

doi 10.18699/vjgb-26-09

Детекция колосков в колосе пшеницы на RGB-изображениях с использованием глубокого машинного обучения

М.А. Генаев ^{1, 2} , И.Д. Бусов^{1, 3}, Ю.В. Кручинина ^{1, 2}, В.С. Коваль^{1, 2}, Н.П. Гончаров^{1, 3}¹ Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия² Курчатовский геномный центр ИЦиГ СО РАН, Новосибирск, Россия³ Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия mag@bionet.nsc.ru

Аннотация. В работе рассматривается задача автоматизированного высокопроизводительного фенотипирования признаков колоса пшеницы с использованием современных методов компьютерного зрения и глубокого обучения. Точная оценка числа колосков является важным компонентом анализа продуктивности растения, однако традиционные методы ручной разметки и подсчета крайне трудоемки, плохо масштабируются и требуют значительных временных затрат. В исследовании предложен подход для эффективной детекции колосков, основанный на использовании упрощенной точечной разметки, при которой эксперт отмечает только центры колосков, без необходимости формировать трудоемкие пиксельные маски или ограничивающие рамки. Такая схема позволяет существенно снизить стоимость подготовки обучающей выборки и ускорить процесс аннотации. Для определения оптимального способа обработки упрощенной разметки были исследованы три метода: сегментация бинарных масок с помощью архитектуры U-Net, регрессия плотностных карт на основе двумерного нормального распределения и функции дивергенции Кульбака–Лейблера, а также детекция областей фиксированного размера с использованием модели YOLOv8. Проведено сравнение точности методов по количественным (MAE, MAPE) и пространственным метрикам (Precision, Recall, F1) на тестовых наборах изображений. Анализ результатов показал, что подходы, основанные на U-Net, обеспечивают высокую точность локализации и подсчета колосков при минимальных затратах на разметку данных, тогда как метод YOLOv8 менее устойчив к геометрической вариативности реальных объектов. Предложенный подход демонстрирует, что комбинация точечной разметки и современных моделей сегментации является эффективным инструментом для автоматизации фенотипирования, что может значительно ускорить селекционные исследования и расширить возможности высокопроизводительного анализа морфологических признаков растений.

Ключевые слова: компьютерное зрение; глубокое обучение; пшеница; колос; число колосков; фенотипирование; детекция объектов

Для цитирования: Генаев М.А., Бусов И.Д., Кручинина Ю.В., Коваль В.С., Гончаров Н.П. Детекция колосков в колосе пшеницы на RGB-изображениях с использованием глубокого машинного обучения. *Вавиловский журнал генетики и селекции*. 2026;30(1):27-35. doi 10.18699/vjgb-26-09

Финансирование. Подготовка данных, разработка и верификация алгоритма выполнены при поддержке Российского научного фонда, проект № 23-14-00150.

Благодарности. Расчеты выполнялись с использованием вычислительных ресурсов ЦКП «Биоинформатика» при поддержке бюджетного проекта № FWNR-2022-0020.

Wheat spikelet detection on RGB images using deep machine learning

М.А. Genaev ^{1, 2} , I.D. Busov^{1, 3}, Yu.V. Kruchina ^{1, 2}, V.S. Koval^{1, 2}, N.P. Goncharov^{1, 3}¹ Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia² Kurchatov Genomic Center of ICG SB RAS, Novosibirsk, Russia³ Novosibirsk State University, Novosibirsk, Russia mag@bionet.nsc.ru

Abstract. This study addresses the challenge of automated high-throughput phenotyping of wheat spike characteristics using modern computer vision and deep learning methods. Accurate estimation of spikelet number is a key indicator of plant productivity, yet traditional manual counting approaches are labor-intensive, slow, and difficult to scale to large breeding datasets. To overcome these limitations, we propose a spikelet detection strategy based on simplified point annotations, where an expert marks only the centers of spikelets rather than drawing detailed segmentation masks or bounding boxes. This significantly reduces annotation time and lowers the overall cost of preparing training datasets for machine learning models. To determine the most effective way of utilizing such simplified annotations, three computational methods were explored: segmentation of binary masks using a U-Net architecture, density regression based on two-dimensional Gaussian distributions optimized via Kullback–Leibler divergence, and detection of fixed-size bounding regions using the YOLOv8 object detection framework. The models were evaluated on dedicated test datasets using both quantitative metrics (MAE,

MAPE) and spatial localization metrics (Precision, Recall, F1 score). The results demonstrate that U-Net-based approaches provide consistently high accuracy in spikelet localization and counting while maintaining robustness to annotation imperfections. In contrast, the YOLOv8-based method showed reduced performance, likely due to the geometric mismatch between fixed-size boxes and the natural elongated shape of spikelets. Overall, the proposed methodology highlights the effectiveness of combining minimalistic point-level annotation with advanced segmentation models for automating phenotyping workflows. This approach has the potential to accelerate breeding programs, enhance the efficiency of large-scale phenotypic data collection, and support further development of robust computer-vision tools for plant science applications.

Key words: computer vision; deep learning; wheat; spike; spikelets per spike; phenotyping; object detection

For citation: Genaev M.A., Busov I.D., Kruchinina Yu.V., Koval V.S., Goncharov N.P. Wheat spikelet detection on RGB images using deep machine learning. *Vavilovskii Zhurnal Genetiki i Selekcii* = *Vavilov J Genet Breed.* 2026;30(1):27-35. doi 10.18699/vjgb-26-09

Введение

Рост населения и локальное и глобальное изменение климата обуславливают необходимость в ускоренной селекции сельскохозяйственных культур, таких как пшеница, для повышения урожайности и устойчивости к биотическим и абиотическим факторам внешней среды (Ефимов и др., 2024). Оценка признаков, в том числе фенотипирование параметров колоса, – один из ключевых этапов селекции. Среди них число колосков в колосе является наиболее важным показателем продуктивности растения пшеницы (Афонников и др., 2016; Скрипка и др., 2016; Маслова и др., 2018; Vahamidis et al., 2019; и др.).

