several models for estimating crop yield for wheat and spelt.
- We analyse the impact of fusing different crop features with computer vision models to predict crop yield using different approaches.
- We release our code in a public repository to facilitate the application of our methods to other projects[1].

1. Materials and methods

In this section, we present both the dataset and computational materials and methods employed in this work.

1.1. Dataset

In this work, we have developed models to automatically estimate yield production from photographs of two species: *Triticum aestivum var. Florence Aurora* (wheat) and *Triticum spelta var. Franckenkorn* (spelt).

Photographs were captured using a digital RGB camera (EVIL Canon EOS M200) mounted to a monopod (Hama Monopod Star 78 Mono), see Figure 1 for samples of the captured photographs. To ensure a consistent distance to capture the photos, a fishing line measuring approximately 1 metre was added to the monopod. The fishing line was secured with a fishing weight at the other end to maintain tautness and a fixed distance during photography. Zenithal photos were taken by positioning the end of the fishing line at the highest point of the crops, ensuring a constant distance of 1 meter. A total of 829 photos were captured, 558 images were used for training different models, 62 images for validation, and the rest for evaluating the models.

[1]The code of the project is available at `https://github.com/joheras/yield-prediction/`

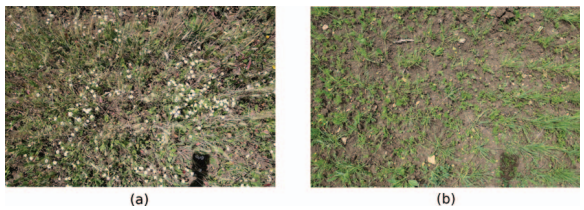

Figure 1. Samples of the captured photographs: (a) wheat and (b) spelt.

Feature	Values	Obtained by	Correlation
species	{wheat, spelt}	manual	-
water treatment	{d,ww}	manual	-
management	{conventional, ecological}	manual	-
avg water (L m^2)	$[2752.8, 3096.2]$	manual	0.257
Nitrogen (kg)	$[0, 0.168]$	manual	0.974
Phosphorus (kg)	$[0, 0.072]$	manual	0.974
Potassium (kg)	$[0, 0.072]$	manual	0.974
GA	$[0, 1]$	CerealScanner	0.314
GGA	$[0, 1]$	CerealScanner	0.398
CSI	$[4, 93]$	CerealScanner	-0.143
NGRDI	$[0, 0.135]$	CerealScanner	-0.004
TGI	$[776.6, 4993.7]$	CerealScanner	-0.712
Final yield (kg)	$[0.8579, 11.35]$	manual	-

Table 1. Features employed in the study and their correlation with final yield.

Once we had the photos, we automatically calculated several features using the CerealScanner plugin in Fiji[2]. These features include Green Area (GA), Greener Green Area (GGA), Crop Senescence Index (CSI), Normalised Green-Red Difference Index (NGRDI), and Triangular Greenness Index (TGI). GA and GGA estimate the photosynthetic surface area of the canopy, while CSI indicates the degree of canopy senescence [11]. NGRDI is correlated with aboveground biomass, and TGI allows estimation of chlorophyll concentration in the canopy [17,18]. All these features are correlated with plant health and photosynthetic activity and can be correlated with biomass and yield. In addition, for each photo the following features from the crops were annotated: water treatment, management, average water, Nitrogen, Phosphorus and Potassium, and average height. A table with all the features considered in this study, their values, and the correlation with the final yield in kg is provided in Table 1.

1.2. Computational methods

We have conducted a study of six deep learning architectures for estimating yield production. The architectures studied included three convolutional neural networks (namely, EfficientNet v2 medium [29], ResNet-50 [14] and ConvNext base [22]), and three transformer-based architectures (in particular, Swin v2 base [21], Vit base patch 16 [9] and VOLO d2 [35]) — the selection of deep learning architectures was based on their outstanding performance on

[2]`https://integrativecropecophysiology.com/software-development/cerealscanner/`

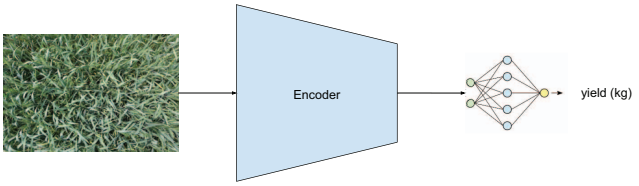

Figure 2. Base approach to train the models that estimate yield production.

other tasks [15], but the data fusion approach proposed in this paper can be applied with any deep learning architecture. All the networks used in our experiments were implemented in Pytorch [26], and have been trained thanks to the functionality of the FastAI [16] and Timm [33] libraries using a GPU Nvidia RTX 2080 Ti.

