

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Методи напівкерованого адаптивного виявлення об'єктів
для підвищення ефективності сільського господарства

(тема)

Виконав:

Студент II курсу, групи СПМ-20-1
Кравченко Н.В.
(прізвище, ініціали)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва спеціальності)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Системне програмування
(повна назва освітньої програми)

Керівник: доц. Ляшенко О.С.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ЕОМ

(підпис)

Коваленко А.А.

(прізвище, ініціали)

2022 р.

Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання

Кафедра електронних обчислювальних машин

Рівень вищої освіти другий (магістерський)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва)

Тип програми освітньо-наукова
(освітньо-професійна або освітньо-наукова)

Освітня програма Системне програмування
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Кравченку Нікиті Вікторовічу
(прізвище, ім'я, по батькові)

1. Тема роботи Методи напівкерованого адаптивного виявлення об'єктів
для підвищення ефективності сільського господарства

затверджена наказом по університету від “ 25 ” березня 2022 р. № 33 Стз

2. Термін подання студентом роботи до екзаменаційної комісії 10 травня 2022 р.

3. Вхідні дані до роботи еко робот

Машине навчання

Комп'ютерний зір

Система розпізнавання

4. Перелік питань, що потрібно опрацювати у роботі _____

Постановка задачі

Огляд інструментів штучного інтелекту

Опис технічних вимог

Розробка методології

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) 14

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Отримання завдання	27.03	вик
2	Аналіз літератури	28.03-4.04	вик
3	Аналіз алгоритмів штучного інтелекту	4.04-16.04	вик
4	Реалізація проекту	17.04-24.04	вик
5	Підготовка пояснювальної записки	25.04-30.04	вик
6	Рецензування	1.05-3.05	вик
7	Подача в ЕК	10.05	вик

Дата видачі завдання 27 березня 2022 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

доц. Ляшенко О.С..
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 68 с., 15 рис., 1 дод., XX джерел.

КОМП'ЮТЕРНА МЕРЕЖА, ІНТЕРНЕТ, МАРШРУТИЗАТОР, ПРОТОКОЛ, СЕРВЕР, ШЛЮЗ, FIREWALL, WI-FI, WLAN.

Метою кваліфікаційної роботи є адаптувати модель виявлення об'єктів, що працює на автономному роботі з бур'янами, до нових умов. Найефективнішим рішенням проблеми є ручна анотація даних, але це рішення не є ефективним у часі. Таким чином, усі вимоги проекту спрямовані на мінімізацію ручних зусиль при одночасному підвищенні точності виявлення.

У ході виконання кваліфікаційної роботи запропоновано, реалізовано та оцінено систему напівконтрольованого адаптивного навчання (SAL), яка складається з двох компонентів, а саме інструменту анотації та виявлення об'єктів. За допомогою експериментальної оцінки показано, що система відповідає всім вимогам.

ABSTRACT

Master's thesis: 68 pages, 15 figures, , 1 appendices, 31 sources.

FIREWALL, GATE, INTERNET, PROTOCOL, ROUTER, SERVER, WI-FI, WIRELESS NETWORK, WLAN.

The major goal of the qualification work is to adapt the object detection model, which works on autonomous work with weeds, to new conditions. Manual data annotation is the most efficient solution to the problem, but this solution is not time-efficient. Thus, all project requirements are aimed at minimizing manual effort while increasing detection accuracy.

In the course of the qualification work, a semi-supervised adaptive learning (SAL) system was proposed, implemented and evaluated, which consists of two components, namely, an annotation tool and object detection. Experimental evaluation shows that the system meets all requirements.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	8
ВСТУП	9
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ	12
1.1 Огляд проекту.....	12
1.1.1 Схема проблеми	12
1.1.2 Клієнт	12
1.1.3 Мета проекту	12
1.1.4 Можливий внесок.....	13
1.2 Опис проблеми	14
1.2.1 Доступні дані	15
1.2.2 Постановка проблеми	15
1.3 Проблема адаптації середовища.....	16
1.4 Вимоги.....	17
2 ШТУЧНИЙ ІНТЕЛЕКТ ТА МАШИНЕ НАВЧАННЯ	18
2.1 Машинне навчання	18
2.1.1 Огляд машинного навчання	18
2.1.2 Оцінювання.....	19
2.2 Штучні нейронні мережі	23
2.3 Згорткові нейронні мережі	26
2.3.1 Згортка.....	26
2.3.2 Згортка в нейронних мережах.....	27
2.3.3 Згортковий шар	27
2.3.4 Структура CNN	28
2.3.5 Архітектури CNN.....	29
2.4 Виявлення об'єктів	30
3 ПРАКТИЧНІ ВИМОГИ К ДОСЛІДЖЕННЮ	32