Число колосков в колосе у пшеницы имеет сложный генетический контроль (Zhang B. et al., 2015) и часто связано с плотностью колоса – не менее важным селекционным и таксономическим признаком (Vavilova et al., 2017; Вавилова и др., 2019; Савин, 2019).

Традиционные методы фенотипирования параметров колоса, базирующиеся на визуальной оценке экспертами, являются медленными, дорогостоящими и субъективными (Konopatskaia et al., 2016). Это стимулирует разработку автоматизированных, использующих компьютерное зрение для высокопроизводительного анализа (Li et al., 2017; Liu et al., 2017; Genaev et al., 2019).

Методы компьютерного зрения на основе глубокого обучения доказали свою эффективность в автоматизации задач фенотипирования сельскохозяйственных растений (Artemenko et al., 2024). Существующие подходы к решению задачи подсчета колосков можно условно разделить на два типа. Первый опирается на детекцию объектов. Например, Ф. Хорошевская с коллегами (Khoroshevsky et al., 2021) использовали архитектуру RetinaNet для детекции и подсчета колосков непосредственно в полевых условиях; достигнуты средние значения процентной ошибки (MAPE) в диапазоне от 9.2 до 11.5 %. Л. Ши с соавторами (Shi et al., 2023) применили модель YOLOv5s для детекции числа колосков на изображениях колосьев, получив на тестовой выборке для зрелой пшеницы среднее значение абсолютной ошибки (MAE) числа колосков, равное 0.43. Второй тип подходов использует семантическую сегментацию. В работе (Misra et al., 2020) предложена архитектура SpikeSegNet на основе U-Net, которая достигла точности подсчета (accuracy) 95 %. Однако общим для этих методов является необходимость в трудоемкой и дорогостоящей полной разметке данных (использование ограничивающих

прямоугольников или пиксельных масок), что становится ключевым ограничением при масштабировании исследований.

В качестве альтернативы для снижения затрат на разметку предлагается использовать точечную аннотацию, когда эксперту требуется отметить лишь центр объекта. Ф. Чен с коллегами (Chen et al., 2021) продемонстрировали, что обучение с неполной точечной разметкой, где аннотировано лишь 50 % колосков, приводит к потере точности на 6.5 % (F1 падает с 84.15 до 78.65 %) по сравнению с полной разметкой. А уменьшение доли размеченных объектов до 10 % снижает точность на 16.5 %. Таким образом, уменьшение трудозатрат на разметку в 10 раз сохраняет 83.5 % исходной точности, что подтверждает перспективность этого направления. Другой путь упрощения подготовки данных – полуавтоматическая разметка (Alkhudaydi et al., 2019). Р. Куи с соавторами (Qiu et al., 2022) предложили метод, где первоначальная модель обучается на слабо размеченных данных, а затем используется для автоматической генерации разметки, на которой повторно обучается новая модель.

Эти исследования демонстрируют прогресс в области автоматизации фенотипирования параметров колоса, но также указывают на главную проблему – сложность и стоимость подготовки данных. Несмотря на прогресс, задача разработки точного и робастного алгоритма, эффективно работающего с упрощенной разметкой и сохраняющего высокую точность, остается актуальной.

В настоящей работе мы исследуем альтернативный подход, основанный на использовании упрощенной точечной разметки лишь центров колосков колоса, что значительно сокращает время и стоимость подготовки данных (Chen et al., 2021) и может стать эффективным и практичным решением для широкого применения в сельскохозяйственных исследованиях.

Материалы и методы

Биологический материал и набор данных

В работе использовались колосья растений пшеницы из коллекции Н.П. Гончарова. Выборка растений включала представителей различных видов ди- ($2n = 2x = 14$), тетра- ($2n = 4x = 14$) и гексаплоидных ($2n = 6x = 42$) пшениц. Изображения колосьев были получены в лабораторных условиях согласно протоколу, описанному ранее (Генаев

и др., 2018). Для съемки использовалась цифровая камера Canon 350D, объектив EF-S 18–55 мм f/3.5–5.6. Параметры съемки: выдержка 1/160, диафрагма 11, ISO 100, фокусное расстояние 55 мм. Колос пшеницы располагался на синем фоне рядом с картой цветовой палитры X-Rite Mini ColorChecker Classic (<http://xritephoto.com/colorchecker-targets>). Использовались изображения колоса, полученные на прищепке, и колоса, расположенного на столе. Пример изображений, включенных в анализ, показан на рис. 1.

Всего набор данных включал 1745 цифровых изображений колосьев пшеницы (протокол «на столе» – 82 %, протокол «на прищепке» – 18 % изображений). Для оценки обобщающей способности моделей была взята отдельная отложенная выборка из 14 изображений, полученных по протоколу «на столе», которые не были включены нами ни в обучение, ни в тестирование алгоритмов.

Разметка колосков на изображениях

В работе было использовано три типа разметки: на основе бинарных масок, гауссовых масок и разметки на основе ограничивающих квадратов. Первоначально разметка изображений осуществлялась в программе ImageJ (Schneider et al., 2012) вручную. Точками размечались центры колосков, их координаты сохранялись в отдельном файле для каждого изображения. На основе координат центров колосков применялись три варианта генерации разметок для машинного обучения (рис. 2).

Разметка в виде бинарных масок (см. рис. 2, б). Для каждого колоска генерировалась круговая область радиусом 2 мм. В этом случае мы использовали алгоритм сегментации изображений: нейросетевая модель обучалась предсказывать области на изображении, соответствующие бинарным маскам.

Разметка в виде гауссовой маски (см. рис. 2, в). Можно предположить, что центры колосков, размеченные вручную, не соответствуют в точности их геометрическим центрам, а отклоняются случайным образом. Мы предположили, что плотность этого случайного расположения меток вокруг центра имеет вид радиального распределения Гаусса. Поэтому в качестве второго варианта разметки генерировалось двумерное гауссово распределение, центр которого совпадал с размеченным центром колоска.