In order to establish a baseline for our models, we have used the transfer-learning method presented in [16] for training both the convolutional and transformer-based architectures. The training method is a two-stage procedure that starts from a model pretrained in the ImageNet challenge, and can be summarised as follows, see Figure 2. In the first stage, we replaced the head of the model (that is, the layers that give us the classification of the images), with a new head adapted to estimate crop production in our particular dataset. Then, we trained these new layers (the rest of the layers stayed frozen) with our data for two epochs. In the second stage, we unfreezed the whole model and retrained all the layers of the model with our dataset for 30 epochs. In order to find a suitable learning rate for both the first and second stage, we used cyclical learning rates for optimisation [28]. Moreover, we employed data augmentation [27] (using vertical and horizontal flips, rotations from -180^{o} to 180^{o}, zooms and lighting transformations) to prevent overfitting. All convolutional models were trained with images of size 512×512, and transformers with images of size 384×384. Finally, we analysed the impact of Min-Max normalisation and standardisation for yield estimation [12].

Moreover, we have applied two data fusion techniques, one at the input level, and another at the output stage. In both approaches, we start with a model pretrained in the ImageNet challenge. As in the baseline, in the multi-input fusion approach, we replaced the head of the model with a new head adapted to estimate crop production in our particular dataset; the difference is that the new head not only receives its input from the the ImageNet model but also from the crop features, see Figure 3. In the case of the multi-output fusion approach, we replaced the head of the model with several heads to predict not only crop production but also other crop features, see Figure 4. In both approaches, we used three sets of crop features: the features obtained with CerealScanner, the crop treatment features, and all the features — since the crop features have different scales, we studied both min-max normalisation and standardisation. For training, we followed the same two-stage regime explained previously (we trained the added layers for two epochs, unfreezed the model, and retrained all the layers for 30 epochs), applied cyclical learning rates for optimisation and data augmentation, and used images of the same resolutions explained previously.

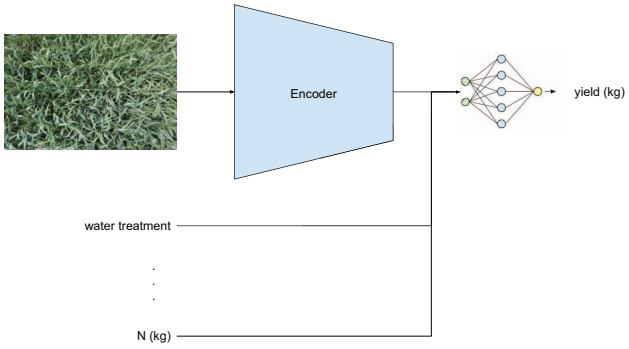

Figure 3. Multi-input approach to estimate yield production

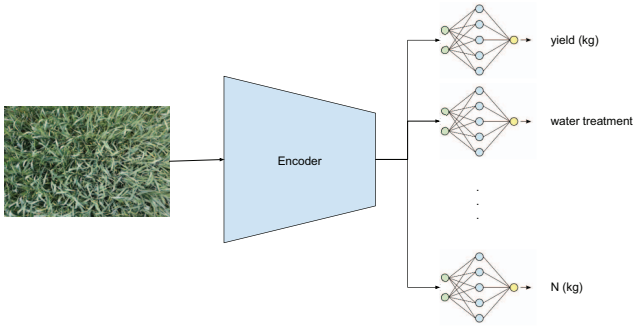

Figure 4. Multi-output approach to estimate yield production

For evaluation, we used two popular error-based metrics, root mean squared error (RMSE) and mean absolute error (MAE), to study the performance of each method in the test set.

2. Results and Discussion

The performance of the baseline models is presented in Table 2. In that table, we also include the impact of applying Min-Max normalisation and standardisation. From those results, we can draw several conclusions. First, both convolutional and transformer-based models achieved a MAE close to 1 Kg without applying any normalisation; being VOLO the architecture that obtained the best result (MAE of 0.936). Moreover, the application of both Min-Max normalisation and standardisation improved the results of all models except for the Swin architecture, but standardisation allowed us to obtain better results in general. Finally, the VOLO architecture applying the standardisation step produced the best model with a MAE of 0.866.