3.1	Виявлення бур'янів і культур на основі навчання	32
3.2	Швидша адаптація	34
3.2.1	Машинне навчання	34
3.2.2	Точне землеробство	35
4	МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ	39
4.1	Напівконтрольоване адаптивне навчання	39
4.2	Апаратне та програмне забезпечення	42
4.3	Інструмент анотації.....	43
4.3.1	Наївний висновок.....	44
4.3.2	Воксельний висновок.....	45
4.3.3	Висновок на основі сегментації.....	52
	ВИСНОВКИ.....	54
	ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	57
	ДОДАТОК А Графічний матеріал кваліфікаційної роботи.....	61

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ
І ТЕРМІНІВ

ШІ – штучний інтелект

ШНМ – штучні нейронні мережі

АС – автономна система

ПЛК – програмований логічний контролер

МН – машинне навчання

ВСТУП

Оскільки у 2050 році чисельність населення світу досягне 9,73 мільярда осіб, Продовольча та сільськогосподарська організація ООН підкреслює, що виробництво сільськогосподарської продукції (харчування, кормів, біопалива) має збільшитися майже на 50% у 2012-2050 роках, щоб задовольнити попит. Цю мету може бути важко здійснити через ряд причин, таких як зміна клімату, дефіцит навколишнього середовища, недостатні інвестиції в сільське господарство та прогалини в технологіях [9]. Крім того, швидко зростаючий середній клас потребуватиме різноманітної високоякісної дієти, оскільки зростає обізнаність про харчування. Як наслідок, прогнозується, що ці тенденції продовжаться, а чисельність сільськогосподарської робочої сили продовжує скорочуватися [10].

Обговорені вище проблеми, збільшення розміру сільськогосподарських угідь [10] у поєднанні з сучасними передовими технологіями призвели до нової хвилі технологічних досягнень у сільському господарстві, яка називається точним землеробством. Європейський досвід показує, що фермери вважають точне землеробство найвпливовішим трендом, який впливатиме на практику та структури землеробства до 2030 року [10]. Відповідно до Addicott, точне землеробство можна розглядати як «підрозділ нещодавно виниклих «точних практик», які включали точну війну, точну астрофізику, точне проектування, точне озеленення, точність, медицину, точне розведення, точне зрошення або точне обрізання тощо. мало [11].

Точне землеробство Простими словами, точне землеробство можна розглядати як метод «застосування правильної обробки в потрібному місці в потрібний час» [12]. Однак на точне землеробство можна дивитися з різних сторін. Як приклад, Саймон Блекмор порівняв різні системи точного землеробства та дійшов висновку, що «Прецизійне землеробство – це процес управління, а не технологія» [13]. Опис Блекмора применшує акцент на

технологічному аспекті точного землеробства. Тому інші дослідники вирішили акцентувати увагу на технологічних аспектах і визначили це як «комбінацію різних інструментів (таких як вимірювачі врожайності, супутникові зображення та спеціалізоване програмне забезпечення) і методів (таких як картографування врожайності, картографування ґрунту та точне буріння), загальною метою якого є контроль коливань врожайності по полю [14].

Точне землеробство пропонує такі ключові переваги, як підвищення врожайності, зниження витрат і прийняття рішень. Його застосування може охоплювати широкий спектр сільського господарства, починаючи від посіву та обробки ґрунту закінчуючи збором врожаю [10].

Основними рушійними силами точного землеробства є автономні машини та системи, призначені для виконання всіх цих завдань. Дрони, Інтернет речей разом із автономними машинами для прополки бур'янів є перспективними автономними технологіями, які зараз розробляються [10].