Ограничивающие квадраты (см. рис. 2, г). Вокруг каждого размеченного вручную центра колоска автоматически генерировался квадрат фиксированного размера (6×6 мм). Модель обучалась детектировать эти квадраты.

Архитектуры моделей и обучение

Для обучения изображения были случайным образом разбиты на тренировочную (~60 %), валидационную (~20 %) и тестовую (~20 %) выборки.

Были задействованы два подхода к определению центров колосков на изображении с помощью алгоритмов глубокого машинного обучения. Первый заключался в решении задачи сегментации пикселей изображения, согласно которому изображения делились на принадлежащие маске и не принадлежащие маске. Для этого применили модели семантической сегментации изображений на основе сети

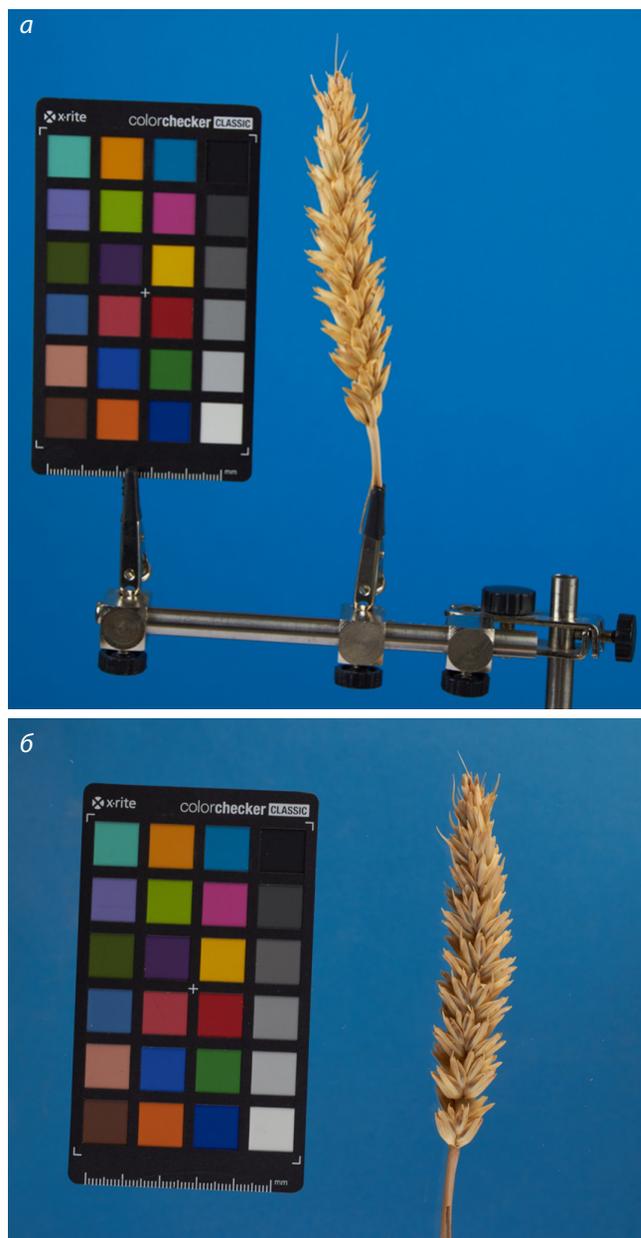


Рис. 1. Примеры изображения колоса пшеницы, закрепленного на прищепке (а) и расположенного на столе (б).

U-Net (Ronneberger et al., 2015). Использовались два типа масок из описанных выше: бинарные и гауссовы. В случае бинарной маски предсказанными центрами колосков считались геометрические центры областей, соответствующих предсказанным маскам. В случае гауссовой маски центр колоска определялся как пиксель с максимальной вероятностью внутри области, удовлетворяющей условию: отношение вероятности каждого пикселя к максимальной вероятности в данной области должно быть не ниже порога C .

Сеть U-Net (Ronneberger et al., 2015) может использовать кодировщики слоев (энкодеры) различной архитектуры. В нашей работе были протестированы несколько вариантов энкодеров: efficientnet-b3, efficientnet-b4, mit_b2, mit_b1,

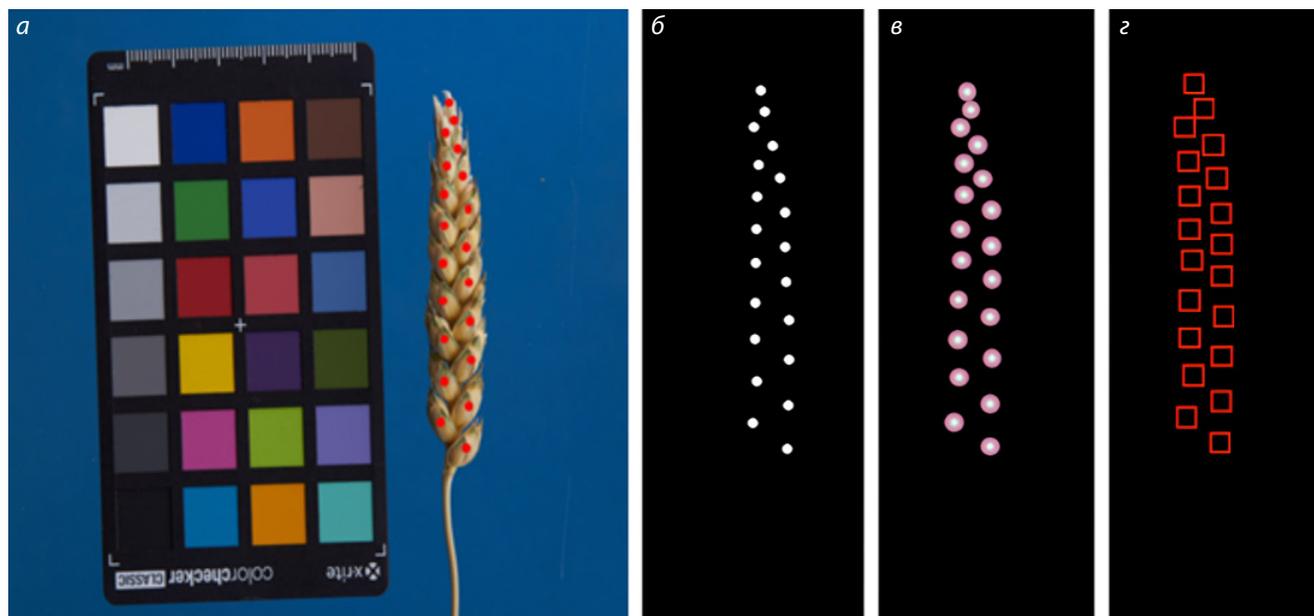


Рис. 2. Методы разметки колосков колоса на основе их центров, указанных вручную.