We focus now on the results obtained by combining images and crop features using both the multi-input and multi-

	No norm		Min-Max		STD norm	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
ConvNext	0.999	1.284	0.912	1.113	0.900	1.093
EfficientNet	1.067	1.248	1.128	1.505	1.038	1.282
Resnet	1.695	1.839	1.126	1.341	1.205	1.405
Swin	1.092	1.252	1.107	1.379	1.096	1.339
Vit	1.111	1.375	0.952	1.156	0.977	1.187
VOLO	0.936	1.171	0.922	1.139	**0.866**	**1.066**

Table 2. Results for the baseline approach. In bold the best results.

output data fusion approach, see Table 3. From those results, we can notice that the multi-input approach serves to improve the performance of all the models. However, it is necessary to either apply Min-Max normalisation or standardise the features; otherwise, the results are worse than those obtained only using the image for training. The preprocessing step is necessary because otherwise the features have different scales, and this affects the performance of the network. In general, Min-Max normalisation provides better results (mean MAE of 0.793) than the standardisation step (mean MAE of 0.823). If we consider transformers and convolutional networks separately, the transformer models worked better with the standardisation step (mean MAE of 0.791) and the convolutional models with Min-Max normalisation (mean MAE of 0.7483). The last conclusion worth highlighting from the multi-input approach is that the ResNet-50 model achieved the best result (MAE of 0.667), this shows that this classical convolutional architecture might still produce better results than new architectures.

In contrast with the results achieved by the multi-input approach, the multi-output data fusion method obtained considerable worse results than those obtained by the baseline or multi-input models. The best multi-output model was a ConvNext model with Min-Max normalisation (MAE of 2.892). The explanation for these results could lie in the number of epochs (30) in which the models were trained. These models have more features due to the multiple heads, and, therefore, training them might require more resources (data and time) or the usage of different configurations for each one of the heads (for instance, the learning rate).

The next set of experiments were focused on the use of features obtained from the images thanks to the CerealScanner plugin, see Table 4. In both the multi-input and multi-output approach, models trained combining images and CerealScanner features got worse results than baseline models, and models trained with a combination of images and crop features. An explanation for those results might be the low correlation of CerealScanner features and yield production, whereas crop features have a higher correlation with yield production, see Table 1. Hence, the combination with CerealScanner features have a negative impact on the performance of the models.

In the last set of experiments, we analysed the results obtained by the models that combine images with all features included in this work; see Table 5. We can observe the same behaviour of the models trained only with crop features; that is, the multi-input models obtained better results than baseline models by applying a pre-processing step, whereas multi-output models fail to estimate crop production. In fact, the multi-input models trained with all the features after Min-Max normalisation obtained the best results in our study (mean MAE of 0.774). This shows that in spite of including some features that might hinder the performance of the networks (as shown with the CerealScanner features), the networks are able to take advantage of relevant information provided as additional features. As the last conclusion, we notice that there are not significant differences between convolutional models (best mean MAE of 0.756) and transformers (best mean MAE of 0.782); so, both kinds of architectures can serve to estimate crop production.

3. Conclusions and further work

In this paper, we have shown that it is feasible to estimate yield production using both convolutional and transformer-based models from photos taken using a digital RGB camera. Moreover, the performance of those models can be considerably improved by taking advantage of some extra information from crops. However, such information must be added carefully to the models. In particular, all extra features feed to the model should be put in the same scale using either Min-Max normalisation or standardisation. In addition, it is necessary to analyse the features that are related to crop production (such as water treatment, or Nitrogen), otherwise the performance of the networks might decay if we only consider unrelated features. Finally, the architecture of the networks should be modified to include those features as input since trying to predict not only yield production but also other features is not straightforward, and might require additional resources.

There are several tasks that remain as further work. First, we would like to study multi-output models more deeply to produce better results. Moreover, we want to analyse the impact of each additional feature provided to the model in order to detect those that are more relevant for estimating yield production. Finally, we are interested in using depth information to improve the performance of the models.

Acknowledgements

This work was partially supported by Grant PID2020-115225RB-I00 funded by MCIN/AEI/10.13039/501100011033 and by grants 00037-IDA2021-45 and IT1682-22 from the Basque Government. Ángela Casado-García has a FPI grant from Community of La Rioja 2020.

