

УДК 004.7

DOI <https://doi.org/10.32782/IT/2023-1-10>

**Олександр МЕЛЬНИЧЕНКО**

аспірант, Хмельницький національний університет, вул. Інститутська, 11, м. Хмельницький, Україна, 29016, [oleksandr.melnychenko@live.com](mailto:oleksandr.melnychenko@live.com)

ORCID: 0000-0001-8565-7092

**Бібліографічний опис статті:** Мельниченко, О. (2023). Методи розпізнавання та обробки зображень за модифікованою YOLOv5-v1. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 74–84, doi: <https://doi.org/10.32782/IT/2023-1-10>

## МЕТОДИ РОЗПІЗНАВАННЯ ТА ОБРОБКИ ЗОБРАЖЕНЬ ЗА МОДИФІКОВАНОЮ YOLOV5-V1

У поданій роботі запропоновано новий підхід згідно глибокого навчання до виявлення, розпізнавання, обробки зображень та обчислення кількості об'єктів, зокрема на прикладі обробки зображень яблук. Відому модель YOLOv5 було взято за основу через її високу точність та швидкість часу обробки даних. YOLOv5 було модифіковано, відповідно до вимог завдання виявлення яблук у робочому середовищі фруктового саду. Запропонований підхід складається із двох етапів: виявлення яблук і обчислення їхньої кількості. На етапі виявлення нова модель YOLOv5-v1 була навчена на створеному вручну наборі даних із зображень яблук для виявлення ознак, які відрізняють яблука від зовнішнього фону. Модель містить нові рівні для модуля BottleneckCSP-v4, замінюючи модуль BottleneckCSP в оригінальній магістральній архітектурі YOLOv5. На етапі обрахування модуль SENet інтегрується в розширену магістральну мережу, щоб краще ідентифікувати ознаки плодів середнього та великого розміру в різних умовах. Початковий розмір прив'язувального блоку початкової мережі було скориговано, щоб запобігти помилковій ідентифікації малих об'єктів на фоні зображення та, таким чином, підвищити точність обчислення кількості. Ефективність підходу було оцінено в результаті проведених обчислювальних експериментів із набором даних зображень яблук. Експериментальні результати на тестовому наборі даних продемонстрували, що вдосконалена модель може ефективно розпізнавати та обчислювати фрукти, зняті відеокамерою безпілотною літальною апарату, із показниками recall, precision, mAP та F1 92,13%, 84,59%, 87,94% та 89,02% відповідно. Запропонований підхід також було порівняно з іншими сучасними методами, такими як YOLOv5, YOLOv3, YOLOv4 та EfficientDet-D0, і встановлено, що запропонована модель перевершила аналоги за точністю та швидкістю. Середня швидкість розпізнавання запропонованої мережі склала 0,015 секунди на 1 кадр відеоряду (66,7 кадрів/с), що у 2,53, 1,13 та 3,53 рази вище, ніж у мережах EfficientDet-D0, YOLOv4 і YOLOv3, відповідно. Отримані результати мають декілька потенційних застосувань у садівництві, де їх можна використовувати для моніторингу посівів, оцінювання врожайності та контролю якості продукції. Подальші дослідження, також, можуть бути проведені для включення додаткових ознак, таких як форма фрукта, використовуючи більший набір даних для навчання моделі.

**Ключові слова:** обробка зображень, виявлення об'єктів, врожайність яблук, YOLOv5, глибоке навчання.

**Oleksandr MELNYCHENKO**

Postgraduate Student, Khmelnytskyi National University, 11, Instytuts'ka str., Khmelnytsky, Ukraine, 29016, [oleksandr.melnychenko@live.com](mailto:oleksandr.melnychenko@live.com)

ORCID: 0000-0001-8565-7092

**To cite this article:** Melnychenko, O. (2023). Metody rozpoznavannya ta obrobky zobrazhen' za modyfikovanoyu YOLOv5-v1 [Image recognition and processing methods according to the modified YOLOv5-v1]. *Information Technology: Computer Science, Software Engineering and Cyber Security*, 1, 74–84, doi: <https://doi.org/10.32782/IT/2023-1-10>

## IMAGE RECOGNITION AND PROCESSING METHODS ACCORDING TO THE MODIFIED YOLOV5-V1

This study proposes a novel deep learning approach for apple detection and counting. The well-recognized YOLOv5 model was employed as a baseline for its high accuracy and fast processing time and further modified to suit the requirements of the apple detection task in the natural environment. The proposed approach involves two stages: apple detection and apple counting. In the detection stage, the novel YOLOv5-v1 model was trained on a manually crafted dataset of apple images to learn the features that distinguish apples from the background. The model contained new layers for the BottleneckCSP-v4 module, replacing the BottleneckCSP module in the original YOLOv5 network's backbone architecture. In the counting stage,

*the SENet module is integrated into the enhanced trunk network to better identify the features of medium and large-sized fruits under various conditions. The initial size of the source network's binding block was adjusted to prevent the misidentification of small objects in the image background and thus improve counting accuracy. To evaluate the performance of the approach, computational experiments were conducted on a dataset of apple images constructed by the authors. The experimental results on the test dataset demonstrated that the improved model could effectively recognize and count fruits captured by the unmanned aerial vehicle camera with recall, precision, mAP, and F1-score of 92.13%, 84.59%, 87.94%, and 89.02%, respectively. The proposed approach was also compared with several other state-of-the-art methods, like YOLOv5, YOLOv3, and YOLOv4 and EfficientDet-D0, and it was found that our model outperformed the analogs in terms of accuracy and speed. The average recognition speed of our model was 0.015 seconds per 1 frame of the video sequence (66.7 frames/s), which was 2.53, 1.13, and 3.53 times higher than that of the EfficientDet-D0, YOLOv4, and YOLOv3 networks, respectively. The obtained results have several potential applications in the agriculture industry, where they can be used for crop monitoring, yield estimation, and quality control. Further research can also be conducted to incorporate additional features, such as fruit shape, using a larger dataset to train the model.*

**Key words:** image processing, object detection, apple yield, YOLOv5, deep learning.

**Вступ.** Яблука є однією з найбільших фруктових культур у світі, а Україна має найбільшу площу плантацій, що займає близько двох мільйонів гектарів і дає найвищий урожай у Європі. В центральних та південних регіонах України є чудові кліматичні та ґрунтові умови, сприятливі для вирощування яблук (Medvedeva et al., 2021). Крім того, технологічні вдосконалення сільськогосподарської практики в Україні призвели до збільшення врожайності яблук і покращення якості.

Виявлення врожайності яблук за допомогою візуального розпізнавання є ефективним та інтуїтивно зрозумілим методом. Однак, оскільки кожне плодове дерево має різний ріст і кількість плодів, для досягнення більшої точності потрібне індивідуальне визначення врожайності (Suresh Kumar & Mohan, 2023). У садах із великою кількістю дерев ефективність має вирішальне значення. Отже, необхідно мати модель виявлення яблук, яка є швидкою, точною та малою, що дозволяє швидше виявляти врожайність яблук під час роботи на різних вбудованих пристроях.