а – исходное изображение колоса с размеченными центрами колосков, показанными красными точками; *б* – бинарная маска для колоска в виде кругов, центры которых совпадают с разметкой центров колосков; *в* – гауссова маска, в которой центры распределения совпадают с центрами размеченных колосков, а плотность вероятности имеет радиальное распределение; *г* – набор ограничивающих квадратов, центры которых совпадают с центрами колосков.

timm-resnest26d, timm-regnetx_032, timm-res2next50, timm-gernet_m, timm-efficientnet-b4, timm-efficientnet-b3 (Tan, Le, 2019; Wightman, 2019; Radosavovic et al., 2020; Zhang H. et al., 2020; Gao et al., 2021; Xie et al., 2021).

Второй подход заключался в детекции областей изображения, ограниченных квадратами и соответствующих центральных областям колоска. Для этого использовалась архитектура сети YOLO в версии YOLOv8m, которая для каждого колоска на изображении определяет ограничивающий прямоугольник (Redmon et al., 2016).

Таким образом, мы применили для определения центров три метода: сегментацию с помощью сети U-Net для бинарных (далее обозначена как U-Net-BIN) и гауссовых (далее обозначена как U-Net-GAUSS) масок, ограничивающих положение центра и детекцию области центра колоска в виде квадрата с помощью сети архитектуры YOLO (далее обозначена как YOLOv8).

Вид функции потерь. Для алгоритма с бинарными масками использовалась функция бинарной кросс-энтропии (torch.nn.BCEWithLogitsLoss). Для гауссовых масок в качестве функции потерь использовалась дивергенция Кульбака–Лейблера (torch.nn.KLDivLoss).

Организация процесса обучения. Все три алгоритма обучались в течение 500 эпох. Веса моделей были инициализированы на основе предобученных параметров, полученных на наборе данных ImageNet (Deng et al., 2009). В качестве метода оптимизации параметров применили алгоритм Adam. В процессе обучения использовали набор аугментаций, реализованных в библиотеке albumentations (Buslaev et al., 2020). Изображение нормализовалось к размеру 512×224 (Resize), затем применялись: горизон-

тальное отражение с вероятностью 0.5 (HorizontalFlip (p=0.5)); вертикальное отражение с вероятностью 0.277 (VerticalFlip (p=0.277)); поворот на случайный угол в диапазоне $-30...+30^\circ$ с вероятностью 0.735 (Rotate(limit=30, p=0.735)); гауссово размытие с размером ядра, случайно выбираемым в диапазоне от 1 до 3, с вероятностью 0.25 (GaussianBlur(blur_limit=(1.3), p=0.25)); добавление гауссового шума с вероятностью 0.15 (GaussNoise(p=0.15)); случайная корректировка яркости и контраста с вероятностью 0.5 (RandomBrightnessContrast(p=0.5)); случайное изменение интенсивности каналов R, G и B в диапазоне ± 15 с вероятностью 0.5 (RGBShift(r_shift_limit=15, g_shift_limit=15, b_shift_limit=15, p=0.5)); цветовые преобразования – случайные изменения яркости, контраста, насыщенности и оттенка (ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.2, p=0.703)); преобразование изображения в градации серого с вероятностью 0.1 (ToGray (p=0.1)).

Исследованные нами модели нейронных сетей для идентификации центров колосков пшеницы зависели от целого ряда параметров. К ним относились: тип архитектуры энкодера для сети U-Net, радиус круговой области r для бинарных масок, порог C отношения вероятности пикселя принадлежать центру колоска к максимальной вероятности в случае гауссовой маски, параметры алгоритма оптимизации и некоторые другие. Их полный список и области значений приведены в табл. 1.

В процессе работы для алгоритмов требовалось подобрать наиболее оптимальные значения параметров из табл. 1. Для этого применен алгоритм байесовской оптимизации, реализованный в библиотеке Optuna (Akiba et

Таблица 1. Гиперпараметры, используемые в методах определения положения центра колосков на изображении

Метод	Гиперпараметры	Области значений
U-Net-BIN	Архитектура энкодера	efficientnet-b3, efficientnet-b4, mit_b2, mit_b1, timm-resnest26d, timm-regnetx_032, timm-res2next50, timm-gernet_m, timm-efficientnet-b4, timm-efficientnet-b3
	Коэффициент градиентного спуска на начальном этапе оптимизации	[0.0001, 0.001]
	Коэффициент градиентного спуска на финальном этапе оптимизации	[0.000001, 0.00005]
	Радиус бинарной маски (r, мм)	[0.5, 3]
U-Net-GAUSS	Архитектура энкодера	efficientnet-b3, efficientnet-b4, mit_b2, mit_b1, timm-resnest26d, timm-regnetx_032, timm-res2next50, timm-gernet_m, timm-efficientnet-b4, timm-efficientnet-b3
	Коэффициент градиентного спуска на начальном этапе оптимизации	[0.0001, 0.001]
	Коэффициент градиентного спуска на финальном этапе оптимизации	[0.000001, 0.00005]
	Параметр σ для гауссова распределения	[0.000002, 0.0002]
	Порог вероятности для выбора области положения центра колоска (C)	[1/9, 1/2]
YOLOv8	Коэффициент градиентного спуска на начальном этапе оптимизации	[0.01, 0.00005]
	Коэффициент градиентного спуска на финальном этапе оптимизации	[0.000005, 0.000001]
	Параметр DropBlock	[0, 0.5]

al., 2019) с использованием изображений валидационной выборки, а в качестве целевого параметра оптимизации – метрика точности F1, описанная ниже.