Розвиток робототехніки, комп'ютерного зору, сенсорної технології та потужності обробки зробили роботів сильними кандидатами на полегшення фізичної праці людини. Отже, автономні роботи являють собою перспективну технологію для формування майбутніх методів ведення сільського господарства.

Автономні роботи для прополки бур'янів. Однією з найбільш виснажливих ручних робіт у сільському господарстві є прополка. Бур'яни – це рослини, які безладно ростуть на полях, де вони небажані. Таким чином, бур'яни слід видаляти якомога швидше, оскільки вони конкурують із культурою за ресурси. Це ускладнює виживання культури на ранніх стадіях. Крім того, дослідження показали сильну кореляцію між конкуренцією бур'янів і втратою врожаю [15], тому, якщо не обробляти належним чином, бур'яни також мають економічні наслідки.

Автономна система видалення бур'янів – це зазвичай робот, який виявляє бур'яни за допомогою алгоритмів виявлення на основі зору, а потім

усуває їх за допомогою вибраних технологій боротьби з бур'янами, які можуть бути будь-якими: від вибіркового обприскування або механічного видалення до лазерів, послідовного спалювання або розпилення абразивів [16]. Таким чином, автономні роботи-прополвачі є багатообіцяючою технологією, яка може допомогти вирішувати сучасні виклики стабільно. Це пояснюється тим, що вони забезпечують краще управління ресурсами за допомогою вибіркової прополки, також відомої як точна прополка, яка знімає залежність від агрохімікатів [17]. Створення робота для прополки передбачає широкий спектр технологій, таких як автономна навігація, планова координація та бачення для виявлення посівів/бур'янів. Ця робота в першу чергу буде зосереджена на технологіях бачення, які застосовують підходи, засновані на глибокому навчанні, для виявлення бур'янів і посівів.

Бачення на основі навчання для точної прополки. Сучасні системи бачення для точного землеробства базуються на аналізі зображень [18, 19, 20, 21] або навчанні систем [15, 22, 23, 24]. Підходи бачення на основі навчання підживлюються швидким розвитком області машинного навчання, що вимагає складних моделей для виявлення та сегментації [4, 2, 2]. Як правило, для моделей потрібні анотовані дані, які анотуються вручну, щоб навчитися виявляти рослини. Точність моделей виявлення значною мірою залежить від кількості позначених даних, які використовуються для навчання. Таким чином, висока точність виявлення об'єктів вимагає значної кількості даних. Також важливо, щоб отримані дані були високої якості та були репрезентативними для рослин на різних стадіях росту та в різних середовищах.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Огляд проекту

У цьому проекті розглядається проблема адаптації детекторів об'єктів для автономних роботів-прополювачів до нових умов.

1.1.1 Схема проблеми

Встановлено, що відсутність надійних систем виявлення посівів і бур'янів є основною перешкодою для комерційної адаптації роботизованих систем боротьби з бур'янами [25]. Це пояснюється тим, що точність виявлення знижується, коли система виявлення прогнозує в новому середовищі, відмінному від того, на якому вона навчалася [17].

Щоб вирішити цю проблему, потрібно зібрати та позначити більше даних, що дорого та забирає багато часу. Це призводить до повільнішого та дорожчого розгортання робота для прополки на різних полях.

1.1.2 Клієнт

Екобот знаходиться в авангарді розробки автономного польового робота для механічного контролю бур'янів. Вони вже створили одного з найрозумніших у світі автономних механічних роботів для видалення бур'янів, використовуючи систему зору на основі машинного навчання, див. рисунок 1.1.

1.1.3 Мета проекту

Щоб допомогти Екобот більш легко розгортати свого робота для

прополки на різних полях і середовищах, система зору робота повинна бути краще адаптована до нових середовищ, зводячи до мінімуму потребу в втручанні людини, наприклад анотації даних вручну.