**Аналіз останніх досліджень і публікацій.** Раніше для виявлення фруктів зазвичай використовували звичайні методи зору, такі як обробка зображень і машинне навчання. Ці методи розпізнавали фрукти за кольором, формою та текстурою. Наприклад, в роботі (Yu et al., 2021) досягли рівня точності 89,92% для зеленого лічі та 94,50% для червоного лічі, навчивши модель розпізнавати плід лічі за ознаками кольору та текстури. Подібним чином в роботі (Wan Nurazwin Syazwani et al., 2022) використовували машинне навчання для виявлення зображень крони ананаса, досягнувши точності 94,4% для підрахунку плодів. У роботі (Fu et al., 2019) поєднали функції текстури з машинним алгоритмом опорних векторів для виявлення бананів із середнім рівнем виявлення 89,63%. Однак ці методи, що ґрунтуються на обробленні

зображень, мають недоліки, зокрема низьку швидкість виявлення, низьку точність і обмежену адаптованість до змінного середовища освітлення саду.

Моделі глибокого навчання (ГН), зокрема згорткові нейронні мережі (CNN), мають можливість витягувати характеристики із зображень (Radiuk & Hryprynska, 2020), що може призвести до автоматичного розпізнавання цілей і кращої адаптивності. Зокрема, існує два типи методів виявлення цілей ГН (Li et al., 2021), які можуть бути корисними для виявлення яблук: двоетапний і одноетапний алгоритми виявлення цілей.

Двоетапний підхід до виявлення цілей, що включає R-CNN, SPP і Fast R-CNN, складається з двох гілок мережі: генерації пропозицій регіону (RPN) (Tang et al., 2018) і модуля класифікації (Zhao et al., 2020). RPN пропонує область інтересу (ROI) для класу переднього плану, а модуль класифікації призначає та оцінює обмежувальну рамку для кожної ROI (Radiuk et al., 2022). Однак цей метод навряд чи підходить для вбудованих пристроїв через низьку швидкість виявлення, спричинену великим розміром моделі. Наприклад, Mai et al. (Mai et al., 2020) запропонували структуру Faster R-CNN для виявлення багатокатегорійних фруктів, яка досягла середньої точності 90,72% і часу виявлення 58 мілісекунд (мс) на зображення.

Порівняно з двоетапним підходом, одноетапний підхід виявлення цілей, який включає серію You Only Look Once (YOLO) (Wang, 2021) і модель Single-Shot Detector (SSD) (Bresilla et al., 2019), має більш збалансовану продуктивність у терміни виявлення швидкість і точність. Це призводить до загальної кращої продуктивності моделі. Наприклад, Huang et al. (Huang et al., 2021) запропонували покращений метод виявлення YOLOv3 для недозрілих яблук у саду та отримали точність 61,6. У дослідженні (Chen et al., 2021) досягли точності розпізнавання яблук 97,13% у складних садах за допомогою

YOLOv4 із середнім часом розпізнавання 16,69 мс на одне зображення на одному графічному процесорі.

Карликові та веретеноподібні дерева мають менш щільне листя, тобто менше яблук приховано від огляду. Тим не менш, деякі яблука можуть бути закриті гілками або листям, а освітлення на задньому плані може бути нерівномірним або складним (Behera et al., 2018). Щоб яблука мали більш рівномірний колір на етапі забарвлення плодів, українські аграрії, що вирощують яблука, можуть використовувати світловідбиваючу плівку для покриття землі між рядами дерев (Колокольчикова, 2019). Однак це може призвести до недостатнього освітлення для захоплення зображення на стороні фонового освітлення фруктових дерев, що ускладнює виявлення яблук на зображенні підсвічування, знятому камерою. Попередні дослідження, як-от (Huang et al., 2021; Chen et al., 2021; Behera et al., 2018; Wang et al., 2021) не встановлювали однакові пріоритети щодо точності виявлення, швидкості та розміру моделі. Задля вирішення вище вказаних проблем у цьому дослідженні пропонується нове рішення під назвою YOLOv5-v1, який об'єднує YOLOv5 із легкою мережею EfficientNet (Tan & Le, 2019).

Автоматичне виявлення фруктів на деревах відбувається в складному сітчастому середовищі, де фрукти ростуть на деревах, розташованих послідовними рядами. Під час робочої місії фрукти можуть бути зняті відеокамерою безпілотного літального апарату (БПЛА) та подані на передньому плані робочого сегмента (фруктове дерево), на задньому плані або на деревах інших рядів (Vinci et al., 2023). Як наслідок, фрукти у відеокадрі можуть бути занадто малими, через що автоматизована система може їх помилково розпізнати, або вони можуть потрапити в різні відеокадри під час робочої місії, що призведе до кількох підрахунків однакових фруктів. Крім того, мінливі погодні умови, зміна освітлення протягом дня, падіння листя та гілок у кадр можуть погіршити точність виявлення плодів. Тому, для усунення згаданих вище проблем у цій роботі пропонується новий підхід до виявлення фруктових плодів у вигляді структурних об'єктів. Мета роботи полягає у вдосконаленні методу ідентифікації яблук як структурних об'єктів на деревах та розробленні нового підходу до обрахування кількості цих об'єктів.

Методологія дослідження. Вхідними даними запропонованого у роботі підходу є цифрове зображення структурних об'єктів, представлене у вигляді матриці чисел, кожне з яких відо-

бражає інтенсивність відповідного пікселя на зображенні. Модель цифрового зображення структурних об'єктів представимо у вигляді лінійного представлення:

$$f: I \rightarrow \langle C_{xy}, P_{h \times w}^i, B_{u \times v}^i, B_{h \times w}^i \rangle, \quad (1)$$

де  $C_{xy} = (c_x, c_y)$  – ширина  $c_x$  і висота  $c_y$  координатної сітки  $C_{xy}$  координатної площини XY, сформованої для виявлення структурних об'єктів;  $P_{h \times w}^i = (p_h^i, p_w^i)$  – висота  $p_h^i$  та ширина  $p_w^i$  рамки прив'язки  $i$ -го цільового структурного об'єкта в координатній площині XY;  $B_{u \times v}^i = (b_u^i, b_v^i)$  – координати центру рамки прив'язки  $i$ -го цільового структурного об'єкта в координатній площині XY;  $B_{h \times w}^i = (b_h^i, b_w^i)$  – висота  $b_h^i$  та ширина  $b_w^i$  обмежувальної рамки, що окреслює  $i$ -й цільовий структурний об'єкт у координатній площині XY.

Координати центра обмежувальної рамки  $i$ -го цільового конструктивного об'єкта в координатній площині XY розраховуються за формулою:

$$\begin{aligned} b_u^i &= \sigma^i(t_x^i + c_x); \\ b_v^i &= \sigma^i(t_y^i + c_y), \end{aligned} \quad (2)$$

де  $t_x^i$  та  $t_y^i$  – зміщення центру обмежувальної рамки  $i$ -го цільового структурного об'єкта відносно координатної сітки  $C_{xy}$ .