Метрики оценки точности алгоритмов

Эффективность определения центров колосков анализировали при помощи двух типов метрик. Количественные метрики оценивали ошибку подсчета количества колосков в колосе на изображении. Среднюю абсолютную ошибку (MAE) и абсолютную процентную ошибку (MAPE) подсчета колосков определяли согласно:

$$MAE = \frac{1}{n} \sum_{i=1}^n |t_i - p_i|,$$

$$MAPE = \frac{100}{n} \% \sum_{i=0}^n \left| \frac{t_i - p_i}{t_i} \right|,$$

где t_i – истинное число колосков на изображении, подсчитанное вручную; p_i – число колосков, определенное на основе методов машинного обучения; n – число изображений в выборке, по которым оценивается точность.

Для оценки точности определения положения центра колосков на изображении подсчитывались метрики точности (Precision), полноты (Recall) и F1-мера:

$$Precision = \frac{TP}{TP + FP},$$

$$Recall = \frac{TP}{TP + FN},$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall},$$

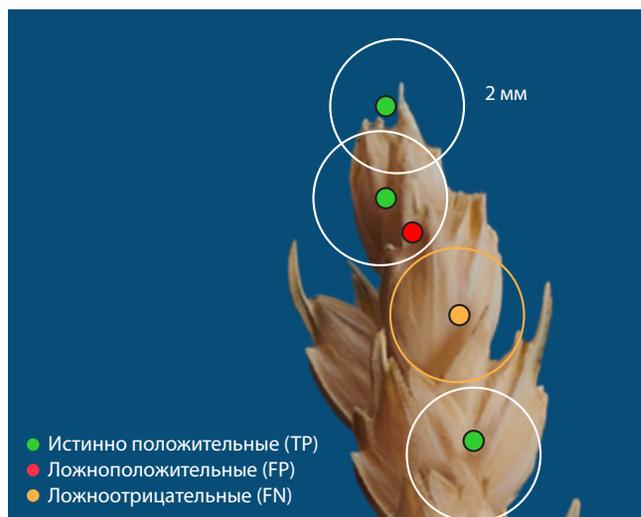


Рис. 3. Визуализация истинно положительных, ложноположительных и ложноотрицательных предсказаний при вычислении пространственных метрик оценки качества детекции колосков.

где TP (true positive) – число верных положительных предсказаний положения колоска; FP (false positive) – число неверных предсказаний положения колоска; FN (false negative) – число ложных предсказаний положения колоска. Для определения параметров TP, FP, FN использовалось следующее правило (рис. 3): предсказанное положение

центра считалось истинно положительным (TP), если находилось в пределах радиуса 2 мм от истинного центра; ложноположительным (FP), если центр колоска предсказан неверно; ложноотрицательным (FN), если в радиусе 2 мм от истинного центра нет предсказанных центров.

Программы были реализованы на языке Python 3.11 с использованием библиотек PyTorch 2.3.0, OpenCV 4.9.0 и scikit-learn 1.5.0. Расчеты выполнялись на рабочей станции с процессором AMD Ryzen 9 7950X, 64 ГБ оперативной памяти и графическим ускорителем NVIDIA GeForce RTX 2080 Ti (11 ГБ).

Результаты и обсуждение

Оптимальные значения гиперпараметров для трех моделей предсказания положения колосков на изображении, подобранные с помощью байесовской оптимизации, перечислены в табл. 2.

Наилучшей архитектурой энкодера для определения центра колоска на основе сегментации и бинарной маски оказалась MiT-B2 (Mix Transformer-B2) (Xie et al., 2021). Данная архитектура сочетает сверточные слои с механизмом самовнимания, что позволяет одновременно захватывать локальные особенности и глобальные контекстные зависимости на изображении. Это особенно важно для точного позиционирования мелких объектов, таких как центры колосков.

Для гауссовой маски наилучшие результаты показала архитектура EfficientNet-B4 (Tan, Le, 2019). Модель использует метод составного масштабирования (compound scaling), который сбалансированно увеличивает глубину, ширину и разрешение входных данных. Такой подход обеспечивает баланс между точностью и вычислительной эффективностью, что делает архитектуру хорошо подходящей для задач регрессии плотности распределения.

Результаты оценки на тестовой выборке (табл. 3) показали, что оба подхода на основе U-Net значительно превосходят метод на основе YOLOv8.

Как видно из табл. 3, оба подхода на основе архитектуры U-Net продемонстрировали высокую и сопоставимую точность как по количественным (MAE ~0.5), так и по пространственным (F1 > 0.96) метрикам на тестовой выборке. Незначительное превосходство модели U-Net-GAUSS по метрике MAE (0.502 против 0.512) и модели U-Net-BIN по метрике F1 (0.965 против 0.962) позволяет считать их одинаково эффективными для данных, распределение которых соответствует обучающей выборке. В то же время модель YOLOv8 дает значительно более низкие результаты (MAE = 3.641, F1 = 0.679), что указывает на ее непригодность для решения задачи в текущей конфигурации с автоматически сгенерированными квадратными ограничивающими рамками.

На отложенной выборке (табл. 4) модель U-Net-BIN обуславливает наилучшую точность, тогда как для модели U-Net-GAUSS точность оказалась ниже. Наибольшая ошибка наблюдается для модели YOLOv8.