1.1.4 Можливий внесок

Ця робота представляє напівконтрольований метод для швидкої та надійної адаптації до середовища. Основний внесок цієї роботи можна підсумувати таким чином:

1. Навчання та впровадження найсучасніших методів виявлення об'єктів для досягнення точності виявлення в реальному часі понад 94%.
2. Пропозиція нової системи анотації для автоматичного маркування даних.
3. Пропозиція, впровадження та оцінка напівконтрольованої системи для надійної адаптації детекторів на основі глибокого навчання в сільськогосподарських полях.



Рисунок 1.1 – Автономний робот для прополки.

1.2 Опис проблеми

Модель виявлення на основі глибокого навчання вивчає анотовані зображення, навчаючись на них. Це означає, що модель співвідносить свої вивчені шаблони з деталями, розпізнаними на зображеннях. Коли детектор об'єктів працює в новому середовищі, його точність виявлення знижується, оскільки він не може розпізнавати нові функції, наявні в новому середовищі [17]. Щоб подолати цю проблему, систему виявлення об'єктів потрібно перенавчити на нове середовище, щоб відновити точність виявлення. Важливо підтримувати високу точність, оскільки нижча точність може обмежити можливості автономного робота для прополки, зробивши його менш надійним і зручним у використанні. Адаптація моделі до нового середовища вимагає збору даних із нового середовища та позначення даних, щоб потім їх можна було використовувати для навчання моделі виявлення. Отже, ця процедура стає фінансовим і тимчасовим тягарем під час масштабування системи виявлення в кількох місцях, рисунок 1.2.

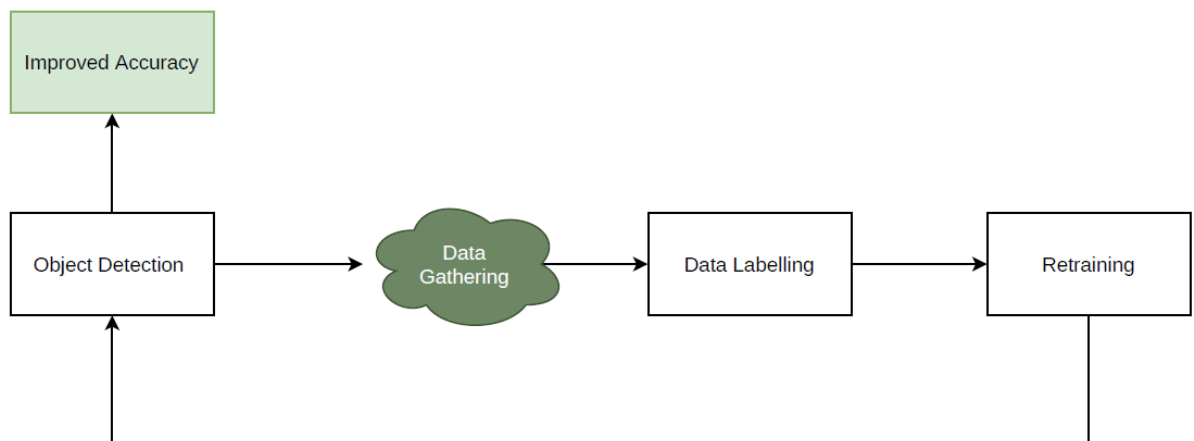


Рисунок 1.2 – Масштабування системи виявлення для нових середовищ

Масштабування системи виявлення для нових середовищ створює постійну проблему. Кожного разу, коли систему виявлення встановлюють у

новому середовищі, її потрібно адаптувати за кроками на схемі, щоб уникнути ризику низької точності виявлення. Зелена хмара символізує нове середовище.

1.2.1 Доступні дані

Система виявлення, розширена в цій роботі, розроблена спеціально для робота-прополювача Ekobot, який має кілька датчиків, таких як камера RGBD, LIDAR, одометрія коліс і RTK-GNSS. Крім того, Ekobot надає анотований набір даних для навчання детекторів об'єктів. Доступний набір даних складається з трьох класів і надається у двох форматах: Pascal VOC і CVAT. Усі зображення в наборі даних витягнуті з відеоматеріалів, що означає, що деякі зображення можуть бути розмитими, що призводить до втрати деталей.