Висота ( $b_h^i$ ) і ширина ( $b_w^i$ ) обмежувальної рамки  $i$ -го цільового конструктивного об'єкта в координатній площині XY розраховуються за формулою:

$$\begin{aligned} b_h^i &= p_h^i e^{c_y}; \\ b_w^i &= p_w^i e^{c_x}. \end{aligned} \quad (3)$$

За механізм накладання прив'язувальної та обмежувальної рамок на цільовий об'єкт взято нейромережеву модель на основі архітектури згорткової нейронної мережі з назвою YOLOv5, до якої додано вдосконалення (рис. 1).

Використання автоматизованої системи в реальних умовах передбачає три основні сценарії розміщення структурних об'єктів на зображенні: а) на передньому плані поточного робочого сегмента, б) на задньому плані поточного робочого сегмента, в) на іншому робочому сегменті, що потрапив у кадр. Відстань між плодами на задньому плані, які розміщені на деревах у віддаленому ряду посадки, та БПЛА може бути вкрай великою, що може негативно вплинути на якість виявлення структурних об'єктів. Тому, відповідні їм структурні об'єкти на зображенні варто вважати слабо прийнятними цілями для виявлення та обчислення кількості. Щоб запобігти помилковому виявленню структурних об'єктів на задньому плані відеокадру та підвищити точність їхнього виявлення на передньому плані, початкові розміри прив'язувальних рамок

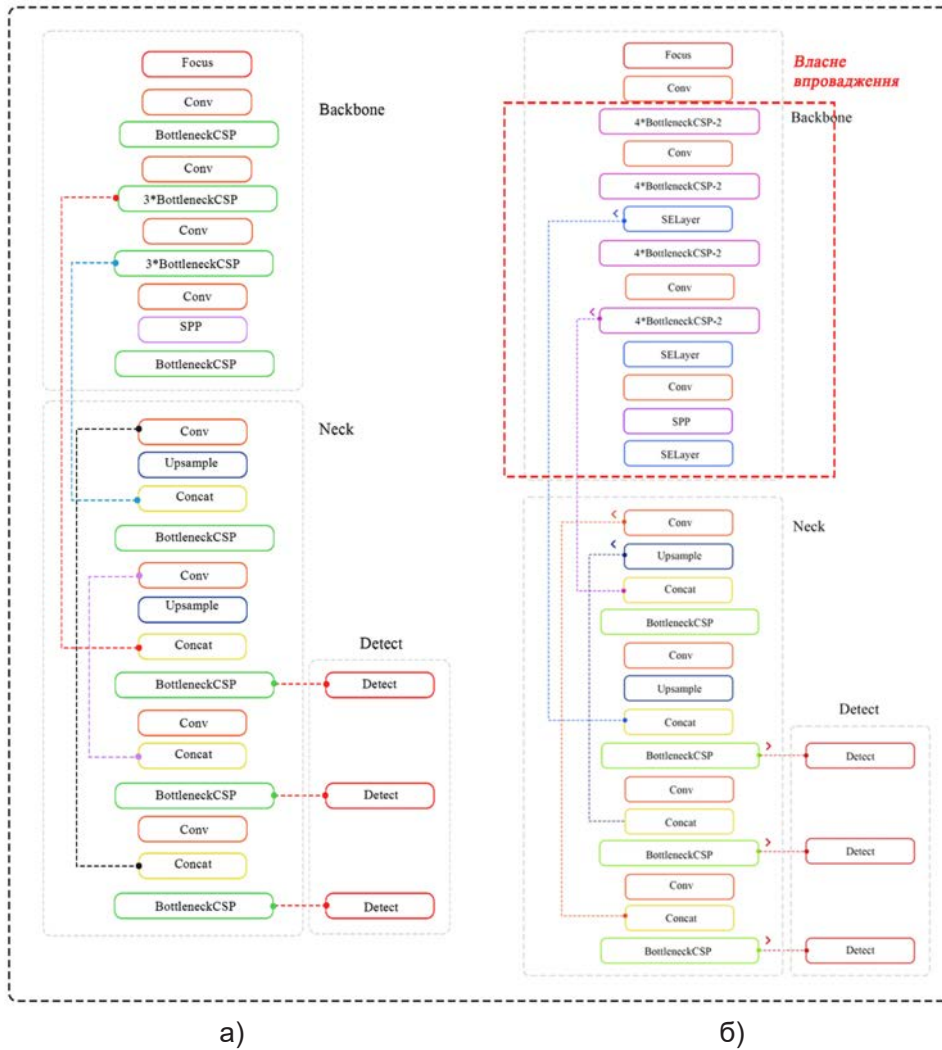


Рис. 1. Порівняння схем класичної архітектури YOLOv5 (а) та вдосконаленої архітектури YOLOv5-v1 (б)

Таблиця 1  
Значення початкових розмірів обмежувальних та прив'язувальних рамок вдосконаленої архітектури YOLOv5-v1

	Карта ознак 1		Карта ознак 2		Карта ознак 3	
	$p_h$	$p_w$	$p_h$	$p_w$	$p_h$	$p_w$
Обмежувальні рамки	20	20	40	40	80	80
Прив'язувальні рамки, малі	80	70	75	75	85	100
Прив'язувальні рамки, середні	95	110	130	110	115	125
Прив'язувальні рамки, великі	116	90	156	198	373	326

малого та середнього масштабу було змінено, з огляду на умови робочого середовища та результати обчислювальних експериментів. Відношення довжини до ширини початкових розмірів обмежувальної рамки та відношення довжини до ширини початкових розмірів прив'язувальної рамки встановлено приблизно 1 до 1. Значення початкових розмірів обмежувальної та прив'язувальної рамок вдосконаленої архітектури YOLOv5-v1 подано в таблиці 1.

Передбачається, що подання початкових розмірів обох типів рамок у таблиці 1 дасть змогу підвищити точність виявлення структурних об'єктів на передньому плані та знизити ймовірність відстеження нецільових об'єктів на задньому плані.

Для досягнення високої точності виявлення структурних об'єктів у режимі реального часу фізичний розмір оригінальної нейромережевої моделі YOLOv5 має бути мінімальним, а струк-

тура нейронної мережі – відповідно максимально стиснутою. Тому, магістральний блок нейронної мережі, який вилучає ознаки цільового об'єкта на вхідних зображеннях, було модифіковано через зменшення кількості вагових параметрів нейронної мережі. Вдосконалення магістрального блоку полягає в наступних змінах.

1. Модифіковано модуль фокусування нейронної мережі для зменшення часу навчання нейронної мережі, а саме:

1) вхідне зображення, що подане трьома каналами RGB та розміром  $3 \times 640 \times 640$ , було розбито на 4 частини розмірами  $3 \times 320 \times 320$  кожна;

2) 4 частини зображення об'єднано в одну карту ознак за RGB каналом із результируючим розміром  $12 \times 320 \times 320$ ;

3) до отриманої карти ознак розміром  $12 \times 320 \times 320$  застосовано операцію згортки із 32 ядрами, яка генерує карту ознак розміром  $32 \times 320 \times 320$ ;

4) до отриманої карти ознак розміром  $32 \times 320 \times 320$  застосовано операцію пакетної нормалізації та активаційну функцію Hardwish; вихідна карта ознак розміром  $32 \times 320 \times 320$  передається на наступний етап оброблення нейронної мережі.