	Multi-input						Multi-output					
	No norm		Min-Max		STD norm		No norm		Min-Max		STD norm	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
ConvNext	1.753	1.986	0.823	0.938	0.760	0.914	5.885	7.131	2.892	3.619	4.124	5.286
EfficientNet	1.458	1.759	0.755	0.929	0.920	1.051	5.928	7.124	7.495	9.180	3.523	3.887
Resnet	1.549	1.693	**0.667**	**0.804**	0.882	1.009	6.005	7.203	7.172	8.263	3.178	3.676
Swin	1.541	1.916	0.926	1.013	0.859	0.988	5.821	7.194	4.806	5.827	3.821	4.375
Vit	1.492	1.782	0.770	0.913	0.724	0.896	5.532	6.849	3.667	4.685	3.721	4.062
VOLO	1.428	1.638	0.819	0.969	0.792	0.958	5.758	6.96	5.245	7.423	4.431	4.830

Table 3. Results for data fusion using crop features. In bold the best results.

	Multi-input						Multi-output					
	No norm		Min-Max		STD norm		No norm		Min-Max		STD norm	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
ConvNext	2.203	2.649	2.202	2.498	2.137	2.419	12.138	12.796	10.245	11.983	4.021	4.471
EfficientNet	2.572	3.135	2.038	2.541	1.631	2.092	5.533	6.488	6.080	8.476	4.073	4.597
Resnet	1.436	1.923	**1.340**	**1.770**	1.591	1.878	6.026	7.313	5.853	7.331	3.494	4.077
Swin	2.299	2.709	2.15	2.608	2.123	2.391	5.745	6.912	6.728	8.241	4.055	4.448
Vit	2.636	3.191	2.191	2.687	2.293	2.632	5.382	6.692	5.788	7.059	3.890	4.468
VOLO	2.428	2.855	1.783	2.107	1.828	2.167	6.003	7.148	5.628	6.615	4.514	4.895

Table 4. Results for data fusion using features extracted with CerealScanner. In bold the best results.

References

[1] Elizabeth A Ainsworth and Stephen P Long. 30 years of free-air carbon dioxide enrichment (face): what have we learned about future crop productivity and its potential for adaptation? *Global Change Biology*, 27(1):27–49, 2021.

[2] Mohammad Ajlouni, Audrey Kruse, Jorge A Condori-Apfata, Maria Valderrama Valencia, Chris Hoagland, Yang Yang, and Mohsen Mohammadi. Growth analysis of wheat using machine vision: opportunities and challenges. *Sensors*, 20(22):6501, 2020.

[3] José Luis Araus and Jill E Cairns. Field high-throughput phenotyping: the new crop breeding frontier. *Trends in plant science*, 19(1):52–61, 2014.

[4] Jayme Garcia Arnal Barbedo. Data fusion in agriculture: Resolving ambiguities and closing data gaps. *Sensors*, 22(6):2285, 2022.

[5] Beáta Barnabás, Katalin Jäger, and Attila Fehér. The effect of drought and heat stress on reproductive processes in cereals. *Plant, cell & environment*, 31(1):11–38, 2008.

[6] Aakash Chawade, Joost van Ham, Hanna Blomquist, Oscar Bagge, Erik Alexandersson, and Rodomiro Ortiz. High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy*, 9(5):258, 2019.

[7] Mario Corrochano-Monsalve, Adrián Bozal-Leorri, Cristina Sánchez, Carmen González-Murua, and José-María Estavillo. Joint application of urease and nitrification inhibitors to diminish gaseous nitrogen losses under different tillage systems. *Journal of Cleaner Production*, 289:125701, 2021.

[8] Division UND of E and SAP. World population prospects 2022. summary of results. united nation, 2022.

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[10] Dhivya Elavarasan, Durai Raj Vincent, Vishal Sharma, Albert Y Zomaya, and Kathiravan Srinivasan. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and electronics in agriculture*, 155:257–282, 2018.

[11] Jose A Fernandez-Gallego, Shawn C Kefauver, Thomas Vatter, Nieves Aparicio Gutiérrez, María Teresa Nieto-Taladriz, and José Luis Araus. Low-cost assessment of grain yield in durum wheat using rgb images. *European Journal of Agronomy*, 105:146–156, 2019.

[12] Aurélien Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. " O'Reilly Media, Inc.", 2022.

[13] Sharon B Gray and Siobhan M Brady. Plant developmental responses to climate change. *Developmental biology*, 419(1):64–77, 2016.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[15] Jeremy Howard. The best vision models for fine-tuning, 2022.

[16] J. Howard and S. Gugger. Fastai: A layered api for deep learning. *Information*, 11:108, 02 2020.

[17] E Raymond Hunt, Michel Cavigelli, Craig ST Daughtry, James E Mcmurtrey, and Charles L Walthall. Evaluation of digital photography from model aircraft for remote sensing of crop biomass and nitrogen status. *Precision Agriculture*, 6:359–378, 2005.