Дані анотовані за допомогою обмежувальних рамок, але невелика частина даних анотована за допомогою багатокутників, які можна перетворити на маски сегментації. Це означає, що доступні дані можна використовувати не тільки для виявлення об'єктів, але й для семантичної сегментації. Три класи позначають корінь цибулі, корінь бур'яну та бур'ян з анотаціями в рамці. Класи однакові для анотацій семантичної сегментації, але відрізняються лише тим, що анотується вся рослина цибулі, а не лише корінь, дивіться рисунок 1.3 для зразків того, як виглядає анотація. Загалом є 3956 зображень із роздільною здатністю 1280x640, де 3656 зображень анотовано за допомогою обмежувальних рамок, а 300 – за допомогою полігонів.

1.2.2 Постановка проблеми

Враховуючи наявні джерела даних, проблема полягає в тому, щоб знайти ефективний, швидкий і надійний метод прискорення адаптації моделі виявлення в реальному часі робота-прополки до нових умов.



Рисунок 1.3 – Зразки доступних даних: (а) зображення з анотаціями обмежувальної рамки (b) зображення з семантичною сегментацією.

Помаранчевий колір позначає цибулю, а зелений – бур'ян.

1.3 Проблема адаптації середовища

Проблема адаптації середовища цього проекту обмежена налаштуванням системи до середовищ, які схожі одне на одне. Таким чином, не очікується, що нове середовище включатиме різкі зміни, такі як різні види рослин або зовсім інше освітлення (день проти ночі). Натомість очікується, що зміни погіршать модель виявлення, але все одно збережуть її функціональність. Прикладами таких змін є варіації полів і росту рослин.

Під час навчання власних детекторів об'єктів мета полягає в тому, щоб отримати оптимальний висновок у формі обмежувальних рамок, а не семантичних масок. Крім того, ця робота не спрямована на модифікацію або створення спеціальних згорткових нейронних мереж або детекторів об'єктів, а лише на використання існуючих методів. Оскільки робот-пропелювач не використовує корінь класу бур'янів, його було вилучено з набору даних граничної рамки, тобто лише корінь цибулі та бур'ян є проблемними. Інші роботи з посівами/бур'янами використовували різні налаштування для найкращої взаємодії з навколишнім середовищем під час захоплення даних, наприклад використання камери RGB+NIR [26], а також завіси зі штучним

джерелом світла для керування освітленням [23]. Тим не менш, цей проект спрямований лише на найкраще використання переваг поточного налаштування робота-прополювача, а не на його модифікацію.

1.4 Вимоги

Ця робота має на меті адаптувати модель виявлення об'єктів, що працює на автономному роботі з бур'янами, до нових умов. Найефективнішим рішенням проблеми є ручна анотація даних, але це рішення не є ефективним у часі. Таким чином, усі вимоги проекту спрямовані на мінімізацію ручних зусиль при одночасному підвищенні точності виявлення, і їх можна описати так:

1. Ефективна адаптація: рішення вважається ефективним, якщо воно покращує точність виявлення в новому середовищі. Поліпшення відносно зниження точності.

2. Швидка адаптація: рішення вважається швидким, якщо воно краще адаптує модель за короткий період часу порівняно з часом анотації.

3. Надійність: наявність надійного рішення означає, що рішення повинно бути здатним адаптуватися до різних сценаріїв з мінімальними змінами. Це призводить до того, що на адаптацію рішення до нових середовищ витрачається менше часу.

2 ШТУЧНИЙ ІНТЕЛЕКТ ТА МАШИНЕ НАВЧАННЯ

Термін ШІ (штучний інтелект) зараз широко використовується. Його фактичні наслідки більше пов'язані з машинним навчанням, яке є підмножиною ШІ. Метою цієї роботи є адаптація моделі виявлення культур/бур'янів на основі машинного навчання до нових середовищ. Таким чином, машинне навчання є важливим аспектом цього проекту.

Цей розділ починається зі знайомства з загальними основами машинного навчання, необхідними для розуміння моделі навчання та природи цієї роботи. Після обговорення машинного навчання вводяться штучні нейронні мережі (ШНМ), будівельні блоки глибокого навчання. Розуміння ШНМ створює основу, яка допомагає зрозуміти згорточні нейронні мережі (CNN), оскільки згорточні нейронні мережі є похідними від ШНМ. CNN зазвичай використовуються в обробці зображень на основі глибокого навчання і тому є фундаментальним компонентом системи виявлення.