2. Модуль BottleneckCSPv1. Вдосконалення модуля BottleneckCSPv1 в архітектурі YOLOv5-v1 (див. рис. 1) полягає у видаленні згорткового шару, що в оригінальному модулі BottleneckCSP (Zhou et al., 2020) поєднував вхідну карту ознак з операцією конкатенації.

3. Об'єднання карт ознак. Щоби підвищити точність виявлення структурних об'єктів на відеокадрів, до архітектури YOLOv5-v1 застосовано операцію об'єднання карт ознак різних розмірів в одну карту ознак із вищою дискримінаційною здатністю. З огляду на спрощену структуру модифікованого модуля BottleneckCSPv1 та об'єднання карт ознак, що виводяться зі згорткових шарів вдосконаленої архітектури YOLOv5-v1, об'єднання шарів 4 і 15, 6 і 11, 10 і 21 оригінальної архітектури YOLOv5 було замінено на об'єднання шарів 5 і 18, 8 і 14, 13 і 24 вдосконаленої архітектури YOLOv5-v1.

В результаті виконаних модифікацій отримано вдосконалений метод, який полягає в автоматизованому перетворенні вхідної інформації, що подана як цифрове зображення структурних об'єктів однієї природи (плодів фруктових дерев), у результируючу інформацію у вигляді множини координат, що відповідають розташуванню цих об'єктів на зображенні.

Підготовка робочого середовища та навчальних даних. Для підготовки та форму-

вання навчальних даних було використано робоче середовище сучасного фруктового саду, яке знаходиться в Хмельницькій області, Городоцького району за координатами 49.265011, 26.548304. Відстань між рядами яблунь робочого середовища становить 4 м, відстань між плодовими деревами – близько 1 м, а висота дерев – близько 2 м. Також в експериментальному полі встановлено базову RTK-станцію (Kim & Bae, 2019), яка складається з GNSS-приймача (Dąbrowski et al., 2021), супутникової антени, що посиляє через GPRS-з'єднання сигнал для підвищення точності звичайного супутникового позиціонування до 2.5 см.

Для навчання нейронних мереж використано фруктові плоди яблуні Фуджі, що покриті штрихованим малиновим рум'янцем із жовтими підшкірними цятками та мають округлену форму. Оригінальні зображення яблуневих плодів було отримано в сонячні та похмурі дні. Для отримання зображень було використано камеру смартфона iPhone 13 Pro Max (Dempsey, 2021) за різних погодних умов та на різних відстанях відносно дерев (0,5-1,5 м).

Перших 200 зображень (100 сонячних днів і 100 похмурих днів) були випадковим чином відібрані з 1 214 зображень, як тестовий набір даних, а решта 1 014 зображень – як навчальний набір. У таблиці 2 наведено розподіл створеного набору даних зображень плодів.

Таблиця 2

#### Детальна інформація про попередню підготовку даних

Дані	Сонячні	Похмурі	Загалом
Кількість зображень	100	100	200
Зображення із яскраво вираженими плодами	482	525	1007
Зображення із змішаною оклюзією	766	563	1329

До відзнятих зображень застосовано механізми приросту та покращення даних з використанням технології CLaID.AI (Claid.AI, 2023). Також були добавлені зображення яблуневих плодів інших кольорів та форм. Фінальний набір даних мав 16 224 зображень плодів, які використовуються як дані остаточного навчального набору для навчання моделі розпізнавання, включаючи 15 210 покращених зображень і 1014 необроблених зображень. Навчальний та тестовий набори даних не містили збігів. Також в роботі використано сервіс Amazon SageMaker Data Labeling (Mishra, 2019) для автоматичного маркування даних. В результаті було відібрано

200 тестових зображень, які були позначені мітками у ручному режимі. Решта зображень отримувала мітки вже в автоматичному режимі.

Навчання нейронної мережі та експериментальне тестування системи виконано з допомогою апаратного забезпечення: процесор Intel (R) Core (TM) i7-9750H, 2,6 ГГц, 32 ГБ оперативної пам'яті; графічний процесор NVIDIA GeForce RTX 2060, 6 ГБ відеопам'яті. Використане програмне забезпечення включало мову програмування Python, програмний комплекс PyTorch та пакет інструментів CUDA, cudnn, OpenCV під операційною системою Windows 11.

Створення детектора. За метрики оцінювання побудованого детектора взято: 1) precision; 2) recall; 3) середня точність класифікації (mAP); 4) F1-міра. Результати навчання нейронної мережі за вдосконаленою архітектурою YOLOv5-v1 подано на рис. 2.

Значення функцій втрат за валідаційним набором даних і значення функцій втрат за навчальним набором даних показані кривими на рис. 2, вказують на те, що значення функції втрат швидко зменшуються на перших 100 ітераціях навчання нейронної мережі, при цьому більш стабільна тенденція отримується після 250 ітерацій навчання. Таким чином, вихідна модель нейронної мережі утворюється після 300 ітерацій навчання. Показані вище значення діаграми на рис. 2 показують, що модель нейронної мережі добре навчена без перенавчання.

Модель нейронної мережі після виявлення певної множини структурних об'єктів на відеокадри, виконує фільтрування за допомогою попередньо встановленого значення порогу вірогідності прогнозування. Влучність та повнота результатів виявлення різняться беручи за основу

одну модель розпізнавання, але різне значення порогів ймовірностей розпізнавання. Якщо поріг достовірності моделі не було встановлено належним чином, то отримуються непередбачувані результати, як показано на рис. 3.

Об'єкти, які були виявлені на передньому фоні вважаються помилковими (позначено жовтим еліпсом на рис. 3а), у випадку якщо значення порогу вірогідності було встановлено занадто низько, при цьому цільовий структурний об'єкт на передньому плані не буде виявлено, або з помилками (позначено жовтим еліпсом на рис. 3б), якщо значення порогу ймовірності буде встановлено надто високим. Значення порогу достовірності визначено на основі таких сценаріїв: а) порогове значення достовірності було встановлено нижче 0.5. У цьому випадку значення влучності розпізнавання була низькою та становило менше, а ніж 80%; б) порогове значення достовірності було встановлено вище 0.5. У цьому випадку значення середньої точності класифікації поступово знижувалася; в) порогове значення ймовірності було встановлено 0.5.