[18] E Raymond Hunt Jr, Paul C Doraiswamy, James E McMurtrey, Craig ST Daughtry, Eileen M Perry, and Bakhyt

	Multi-input						Multi-output					
	No norm		Min-Max		STD norm		No norm		Min-Max		STD norm	
	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
ConvNext	0.845	1.095	0.749	0.927	0.785	0.903	13.834	14.443	3.785	4.179	3.597	3.899
EfficientNet	1.536	1.678	**0.748**	**0.917**	0.786	0.901	15.383	16.431	3.580	3.956	4.023	4.318
Resnet	1.595	1.889	0.772	0.938	0.821	0.999	15.684	16.250	4.241	4.724	3.700	3.932
Swin	1.584	1.717	0.770	0.919	0.817	0.983	430.088	430.255	3.749	4.091	3.798	4.040
Vit	2.279	2.578	0.846	0.995	0.770	0.955	430.324	430.403	3.819	4.137	3.794	4.029
VOLO	2.167	2.681	0.757	0.933	0.758	0.923	14.226	14.757	3.887	4.189	3.903	4.132

Table 5. Results for data fusion using all the features. In bold the best results.

Akhmedov. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International journal of applied earth observation and Geoinformation*, 21:103–112, 2013.

[19] Diana Ivanova, John Barrett, Dominik Wiedenhofer, Biljana Macura, Max Callaghan, and Felix Creutzig. Quantifying the potential for climate change mitigation of consumption options. *Environmental Research Letters*, 15(9):093001, 2020.

[20] Shawn C Kefauver, Rubén Vicente, Omar Vergara-Díaz, Jose A Fernandez-Gallego, Samir Kerfal, Antonio Lopez, James PE Melichar, Maria D Serret Molins, and José L Araus. Comparative uav and field phenotyping to assess yield and nitrogen use efficiency in hybrid and conventional barley. *Frontiers in Plant Science*, 8:1733, 2017.

[21] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, et al. Swin transformer v2: Scaling up capacity and resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12009–12019, 2022.

[22] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11976–11986, 2022.

[23] Derek H Lynch, Rod MacRae, and Ralph C Martin. The carbon and global warming potential impacts of organic farming: does it have a significant role in an energy constrained world? *Sustainability*, 3(2):322–362, 2011.

[24] Maitiniyazi Maimaitijiang, Vasit Sagan, Paheding Sidike, Sean Hartling, Flavio Esposito, and Felix B Fritschi. Soybean yield prediction from uav using multimodal data fusion and deep learning. *Remote sensing of environment*, 237:111599, 2020.

[25] Rajendra K Pachauri, Myles R Allen, Vicente R Barros, John Broome, Wolfgang Cramer, Renate Christ, John A Church, Leon Clarke, Qin Dahe, Purnamita Dasgupta, et al. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc, 2014.

[26] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[27] P. Simard, D. Steinkraus, and J. C. Platt. Best practices for convolutional neural networks applied to visual document analysis. In *12th International Conference on Document Analysis and Recognition*, volume 2, pages 958–964, 2003.

[28] L. Smith. Cyclical learning rates for training neural networks. In *IEEE Winter Conference on Applications of Computer Vision*, pages 464–472, 2017.

[29] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021.

[30] Danilo Tedesco-Oliveira, Rouverson Pereira da Silva, Walter Maldonado Jr, and Cristiano Zerbato. Convolutional neural networks in predicting cotton yield from images of commercial fields. *Computers and Electronics in Agriculture*, 171:105307, 2020.

[31] Thomas Van Klompenburg, Ayalew Kassahun, and Cagatay Catal. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709, 2020.

[32] VS Weber, José Luis Araus, Jill E Cairns, C Sanchez, Albrecht E Melchinger, and Elena Orsini. Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. *Field Crops Research*, 128:82–90, 2012.

[33] R. Wightman et al. Pytorch image models, 2021.

[34] Qi Yang, Liangsheng Shi, Jinye Han, Yuanyuan Zha, and Penghui Zhu. Deep convolutional neural networks for rice grain yield estimation at the ripening stage using uav-based remotely sensed images. *Field Crops Research*, 235:142–153, 2019.

[35] Li Yuan, Qibin Hou, Zihang Jiang, Jiashi Feng, and Shuicheng Yan. Volo: Vision outlooker for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 45(5):6575–6586, 2022.

[36] Qiyu Zhou and Douglas J Soldat. Creeping bentgrass yield prediction with machine learning models. *Frontiers in Plant Science*, 12:749854, 2021.