2.1 Машинне навчання

Машинне навчання можна розглядати як програму, яка покращує свою продуктивність при виконанні заданого завдання через досвід [27]. Властивість вдосконалюватися за допомогою досвіду або «навчання» дозволила машинному навчанню вирішувати різноманітні завдання, такі як класифікація, регресія або машинний переклад різних форм даних, наприклад, текст, аудіо та зображення [28].

2.1.1 Огляд машинного навчання

Оскільки машинне навчання можна використовувати для різних завдань, галузь розділена на підрозділи, які спеціалізуються на вирішенні конкретної

проблеми. Серед підрозділів, що мають відношення до цієї роботи, є:

- контрольоване та неконтрольоване машинне навчання: розрізнити контрольоване та неконтрольоване навчання тривіально, оскільки вони в основному відрізняються доступним набором вхідних даних. Контрольоване машинне навчання очікує доступної основної правдивої інформації або міток, які відповідають даним навчання. Основна істина використовується як еталон для оцінки ефективності моделі. Таким чином, контрольована система намагається знайти зв'язок між вхідним набором даних і одним або декількома відомими класами шляхом мінімізації різниці (помилки) між прогнозами та мітками. З іншого боку, система машинного навчання без нагляду не має доступних міток, тому вона займається вивченням корисних властивостей структури набору даних [28];

- напівконтрольоване машинне навчання: напівконтрольоване навчання можна розглядати як гібрид двох попередніх методів навчання, де позначені та немарковані дані доступні для прогнозування [28].

- трансферне навчання: трансфертне навчання покращує вивчення нового завдання шляхом передачі знань із попередньо підготовленого вихідного завдання. Вихідне завдання має бути навчено на подібному, але бажано більш загальному наборі даних. Коли вихідне завдання відповідає цим вимогам, навчання стає швидшим і точнішим [29].

- активне навчання: ключова ідея активного навчання полягає в тому, щоб дозволити алгоритму машинного навчання вибирати дані, з яких він навчається. В результаті модель може досягти більшої точності з меншою кількістю мічених екземплярів навчання [30].

2.1.2 Оцінювання

Основна мета машинного навчання — навчитися правильно узагальнювати невидимі дані. Оскільки цей проект тренує низку моделей машинного навчання з використанням методів навчання під наглядом,

важливо розпізнавати фундаментальні концепції під час оцінювання моделей.

Перевірка. Щоб перевірити результати моделі під час навчання, модель не навчається на всьому наборі даних, а на його частині. Зазвичай набір даних розбивають на три набори: навчальний набір, який використовується для навчання моделі, набір перевірки для налаштування гіперпараметрів під час навчання та тестовий набір для оцінки ефективності моделі. Існують інші методи валідації, такі як валідація затримки, n -кратна перехресна валідація та перевірка за відсутності одного. Однак ця робота розділяє набір даних на три набори як перший згаданий метод.

Крива навчання. У процесі навчання модель прагне мінімізувати загальну помилку, щоб навчитися узагальнювати конкретне завдання. Надмірно складна модель не може узагальнити, що призводить до переобладнання даних. Переобладнання є ознакою того, що розмір набору даних, який використовується для навчання моделі, є недостатнім і що більше даних підвищить продуктивність. На відміну від надмірної комплектації, дуже проста модель призводить до недостатньої комплектації. Як вимірювання того, чи є у моделі надмірна чи недостатня підготовка, можна побудувати графік похибки навчання та перевірки щодо кількості епох, див. рисунок 2.1, де епоха — це ітерація, коли модель навчається на всіх даних у наборі навчальних даних. З графіка видно, що модель починає втрачати своє узагальнення, коли помилка перевірки починає зростати. Рання зупинка може бути використана, щоб уникнути переобладнання шляхом припинення навчання, якщо помилка підтвердження зростає. На практиці зазвичай зберігаються контрольні точки, які описують модель, коли вона навчається. Це дозволяє продовжити тренування за бажанням, щоб гарантувати досягнення найкращої ваги.