Результати експериментальних досліджень. Перевірка продуктивності та ефективності моделі нейронної мережі у задачі по розпізнаванню образів у режимі реального часу на основі вдосконаленого дизайну YOLOv5-v1, були додатково проаналізовані на результатах виявлення використовуючи 200 зображень тестового набору. Всього є 2336 зображень з фруктами, серед яких цільова кількість, які можна захопити, становить 1007, а цільова кількість фруктів, які неможливо захопити, – 1329. Специфічні результати розпізнавання методу, запропонованого в дослідженні, показано в таблиці 3, яка вказує на те, що для фруктів, які можна захопити, точність, запам'ятовування, значення

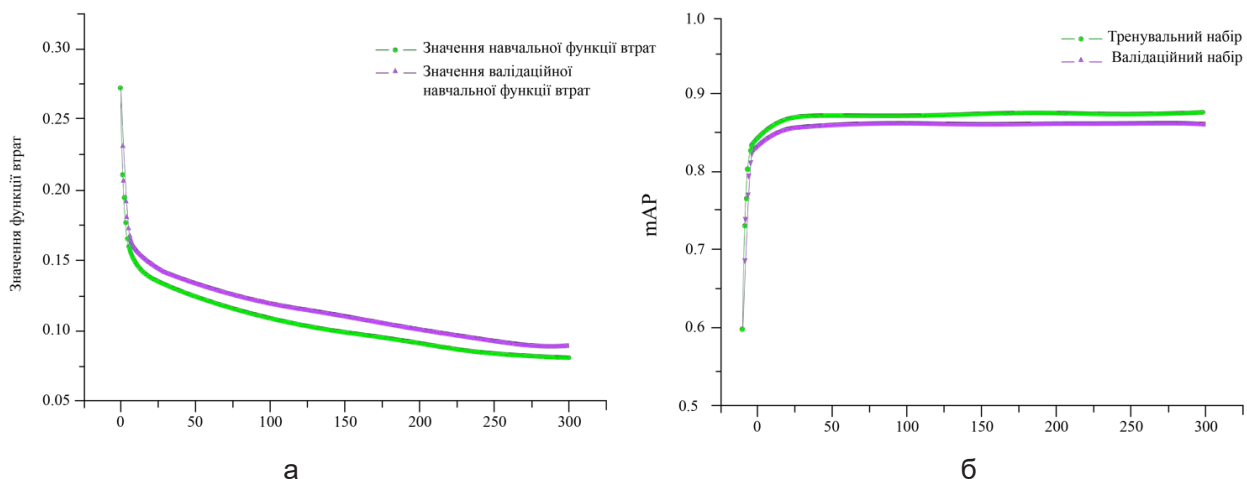


Рис. 2. Результати навчання мережі: (а) втрати під час навчання та перевірки; (б) середня точність класифікації навчальних наборів

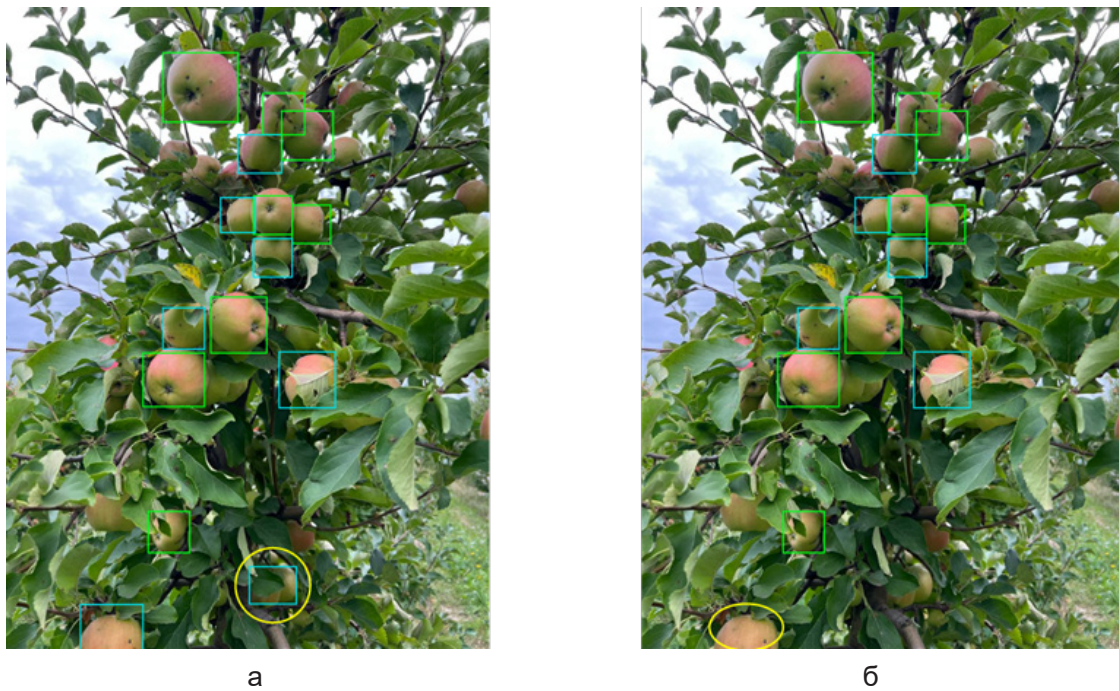


Рис. 3. Вплив порогу вірогідності на результат виявлення:  
(а) поріг занадто низький; (б) поріг занадто високий

Таблиця 3

**Результати виявлення яблук за нейромережевою моделлю YOLOv5-v1**

Дані	Кількість об'єктів	Precision (%)	Recall (%)	mAP (%)	F1 (%)
К-сть, яку можна захопити	1007	85,51	94,33	89,23	89,70
К-сть, яку не можна захопити	1329	82,56	89,32	84,87	85,81
Загалом	2336	83,83	91,48	86,75	87,49