Результаты на отложенной выборке, собранной в иное время и имеющей несколько иное распределение, демонстрируют более выраженные различия между методами. Модель U-Net-BIN сохранила высокую точность, показав лишь незначительный рост ошибки (MAE с 0.512 до 0.538), что свидетельствует о ее высокой обобщающей способности и надежности. В отличие от нее, точность модели U-Net-GAUSS на отложенной выборке заметно снизилась (MAE = 0.846 против 0.502 на тестовой выборке). Это позволяет предположить, что подход может быть более чувствителен к изменчивости данных. Модель YOLOv8, как и на тестовой выборке, показала наихудший результат (MAE = 3.0), дополнительно подтверждая вывод о том, что выбранный способ генерации ограничивающих квадратов не подходит для описания формы колосков.

Оба предложенных метода на основе U-Net позволили получить хорошие результаты на тестовой выборке. Однако модель, обученная на бинарных масках, продемон-

Таблица 2. Гиперпараметры для моделей нейронных сетей идентификации положения центров колосков на изображении, подобранные с помощью байесовской оптимизации

Модель	Гиперпараметры	Оптимальные значения
U-Net-BIN	Архитектура энкодера	mit_b2
	Коэффициент градиентного спуска на начальном этапе оптимизации	0.001
	Коэффициент градиентного спуска на финальном этапе оптимизации	0.000001
	Радиус бинарной маски (r, мм)	1.5999
U-Net-GAUSS	Архитектура энкодера	efficientnet-b4
	Коэффициент градиентного спуска на начальном этапе оптимизации	0.0001
	Коэффициент градиентного спуска на финальном этапе оптимизации	0.0000063723
	Параметр σ для гауссова распределения	0.00001121
	Порог вероятности для выбора области положения центра колоска (C)	0.214564
YOLOv8	Коэффициент градиентного спуска на начальном этапе оптимизации	0.00042
	Коэффициент градиентного спуска на финальном этапе оптимизации	0.000001
	Параметр DropBlock	0.48

Таблица 3. Оценка эффективности методов определения положения центра колоска на тестовой выборке

Модель	MAE	MAPE	F1
U-Net-BIN	0.512	0.032	0.965
U-Net-GAUSS	0.502	0.032	0.962
YOLOv8	3.641	0.280	0.679

Таблица 4. Оценка эффективности методов определения положения центра колоска на отложенной выборке ($n = 14$)

Модель	MAE	MAPE	F1
U-Net-BIN	0.538	0.035	0.972
U-Net-GAUSS	0.846	0.062	0.942
YOLOv8	3.000	0.231	0.704

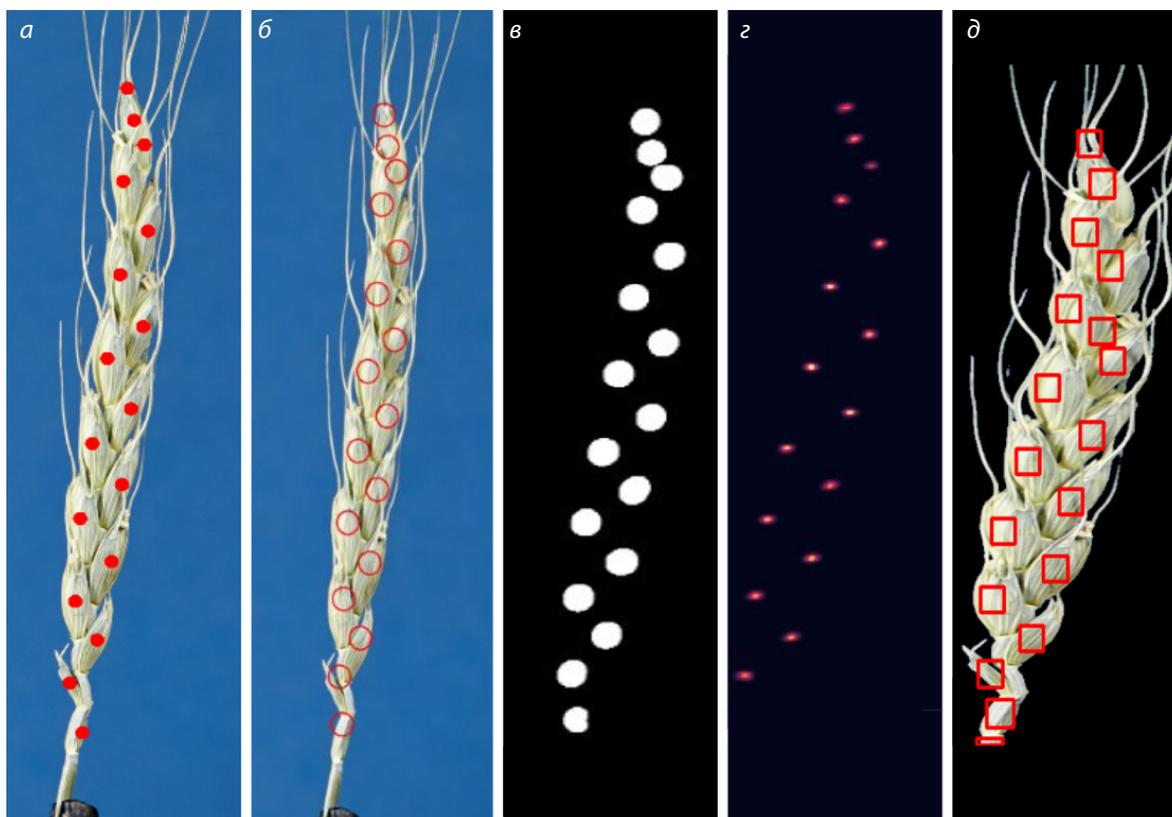


Рис. 4. Результаты работы моделей: *а* – размеченные вручную центры колосков в колесе; *б* – результат работы модели, обученной на бинарных масках, наложенный на реальное изображение колоса; *в* – результат работы модели, обученной на бинарных масках; *г* – результат работы модели, обученной на масках радиального распределения Гаусса; *д* – результат работы модели, обученной на ограничивающих квадратах.

стрировала значительно лучшую обобщающую способность на отложенной выборке, что говорит о ее большей надежности для работы с данными, имеющими иное распределение. Регрессионный анализ предсказанного и истинного числа колосков показал сильную линейную зависимость для моделей U-Net ($R^2 > 0.95$ на тестовой выборке), подтверждая высокую точность подсчета.

