

Автоматизация детекции центров колосков на RGB изображениях колоса пшеницы

Бусов И.Д.^{1, 2*}, Генаев М.А.^{1, 3}, Комышев Е.Г.¹, Пронина В.А.², Коваль В.С.^{1, 3},
Афонников Д.А.^{1, 2, 3}

¹ Институт цитологии и генетики СО РАН, Новосибирск, Россия

² Новосибирский государственный университет, Новосибирск, Россия

³ Курчатowski геномный центр ИЦиГ СО РАН, Новосибирск, Россия

* i.busov@g.nsu.ru

Ключевые слова: пшеница; колос; подсчет колосков; сегментация; детекция; компьютерное зрение; глубокое машинное обучение

Мотивация и цель: Пшеница – это важная сельскохозяйственная культура, которая является одним из основных продуктов питания. Количество колосков в колосе растения является значимым признаком при селекции пшеницы, поскольку это свойство непосредственно связано с урожайностью. Помимо количества колосков, их взаимное расположение также может иметь биологическое значение. Цель работы заключается в разработке точного и эффективного метода для автоматического обнаружения и подсчета центров колосков пшеницы на RGB изображениях колоса.

Методы и алгоритмы: В рамках данного исследования были рассмотрены несколько подходов к решению поставленной задачи: 1) использование нейронной сети с архитектурой YOLO [1], с ограничивающими квадратами с фиксированной длиной стороны, в центре которых находились центры колосков; 2) использование нейронной сети с архитектурой U-Net [2], обученной на бинарных масках с использованием бинарной кросс-энтропии [3]; 3) использование нейронной сети с архитектурой U-Net, обученной на масках из нормального двумерного распределения с применением дивергенции Кульбака–Лейблера [4] в качестве функции потерь. Гиперпараметры моделей подбирались с использованием Байесовской оптимизации (Гауссовских процессов) [5]. Для оценки качества моделей использовались пространственные метрики детекции колоска на изображении (precision, recall, f1) и количественные метрики предсказания числа колосков в колосе (MAE, MSE). Пространственные метрики считались на основе характеристик TP, FP и FN. Случай TP – это предсказанное положение центра колоска, находящееся в радиусе 3 мм от ближайшего реального центра. Если количество предсказанных центров, находящихся в этой окрестности радиуса 3 мм, превышало 1, то оставшиеся предсказанные центры определялись как FP. В случае, если предсказанные центры не находились в указанной окрестности, они относились к FN.

Результаты: Обученные модели были протестированы на двух выборках: тестовой и отложенной. Тестовая выборка состояла из изображений, полученных одновременно с изображениями, на которых обучались нейронные сети, выборка включала в себя 348 изображений колосьев пшеницы. Отложенная выборка была получена намного позже тренировочной и включала в себя 14 изображений колосьев пшеницы. Результаты на тестовой выборке представлены в табл. 1,

результаты на отложенной выборке представлены в табл. 2. Наилучшие результаты продемонстрировала модель с архитектурой U-Net, обученная на нормальных масках с помощью дивергенции Кульбака–Лейблера в качестве функции потерь. Модель с архитектурой U-Net, обученная на бинарных масках с помощью бинарной кросс-энтропии в качестве функции потерь, продемонстрировала результаты немного хуже. Модель, основанная на архитектуре YOLO, показала наихудшие результаты.

Таблица 1. Результаты на тестовой выборке

модель (3 мм)	precision	recall	f1	mae	mse
U-Net (kld)	0.977	0.976	0.975	0.557	1.086
U-Net (bce)	0.983	0.963	0.970	0.597	1.551
YOLO	0.733	0.760	0.734	3.641	26.486

Примечание. Здесь и в табл. 2: U-Net (kld) – модель нейронной сети с архитектурой U-Net, обученной на масках из нормального двумерного распределения с применением дивергенции Кульбака–Лейблера в качестве функции потерь; U-Net (bce) – нейронной сети с архитектурой U-Net, обученной на бинарных масках с использованием бинарной кросс-энтропии; YOLO – нейронной сети с архитектурой YOLO

Таблица 2. Результаты на отложенной выборке

модель (3 мм)	precision	recall	f1	mae	mse
U-Net (kld)	0.971	0.953	0.959	0.929	2.071
U-Net (bce)	0.975	0.946	0.957	1.071	3.214
YOLO	0.678	0.753	0.704	3.0	15.714

Выводы: В ходе работы была достигнута поставленная цель, заключающаяся в разработке точного автоматического метода для локализации и подсчета колосков пшеницы на изображениях.