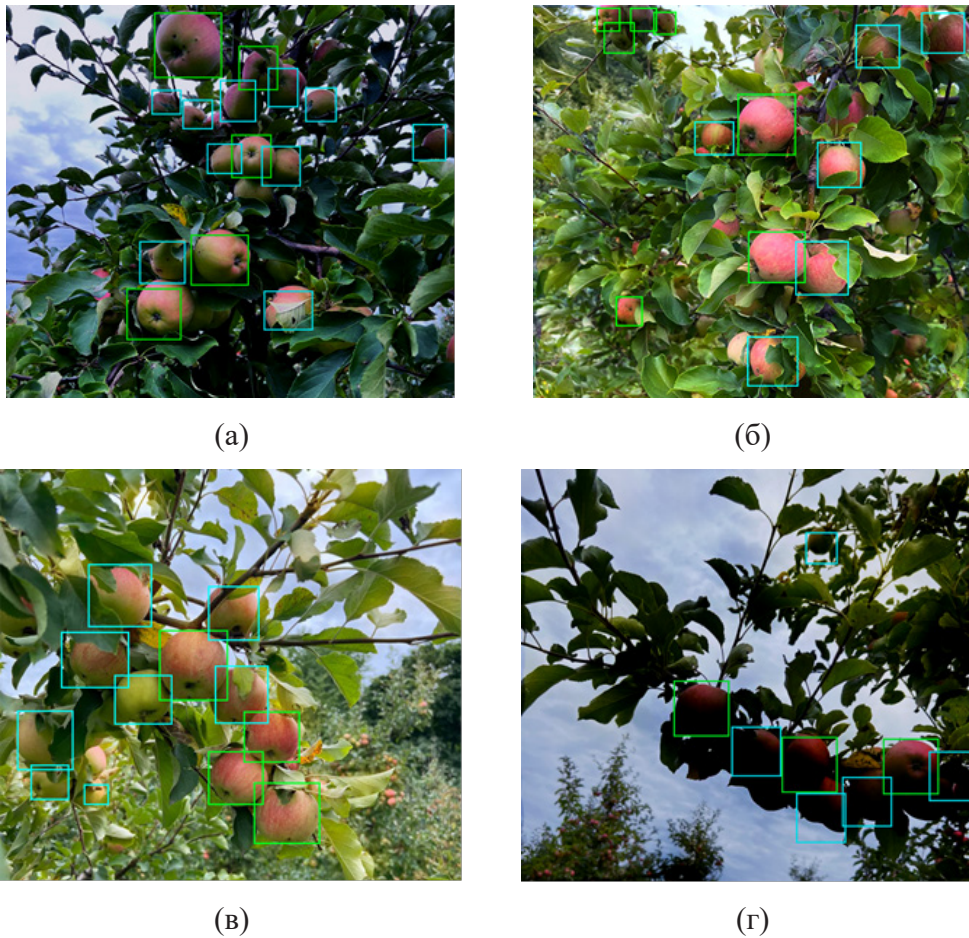
mAP і оцінка F1 запропонованої моделі становили 85,51%, 94,33%, 89,23% і 89,70%, відповідно; для незрозумілих плодів, результати розпізнавання становили 82,56%, 89,32%, 84,87% і 85,81% відповідно. Загальне значення влучності виявлення структурних об'єктів, повноти, mAP і F1 становили 83,83%, 91,48%, 86,75% і 87,49% відповідно.

Приклади результатів розпізнавання запропонованої моделі нейронної мережі фруктів, які можна захопити, і фруктів, які не можна захопити, за різних погодних умов і умов освітлення показані на рис. 4.

Обмежувальні рамки зеленого кольору використовувались, як мітки, які можна захопити, тоді як сині використовувались, як мітки, які неможливо захопити. Як видно на рис. 4, запропонована модель нейронної мережі виявляє структурні об'єкти на зображеннях знятих при рівномірному освітленні, похмурому освітленні, а також при сонячному освітленні. Крім цього, поточна модель виявляє структурні об'єкти в умовах переднього, заднього та бічного освітлення при сонячному освітленні.

Аналіз результатів. Для подальшого аналізу ефективності розпізнавання запропонованого підходу у роботі удосконалену мережу YOLOv5-v1 було порівняно з оригінальною мережею YOLOv5 (Behera et al., 2018), YOLOv3 (Huang et al., 2021), YOLOv4 (Chen et al., 2021) та EfficientDet-D0 (Wang et al., 2021) на 200 зображеннях тестового набору. В якості оціночних показників були прийняті значення та середня швидкість розпізнавання моделі. Результати виявлення, розмір і кількість параметрів кожної моделі мережі наведені в табл. 4.

Відповідно до таблиці 4, значення mAP для вдосконаленої моделі розпізнавання YOLOv5-v1, запропонованої в дослідженні, було найвищим, і воно було вищим, ніж у вихідної мережі YOLOv5, і вищим, ніж у мережах YOLOv3, YOLOv4 і EfficientDet-D0 відповідно, що вказує на те, що запропонований алгоритм є найкращим для задачі розпізнавання структурних об'єктів серед п'яти інших методів. Для швидкості розпізнавання моделі середня швидкість виявлення вдосконаленої моделі YOLOv5-v1 за зображенням (66,7 кадрів/с)



**Рис. 4. Результати розпізнавання фруктових плодів за допомогою вдосконаленої мережі YOLOv5-v1. (а) Хмарно (б) Бічне сонячне світло (в) Підсвічування сонцем (г) Сонячний фронт**

Таблиця 4

**Результати порівняння моделей виявлення фруктових плодів**

Мережі виявлення об'єктів	mAP (%)	Середня швидкість виявлення (s/pic)	Кількість параметрів	Розмір моделі (Мб)
YOLOv5	81,70	0,013	7,25×10 <sup>6</sup>	14,0
YOLOv3	71,80	0,053	6,15×10 <sup>7</sup>	235,0
YOLOv4	82,01	0,017	6,39×10 <sup>7</sup>	244,0
EfficientDet-D0	80,02	0,038	3,83×10 <sup>6</sup>	15,0
YOLOv5-v1	86,75	0,015	6,52×10 <sup>6</sup>	12,7

(кадр/с, кадрів на секунду), яка становила 2,53 і час мережі EfficientDet-D0, YOLOv4 і YOLOv3 відповідно, вказуючи, що запропонована модель нейронної мережі може задовольнити вимоги групи БПЛА для розпізнавання плодів у режимі реального часу.

У підсумку, побудована модель YOLOv5-v1 у рамках запропонованого підходу краще підходить для розгортання на вбудованих пристроях і може легко адаптуватися до змінних погодних умов у фруктовому саду.

**Висновки.** Розроблено новий підхід до виявлення структурної архітектури YOLOv5-v1. Під-

хід передбачає додавання рівнів до модуля BottleneckCSP-v4, який замінює оригінальний модуль BottleneckCSP. Покращений модуль SENet, також, був інтегрований у магістральну мережу для кращої ідентифікації середніх і великих плодів у різних умовах. Крім того, покращено початковий розмір блоку прив'язки вихідної мережі, щоб запобігти помилковій ідентифікації малих об'єктів на фоні зображення. Результати експериментів на тестовому наборі даних демонструють, що вдосконала модель мережі може ефективно розпізнавати фрукти, зняті камерами БПЛА. Модель досягла



92,13%, 84,59%, 87,94% і 89,02% показників запам'ятовування, точності, mAP і класифікації F1 відповідно, а середня швидкість виявлення становила 0,015 секунди на кадр у відеопослідовності.

В результаті, запропоновано новий підхід згідно глибокого навчання до виявлення, розпізнавання, обробки зображень та обчислення кількості об'єктів, зокрема на прикладі обробки

зображень яблук. Отримані рішення є частиною автоматизованої системи виявлення, розпізнавання, обробки зображень та обчислення кількості об'єктів, в складі якої використовуються БПЛА для отримання зображень.

Напрямом подальших досліджень буде розробка методу обробки зображень, які отримані з декількох камер різних БПЛА і містять спільні області зображень.