Измерение скорости инференса на одном графическом ускорителе показало, что время обработки одного изображения для модели, обученной на гауссовых масках, состав-

ляет примерно 0.67 с, тогда как для модели на бинарных масках – около 0.64 с. Таким образом, подход U-Net-BIN демонстрирует не только более высокую точность, но и несколько большую скорость работы.

Низкая производительность YOLOv8, вероятно, связана с неоптимальностью автоматически сгенерированных квадратных рамок для описания продолговатой и изогнутой формы колосков (рис. 4). Это указывает на то, что для методов детекции требуется более тщательная и, возможно, ручная корректировка разметки.

Заключение

Настоящее исследование демонстрирует, что использование упрощенной точечной разметки центров колосков в сочетании с современными архитектурами глубокого обучения позволяет достичь точности, сопоставимой с лучшими современными методами, использующими более сложную и дорогую разметку (Misra et al., 2020; Khoroshevsky et al., 2021; Zhou et al., 2021; Shi et al., 2023). Упрощенная схема разметки не только снижает стоимость экспериментов, но и открывает возможность быстрого расширения наборов данных. Предложенный подход может быть адаптирован для решения других задач подсчета морфологических признаков растений и интегрирован в системы автоматизированного фенотипирования для ускорения селекционных программ.

Список литературы / References

- Афонников Д.А., Генаев М.А., Дорошков А.В., Комышев Е.Г., Пшеничникова Т.А. Методы высокопроизводительного фенотипирования растений для массовых селекционно-генетических экспериментов. *Генетика*. 2016;52(7):788-803. doi 10.7868/S001667581607002X
[Afonnikov D.A., Genaev M.A., Doroshkov A.V., Komyshev E.G., Pshenichnikova T.A. Methods of high-throughput plant phenotyping for large-scale breeding and genetic experiments. *Russ J Genet*. 2016;52(7):688-701. doi 10.1134/S1022795416070024]
- Вавилова В.Ю., Конопатская И.Д., Блинов А.Г., Гончаров Н.П. Межвидовой полиморфизм генов *DEP1* и форма колоса у пшениц. *Генетика*. 2019;55(7):837-843. doi 10.1134/S0016675819070142
[Vavilova V.Yu., Konopatskaia I.D., Blinov A.G., Goncharov N.P. Interspecific polymorphism of *DEP1* genes and the spike shape in wheats. *Russ J Genet*. 2019;55(7):908-913. doi 10.1134/S1022795419070147]
- Генаев М.А., Комышев Е.Г., Хао Ф., Коваль В.С., Гончаров Н.П., Афонников Д.А. SpikeDroidDB – информационная система для аннотации морфометрических характеристик колоса пшеницы. *Вавиловский журнал генетики и селекции*. 2018;22(1):132-140. doi 10.18699/VJ18.340
[Genaev M.A., Komyshev E.G., Khao F., Koval V.S., Goncharov N.P., Afonnikov D.A. SpikeDroidDB: an information system for annotation of morphometric characteristics of wheat spike. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2018;22(1):132-140. doi 10.18699/VJ18.340 (in Russian)]
- Ефимов В.М., Речкин Д.В., Гончаров Н.П. Многомерный анализ многолетних климатических данных в связи с урожайностью, скороспелостью и проблемой глобального потепления. *Вавиловский журнал генетики и селекции*. 2024;28(2):155-165. doi 10.18699/vjgb-24-18
[Efimov V.M., Rechkin D.V., Goncharov N.P. Multivariate analysis of long-term climate data in connection with yield, earliness and the problem of global warming. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov J Genet Breed*. 2024;28(2):155-165. doi 10.18699/vjgb-24-18]
- Маслова Г.Я., Абдраев М.Р., Шараров И.И., Шарарова Ю.А. Корреляционный анализ урожайности и элементов продуктивности сортов озимой мягкой пшеницы в засушливых условиях лесостепной зоны среднего Поволжья. *Известия Самарского научного центра Российской академии наук*. 2018;20(2-4):680-683. doi 10.24411/1990-5378-2018-00153
[Maslova G., Abdryayev M.R., Sharapov I.I., Sharapova Ju.A. Correlation analysis of yield and elements of productivity of winter soft wheat varieties in the arid conditions of steppe zone of Middle Volga region. *Izvestiya of Samara Scientific Center of the Russian Academy of Sciences*. 2018;20(2-4):680-683. doi 10.24411/1990-5378-2018-00153 (in Russian)]
- Савин В.Ю. Определение усилий, необходимых для очеса колоса пшеницы. *Инженерные технологии и системы*. 2019;29(3):456-466. doi 10.15507/2658-4123.029.201903.456-466
[Savin V.Yu. Determination of the force required for stripping the wheat ear. *Inzhenernyye Tekhnologii i Sistemy = Engineering Technologies and Systems*. 2019;29(3):456-466. doi 10.15507/2658-4123.029.201903.456-466 (in Russian)]
- Скрипка О.В., Самофалов А.П., Подгорный С.В., Громова С.Н. Урожайность и основные элементы продуктивности у сортов озимой пшеницы интенсивного типа селекции ВНИИЗК. *Достижения науки и техники АПК*. 2016;30(9):30-32.
[Skripka O.V., Samofalov A.P., Podgorniy S.V., Gromova S.N. Productivity and main elements of productivity of winter wheat varieties of intensive type, bred in the All-Russian Research Institute of Grain Crops. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-Industrial Complex*. 2016;30(9):30-32 (in Russian)]
- Akiba T., Sano S., Yanase T., Ohta T., Koyama M. Optuna: a next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA. ACM, 2019;2623-2631. doi 10.1145/3292500.3330701
- Alkhudaydi T., Zhou J., De La Iglesia B. SpikeletFCN: counting spikelets from infield wheat crop images using fully convolutional networks. In: Artificial Intelligence and Soft Computing. ICAISC 2019. Lecture Notes in Computer Science. Vol. 11508. Springer, Cham, 2019;3-13. doi 10.1007/978-3-030-20912-4_1
- Artemenko N.V., Epifanov R.U., Genaev M.A., Komyshev E.G., Kruchinina Yu.V., Koval V.S., Goncharov N.P., Afonnikov D.A. Image-based classification of wheat spikes by glume pubescence using convolutional neural networks. *Front Plant Sci*. 2024;14:1336192. doi 10.3389/fpls.2023.1336192
- Buslaev A., Igloukov V.I., Khvedchenya E., Parinov A., Druzhinin M., Kalinin A. Alumentations: fast and flexible image augmentations. *Information*. 2020;11(2):125. doi 10.3390/info11020125
- Chen F., Pound M.P., French A.P. Learning to localise and count with incomplete dot-annotations. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada. IEEE, 2021;1612-1620. doi 10.1109/ICCVW54120.2021.00186
- Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009, Miami. IEEE, 2009;248-255. doi 10.1109/CVPR.2009.5206848
- Gao S.-H., Cheng M.-M., Zhao K., Zhang X.-Y., Yang M.-H., Torr P. Res2Net: a new multi-scale backbone architecture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021;43(2):652-662. doi 10.1109/TPAMI.2019.2938758
- Genaev M.A., Komyshev E.G., Smirnov N.V., Kruchinina Y.V., Goncharov N.P., Afonnikov D.A. Morphometry of the wheat spike by analyzing 2D images. *Agronomy*. 2019;9(7):390. doi 10.3390/agronomy9070390
- Khoroshevsky F., Khoroshevsky S., Bar-Hillel A. Parts-per-object count in agricultural images: solving phenotyping problems via a single deep neural network. *Remote Sens*. 2021;13(13):2496. doi 10.3390/rs13132496
- Konopatskaia I., Vavilova V., Blinov A., Goncharov N.P. Spike morphology genes in wheat species (*Triticum* L.). *Proc Latv Acad Sci B Nat Exact Appl Sci*. 2016;70(6):345-355. doi 10.1515/prolas-2016-0053
- Li Q., Cai J., Berger B., Okamoto M., Miklavcic S.J. Detecting spikes of wheat plants using neural networks with Laws texture energy. *Plant Methods*. 2017;13:83. doi 10.1186/s13007-017-0231-1
- Liu T., Chen W., Wang Y., Wu W., Sun C., Ding J., Guo W. Rice and wheat grain counting method and software development based on