Финансирование: Исследование поддержано грантом РФФИ № 23-14-00150.

Automation of spikelet center detection on RGB images of wheat ear

Busov I.D.^{1, 2*}, Genaev M.A.^{1, 3}, Komyshev E.G.¹, Pronina V.A.², Koval V.S.^{1, 3}, Afonnikov D.A.^{1, 2, 3}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

³ Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

* i.busov@g.nsu.ru

Key words: wheat; spikelet; spikelet counting; segmentation; detection; computer vision; deep machine learning

Motivation and Aim: Wheat is an important crop that is a staple food. The number of spikelets in an ear of a plant is a significant trait in wheat breeding because this trait is

directly related to yield. In addition to the number of spikelets, their mutual arrangement may also have biological significance. The aim of this work is to develop an accurate and efficient method for automatic detection and counting of wheat spikelet centers in RGB images of the spikelet.

Methods and Algorithms: In this research, several approaches have been considered to solve the problem: (1) using a neural network with the YOLO architecture [1], with bounding squares with fixed side lengths at the center of which were the spikelet centers; (2) using a neural network with the U-Net architecture [2] trained on binary masks using binary cross-entropy [3]; and (3) a neural network with U-Net architecture trained on masks from a normal bivariate distribution using the Kulback–Leibler divergence [4] as a loss function. The hyperparameters of the models were selected using Bayesian optimization (Gaussian processes) [5]. Spatial metrics for spikelet detection in the image (precision, recall, f1) and quantitative metrics for predicting the number of spikelets in a spikelet (MAE, MSE) were used to evaluate the quality of the models. The spatial metrics were considered based on TP, FP and FN features. The TP case is the predicted position of the spikelet center that is within 3 mm radius from the nearest real center. If the number of predicted centers located in this vicinity of 3 mm radius was greater than 1, the remaining predicted centers were defined as FP. In case the predicted centers were not in the specified vicinity, they were referred to as FN.

Results: The trained models were tested on two samples: a test sample and a delayed sample. The test sample consisted of images acquired simultaneously with the images on which the neural networks were trained, the sample included 348 images of wheat ears. The delayed sample was obtained much later than the training sample, and included 14 images of wheat ears. The results on the test sample are presented in Table 1, and the results on the deferred sample are presented in Table 2. The model with U-Net architecture trained on normal masks using Kulbak–Leibler divergence as the loss function showed the best results. The model with U-Net architecture trained on binary masks using binary cross-entropy as a loss function performed slightly worse. The model based on YOLO architecture showed the worst results.

Table 1. Results on the test sample

model (3 mm)	precision	recall	f1	mae	mse
U-Net (kld)	0.977	0.976	0.975	0.557	1.086
U-Net (bce)	0.983	0.963	0.970	0.597	1.551
YOLO	0.733	0.760	0.734	3.641	26.486

Note. Here and in Table 2: U-Net (kld) – a neural network with U-Net architecture trained on masks from a normal bivariate distribution using Kullback–Leibler divergence as the loss function. U-Net (bce) – neural network with U-Net architecture trained on binary masks using binary cross-entropy. YOLO – neural network with YOLO architecture

Table 2. Results on delayed sampling

model (3 mm)	precision	recall	f1	mae	mse
U-Net (kld)	0.971	0.953	0.959	0.929	2.071
U-Net (bce)	0.975	0.946	0.957	1.071	3.214
YOLO	0.678	0.753	0.704	3.0	15.714

Conclusions: The work achieved its objective of developing an accurate automatic method for localizing and counting wheat spikelets in images.

Funding: The research was supported by the RSF grant No. 23-14-00150.

Список литературы/References

1. Jiang P., Ergu D., Liu F., Cai Y., Ma B. A Review of Yolo algorithm developments. *Procedia Comput Sci.* 2022;199:1066-1073
2. Siddique N., Paheding S., Elkin C.P., Devabhaktuni V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access.* 2021;9:82031-82057
3. Ruby U., Yendapalli V. Binary cross entropy with deep learning technique for image classification. *Int J Adv Trends Comput Sci Eng.* 2020;9(4):5393-5397
4. Kim T., Oh J., Kim N., Cho S., Yun S.Y. Comparing kullback-leibler divergence and mean squared error loss in knowledge distillation. *arXiv.* 2021
5. MacKay D.J. Introduction to Gaussian processes. 1998. [<http://www.inference.org.uk/mackay/gpB.pdf>]