#### ЛІТЕРАТУРА:

1. Medvedeva Y., Kucher A., Lipsa J. et al. Human health risk assessment on the consumption of apples growing in urbanized areas: Case of Kharkiv, Ukraine. *International Journal of Environmental Research and Public Health*. Vol. 18, № 4. P. 1504. DOI:10.3390/ijerph18041504.
2. Suresh K. M., Mohan S. Selective fruit harvesting: Research, trends and developments towards fruit detection and localization – A review. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*. Vol. 237, № 6. P. 1405–1444. DOI:10.1177/09544062221128443.
3. Yu L., Xiong J., Fang X. et al. A litchi fruit recognition method in a natural environment using RGB-D images. *Biosystems Engineering*. Vol. 204. 2021. P. 50–63. DOI:10.1016/j.biosystemseng.2021.01.015.
4. Wan N. S. R., Muhammad A. H., Megat S. A. M. A. et al. Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alexandria Engineering Journal*. Vol. 61, № 2. P. 1265–1276. DOI:10.1016/j.aej.2021.06.053.
5. Fu L., Duan J., Zou X. et al. Banana detection based on color and texture features in the natural environment. *Computers and Electronics in Agriculture*. Vol. 167. 2019. P. 105057. DOI:10.1016/j.compag.2019.105057.
6. Radiuk P., Hrypynska N. A framework for exploring and modeling neural architecture search methods. *The 4th International Conference on Computational Linguistics and Intelligent Systems (COLINS-2020) : CEUR-Workshop Proceedings*. Vol. 2604. (Lviv, 23-24 April 2020). Lviv, 2020. P. 1060–1074. URL: <http://ceur-ws.org/Vol-2604/>
7. Li W., Feng X. S., Zha K. et al. Summary of target detection algorithms. *Journal of Physics: Conference Series*. Vol. 1757, № 1. P. 012003. DOI:10.1088/1742-6596/1757/1/012003.
8. Tang P., Wang X., Wang A. et al. Weakly supervised region proposal network and object detection. *Computer Vision – ECCV 2018 (Cham, 2018)*. Cham : Springer International Publishing, 2018. P. 370–386. DOI:10.1007/978-3-030-01252-6\_22.
9. Zhao G., Li G., Xu R. et al. Collaborative training between region proposal localization and classification for domain adaptive object detection. *Computer Vision – ECCV 2020 (Cham, 2020)*. Cham : Springer International Publishing, 2020. P. 86–102. DOI:10.1007/978-3-030-58523-5\_6.
10. Radiuk P., Pavlova O., Hrypynska N. An ensemble machine learning approach for Twitter sentiment analysis. *The 6th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2022)*. Volume I: Main Conference : *CEUR-Workshop Proceedings*. Vol. 3171. (Gliwice, Poland, 12-13 May 2022). Gliwice, 2022. P. 387–397. URL: <http://ceur-ws.org/Vol-3171/> C. 11.
11. Mai X., Zhang H., Jia X. et al. Faster R-CNN with classifier fusion for automatic detection of small fruits. *IEEE Transactions on Automation Science and Engineering*. Vol. 17, № 3. P. 1555–1569. DOI:10.1109/TASE.2020.2964289.
12. Wang S. Research towards YOLO-series algorithms: Comparison and analysis of object detection models for real-time UAV applications. *Journal of Physics: Conference Series*. Vol. 1948, № 1. P. 012021. DOI:10.1088/1742-6596/1948/1/012021.
13. Bresilla K., Perulli G. D., Boini A. et al. Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in Plant Science*. Vol. 10. 2019. P. 610. DOI:10.3389/fpls.2019.00611
14. Huang Z., Zhang P., Liu R. et al. Immature apple detection method based on improved YOLOv3. *ASP Transactions on Internet of Things*. Vol. 1, № 1. P. 9–13. DOI:10.52810/TIOT.2021.100028.
15. Chen W., Zhang J., Guo B. et al. An apple detection method based on Des-YOLO v4 algorithm for harvesting robots in complex environments. *Mathematical Problems in Engineering*. Vol. 2021. 2021. P. e7351470. DOI:10.1155/2021/7351470.
16. Behera S. K., Mishra N., Sethy P. K. et al. On-tree detection and counting of apple using color thresholding and CHT. *2018 International Conference on Communication and Signal Processing (ICCSP-2018)*. Vol. 2018, (2018). IEEE Inc., 2018. P. 0224–0228. DOI:10.1109/ICCSP.2018.8524363.

17. Колокольчикова І. В. Промислове садівництво Півдня України в рамках забезпечення продовольчої безпеки. *Science and Education a New Dimension*. VII(24), № 200. P. 7–10. DOI:10.31174/SEND-NT2019-200VII24-01.
18. Wang Y., Qin Y., Cui J. Occlusion robust wheat ear counting algorithm based on deep learning. *Frontiers in Plant Science*. 2021. Vol. 12. P. 645899. DOI:10.3389/fpls.2021.645899
19. Tan M., Le Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the 36th International Conference on Machine Learning, ICML-2019*. Long Beach, California, USA : PMLR. org, 2019. URL: P. 6105–6114. <http://proceedings.mlr.press/v97/tan19a.html>.
20. Vinci A., Brigante R., Traini C. et al. Geometrical characterization of hazelnut trees in an intensive orchard by an unmanned aerial vehicle (UAV) for precision agriculture applications. *Remote Sensing*. Vol. 15, № 2. P. 541. DOI:10.3390/rs15020541.
21. Zhou D., Hou Q., Chen Y. et al. Rethinking bottleneck structure for efficient mobile network design. *Computer Vision – ECCV 2020 (Cham, 2020)*. Cham : Springer International Publishing, 2020. Vol. 12348. P. 680–697. DOI:10.1007/978-3-030-58580-8\_40.
22. Kim H.-U., Bae T.-S. Deep learning-based GNSS network-based real-time kinematic improvement for autonomous ground vehicle navigation. *Journal of Sensors*. Vol. 2019. 2019. P. e3737265. DOI:10.1155/2019/3737265.
23. Dąbrowski P. S., Specht C., Specht M. et al. Integration of multi-source geospatial data from GNSS receivers, terrestrial laser scanners, and unmanned aerial vehicles. *Canadian Journal of Remote Sensing*. Vol. 47, № 4. P. 621–634. DOI:10.1080/07038992.2021.1922879.
24. Dempsey P. Reviews consumer technology: The teardown: Apple iPhone pro 13 smartphone. *Engineering & Technology*. Vol. 16, № 11. P. 68–69. DOI:10.1049/et.2021.1122.
25. Claid.ai: Generate, enhance and edit images at scale via API. 2023. URL: <https://claid.ai/> (дата звернення: 01.04.2023).
26. Mishra A. Machine learning in the AWS cloud: Add intelligence to applications with Amazon SageMaker and Amazon Rekognition. *Amazon SageMaker*. John Wiley & Sons, 2019. P. 353–385. URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119556749>