- Android system. *Comput Electron Agric.* 2017;141:302-309. doi 10.1016/j.compag.2017.08.011
- Misra T., Arora A., Marwaha S., Chinnusamy V., Rao A., Jain R., Narayan R., Ray M., Kumar S., Raju D., Ranjan R., Nigam A., Goel S. SpikeSegNet – a deep learning approach utilizing encoder-decoder network with hourglass for spike segmentation and counting in wheat plant from visual imaging. *Plant Methods.* 2020;16:40. doi 10.1186/s13007-020-00582-9
- Qiu R., He Y., Zhang M. Automatic detection and counting of wheat spikelet using semi-automatic labeling and deep learning. *Front Plant Sci.* 2022;13:872555. doi 10.3389/fpls.2022.872555
- Radosavovic I., Kosaraju R.P., Girshick R., He K., Dollár P. Designing network design spaces. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, Seattle, WA, USA. IEEE, 2020;10428-10436. doi 10.1109/CVPR42600.2020.01044
- Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, Las Vegas, NV, USA. IEEE, 2016;779-788. doi 10.1109/CVPR.2016.91
- Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science. Vol. 9351. Springer, 2015;234-241. doi 10.1007/978-3-319-24574-4_28
- Schneider C.A., Rasband W.S., Eliceiri K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9(7):671-675. doi 10.1038/nmeth.2089
- Shi L., Sun J., Dang Y., Zhang S., Sun X., Xi L., Wang J. YOLOv5s-T: a lightweight small object detection method for wheat spikelet counting. *Agriculture.* 2023;13(4):872. doi 10.3390/agriculture13040872
- Tan M., Le Q.V. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, California, USA. ICML, 2019;6105-6114
- Vahamidis P., Karamanos A.J., Economou G. Grain number determination in durum wheat as affected by drought stress: an analysis at spike and spikelet level. *Ann Appl Biol.* 2019;174(2):190-208. doi 10.1111/aab.12487
- Vavilova V., Konopatskaia I., Kuznetsova A.E., Blinov A., Goncharov N.P. Genomic characterization of *DEP1* gene in wheats with normal and compact spike shape. *BMC Genetics.* 2017;18 (Suppl.1): 106. doi 10.1186/s12863-017-0583-6
- Wightman R. PyTorch Image Models (timm). *GitHub Repository.* 2019. doi 10.5281/zenodo.4414861
- Xie E., Wang W., Yu Z., Anandkumar A., Alvarez J., Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers. *arXiv.* 2021. doi 10.48550/arXiv.2105.15203
- Zhang B., Liu X., Xu W., Chang J., Li A., Mao X., Zhang X., Jing R. Novel function of a putative *MOC1* ortholog associated with spikelet number per spike in common wheat. *Sci Rep.* 2015;5(1):12211. doi 10.1038/srep12211
- Zhang H., Wu C., Zhang Z., Zhu Y., Zhang Z., Lin H., Sun Y., He T., Mueller J., Manmatha R., Li M., Smola A. ResNeSt: split-attention networks. *arXiv.* 2020. doi 10.48550/arXiv.2004.08955
- Zhou H., Riche A.B., Hawkesford M.J., Whalley W.R., Atkinson B.S., Sturrock C.J., Mooney S.J. Determination of wheat spike and spikelet architecture and grain traits using X-ray Computed Tomography imaging. *Plant Methods.* 2021;17(1):26. doi 10.1186/s13007-021-00726-5

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 10.09.2025. После доработки 23.10.2025. Принята к публикации 01.11.2025.