#### REFERENCES:

1. Medvedeva, Y., Kucher, A., Lipsa, J., & Heldak, M. (2021). Human health risk assessment on the consumption of apples growing in urbanized areas: Case of Kharkiv, Ukraine. *International Journal of Environmental Research and Public Health*, 18(4), Article 4. <https://doi.org/10.3390/ijerph18041504>
2. Suresh Kumar, M., & Mohan, S. (2023). Selective fruit harvesting: Research, trends and developments towards fruit detection and localization – A review. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 237(6), 1405–1444. <https://doi.org/10.1177/09544062221128443>
3. Yu, L., Xiong, J., Fang, X., Yang, Z., Chen, Y., Lin, X., & Chen, S. (2021). A litchi fruit recognition method in a natural environment using RGB-D images. *Biosystems Engineering*, 204, 50–63. <https://doi.org/10.1016/j.biosystemseng.2021.01.015>
4. Wan Nurazwin Syazwani, R., Muhammad Asraf, H., Megat Syahirul Amin, M. A., & Nur Dalila, K. A. (2022). Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alexandria Engineering Journal*, 61(2), 1265–1276. <https://doi.org/10.1016/j.aej.2021.06.053>
5. Fu, L., Duan, J., Zou, X., Lin, G., Song, S., Ji, B., & Yang, Z. (2019). Banana detection based on color and texture features in the natural environment. *Computers and Electronics in Agriculture*, 167, 105057. <https://doi.org/10.1016/j.compag.2019.105057>
6. Radiuk, P., & Hrypynska, N. (2020). A framework for exploring and modelling neural architecture search methods. In V. Lytvyn, V. Vysotska, T. Hamon, N. Grabar, N. Sharonova, O. Cherednichenko, & O. Kanishcheva (Eds.), *Proceedings of the 4th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2020)* (Vol. 2604, pp. 1060–1074). CEUR-WS. <http://ceur-ws.org/Vol-2604/>
7. Li, W., Feng, X. S., Zha, K., Li, S., & Zhu, H. S. (2021). Summary of target detection algorithms. *Journal of Physics: Conference Series*, 1757(1), 012003. <https://doi.org/10.1088/1742-6596/1757/1/012003>
8. Tang, P., Wang, X., Wang, A., Yan, Y., Liu, W., Huang, J., & Yuille, A. (2018). Weakly supervised region proposal network and object detection. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (pp. 370–386). Springer International Publishing. [https://doi.org/10.1007/978-3-030-01252-6\\_22](https://doi.org/10.1007/978-3-030-01252-6_22)
9. Zhao, G., Li, G., Xu, R., & Lin, L. (2020). Collaborative training between region proposal localization and classification for domain adaptive object detection. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.),

Computer Vision – ECCV 2020 (pp. 86–102). Springer International Publishing. [https://doi.org/10.1007/978-3-030-58523-5\\_6](https://doi.org/10.1007/978-3-030-58523-5_6)

10. Radiuk, P., Pavlova, O., & Hrypynska, N. (2022). An ensemble machine learning approach for Twitter sentiment analysis. *Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Systems (CoLInS-2022)*. Volume I: Main Conference, 3171, 11. <http://ceur-ws.org/Vol-3171/>

11. Mai, X., Zhang, H., Jia, X., & Meng, M. Q.-H. (2020). Faster R-CNN with classifier fusion for automatic detection of small fruits. *IEEE Transactions on Automation Science and Engineering*, 17(3), 1555–1569. <https://doi.org/10.1109/TASE.2020.2964289>

12. Wang, S. (2021). Research towards yolo-series algorithms: Comparison and analysis of object detection models for real-time UAV applications. *Journal of Physics: Conference Series*, 1948(1), 012021. <https://doi.org/10.1088/1742-6596/1948/1/012021>

13. Bresilla, K., Perulli, G. D., Boini, A., Morandi, B., Corelli Grappadelli, L., & Manfrini, L. (2019). Single-shot convolution neural networks for real-time fruit detection within the tree. *Frontiers in Plant Science*, 10. <https://www.frontiersin.org/articles/10.3389/fpls.2019.00611>

14. Huang, Z., Zhang, P., Liu, R., & Li, D. (2021). Immature apple detection method based on improved YOLOv3. *ASP Transactions on Internet of Things*, 1(1), Article 1. <https://doi.org/10.52810/TIOT.2021.100028>

15. Chen, W., Zhang, J., Guo, B., Wei, Q., & Zhu, Z. (2021). An apple detection method based on Des-YOLO v4 algorithm for harvesting robots in complex environments. *Mathematical Problems in Engineering*, 2021, e7351470. <https://doi.org/10.1155/2021/7351470>

16. Behera, S. K., Mishra, N., Sethy, P. K., & Rath, A. K. (2018). On-tree detection and counting of apple using color thresholding and CHT. *2018 International Conference on Communication and Signal Processing (ICCSP)*, 0224–0228. <https://doi.org/10.1109/ICCSP.2018.8524363>

17. Kolokolchikova, I. V. (2019). Promyslove sadivnytstvo Pivdnia Ukrainy v ramkakh zabezpechennia prodovolchoi bezpeky. *Science and Education a New Dimension*, VII(24)(200), 7–10. <https://doi.org/10.31174/SEND-NT2019-200VII24-01> [in Ukrainian]

18. Wang, Y., Qin, Y., & Cui, J. (2021). Occlusion robust wheat ear counting algorithm based on deep learning. *Frontiers in Plant Science*, 12, 645899. <https://doi.org/10.3389/fpls.2021.645899>

19. Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In K. Chaudhuri & R. Salakhutdinov (Eds.), *Proceedings of the 36th International Conference on Machine Learning, ICML-2019 (Vol. 97, pp. 6105–6114)*. PMLR.org. <http://proceedings.mlr.press/v97/tan19a.html>

20. Vinci, A., Brigante, R., Traini, C., & Farinelli, D. (2023). Geometrical characterization of hazelnut trees in an intensive orchard by an unmanned aerial vehicle (UAV) for precision agriculture applications. *Remote Sensing*, 15(2), Article 2. <https://doi.org/10.3390/rs15020541>

21. Zhou, D., Hou, Q., Chen, Y., Feng, J., & Yan, S. (2020). Rethinking bottleneck structure for efficient mobile network design. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), *Computer Vision – ECCV 2020 (pp. 680–697)*. Springer International Publishing. [https://doi.org/10.1007/978-3-030-58580-8\\_40](https://doi.org/10.1007/978-3-030-58580-8_40)

22. Kim, H.-U., & Bae, T.-S. (2019). Deep learning-based GNSS network-based real-time kinematic improvement for autonomous ground vehicle navigation. *Journal of Sensors*, 2019, e3737265. <https://doi.org/10.1155/2019/3737265>

23. Dąbrowski, P. S., Specht, C., Specht, M., Burdziakowski, P., Makar, A., & Lewicka, O. (2021). Integration of multi-source geospatial data from GNSS receivers, terrestrial laser scanners, and unmanned aerial vehicles. *Canadian Journal of Remote Sensing*, 47(4), 621–634. <https://doi.org/10.1080/07038992.2021.1922879>

24. Dempsey, P. (2021). Reviews consumer technology: The teardown: Apple iPhone pro 13 smartphone. *Engineering & Technology*, 16(11), 68–69. <https://doi.org/10.1049/et.2021.1122>

25. Claid.ai: Generate, enhance and edit images at scale via API. (2023). Retrieved March 18, 2023, from <https://claid.ai/>

26. Mishra, A. (2019). Machine learning in the AWS cloud: Add intelligence to applications with Amazon SageMaker and Amazon Rekognition. In *Amazon SageMaker (pp. 353–385)*. John Wiley & Sons. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119556749>