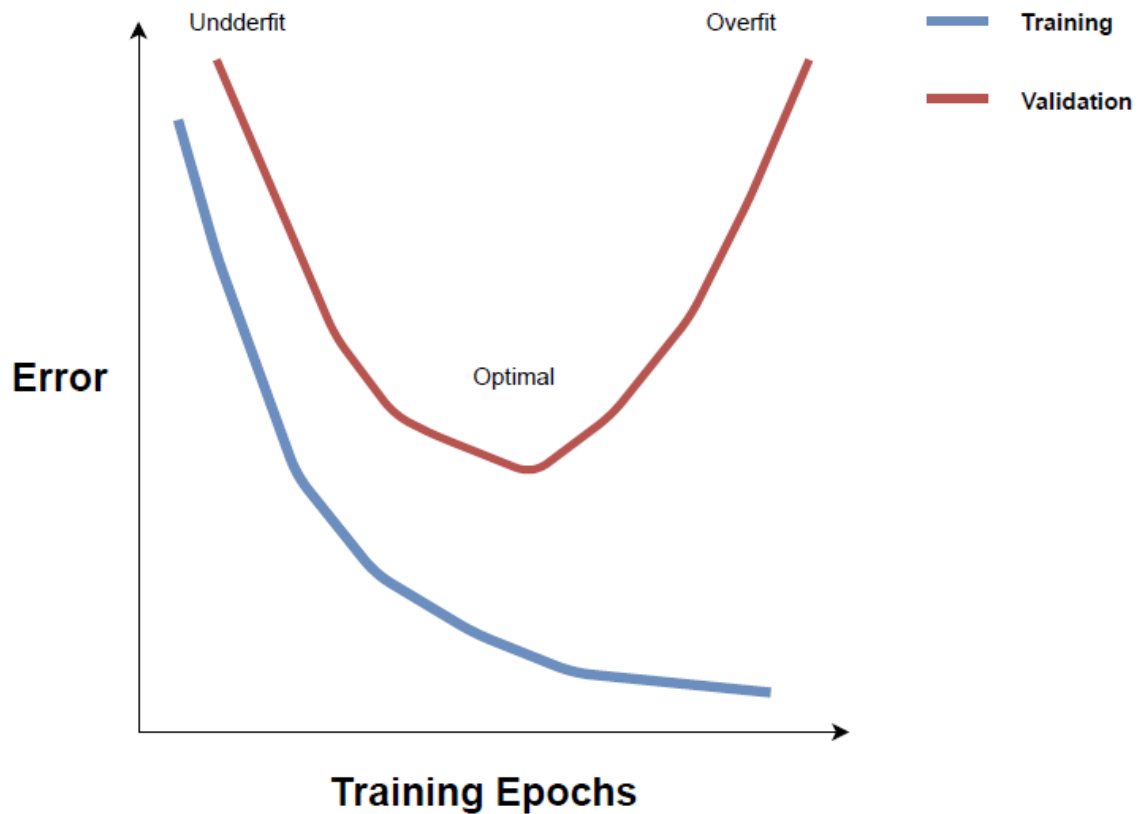


Рисунок 2.1 – Спрощений графік кривої навчання.

Коли похибка навчання та перевірки висока, модель не підходить. Модель є переобладнаною, коли помилка перевірки висока, а помилка навчання низька. Модель має найкращу ефективність, коли помилка підтвердження є якомога меншою.

Виявлення аномалії. Для виявлення спостережень, які відхиляються від набору даних, зазвичай розглядають чотири ключові поняття для інтерпретації результатів прогнозу, тобто істинні позитивні, істинні негативні, хибні позитивні та хибні негативні результати. Істинні позитиви та мінуси (tp & tn) представляють позитивні відповідні негативні прогнози, які відповідають міткам. Помилкові позитивні та негативні результати – це прогнози, які не відповідають відповідній мітці. Основне використання цих показників полягає в обчисленні точності системи:

$$accuracy = \frac{t_p + t_n}{t_p + f_p + f_n + t_n}$$

Точність і відкликання. Більш виразним показником є використання точності та запам'ятовування. Ці два показники можна використовувати, щоб оцінити, чи навчилася модель узагальнювати завдання чи просто запам'ятала навчальний набір. Таким чином, точність – це частка правильно визначених класів (істинних позитивних результатів) від усіх знайдених виявлень, тобто сума істинних і хибних позитивних результатів. Відкликання – це частка правильно визначених класів серед усіх релевантних елементів, тобто сума справжніх позитивних і помилкових негативних результатів:

$$precision = \frac{t_p}{t_p + f_p},$$

$$recall = \frac{t_p}{t_p + f_n}$$

Точність і запам'ятовування часто відображаються разом на кривій точності і запам'ятовування, і їх можна об'єднати, щоб отримати більш складні показники, такі як F-оцінка.

Оцінка точності системи навчання під наглядом залежить від програми. Оскільки в цій роботі використовується машинне навчання головним чином для виявлення об'єктів, важливо зрозуміти, як можна оцінити точність. Виявлення об'єктів відрізняється від класифікації зображень тим, що воно одночасно знаходить і класифікує цікаві об'єкти в наборі даних зображення. Таким чином, оцінка ефективності детектора об'єктів також складається з оцінки того, наскільки добре локалізовано об'єкт. Це можна виконати шляхом обчислення перетину через об'єднання (IoU) виявлення щодо його мітки. IoU отримується шляхом ділення площі перетину передбачення з його міткою на об'єднання передбачення та мітки.

$$IoU = \frac{A(\text{detection}) \cap A(\text{label})}{A(\text{detection}) \cup A(\text{label})}$$

IoU супроводжує часто використовувану метрику для оцінки виявлення об'єктів під назвою Average Precision, вперше представлену у VOC2007 [31]. AP є mAP являє собою площу під інтерпольованою кривою точності відкликання [31]. Під час оцінки детекторів AP розглядається, лише якщо він перевищує поріг IoU. Зазвичай для детекторів об'єктів оцінюють IoU, еквівалентний 0,5, 0,75 і 0,5:0,95, де останні представляють середнє значення обчислень AP із пороговими значеннями IoU від 0,5 до 0,95 із кроком 0,05. Зазвичай AP оцінюється для кожного класу, а потім усереднюється для отримання середньої середньої точності (mAP).

2.2 Штучні нейронні мережі

Перша форма штучних нейронних мереж (ШНМ), яку зазвичай називають нейронними мережами (НМ), була вперше представлена Уорреном МакКалохом і Волтером Піттсом у 1943 році [32]. Назва «нейронні мережі» походить від нейронних мереж, які можна знайти в біологічному мозку. ШНМ — це набір щільно пов'язаних між собою вузлів, які називаються штучними нейронами. Кожен штучний нейрон приймає вхідний вектор x , щоб створити вихідний сигнал y , який можна надіслати до кількох нейронів. Кожне введене значення множиться на вагу, W , яка визначає, наскільки вхід має впливати на вихід нейрона. Сума цієї функції, додана до члена зсуву b , вводиться у функцію активації φ (наприклад, сигмоїда, випрямлені лінійні одиниці). Для n входів цю операцію можна підсумувати таким чином:

$$\varphi \left(\sum_{i=1}^n W_i x_i + b \right) = \varphi(Wx + b).$$

Рівень нейронної мережі складається з кількох вузлів з однаковими вхідними даними. Таким чином, для шару l вектор вагових коефіцієнтів у рівнянні, розширюється до матриці $W^{(l)}$, яка містить вагові коефіцієнти всіх вузлів шару. Подібним чином зміщення також поширюється на вектор $b^{(l)}$. Це означає, що $W_{ij}^{(l)}$ позначає вагу між вузлом j у шарі l та вузлом i у шарі $l + 1$. Подібним чином $b_i^{(l)}$ є зміщенням для вузла i у шарі l . Це призводить до наступного множення матриці:

$$\varphi(W^{(n)}x + b^{(n)}).$$

Глибокі нейронні мережі. Використання кількох рівнів нейронної мережі, які називаються прихованими шарами, створює глибоку нейронну мережу. Отже, глибока нейронна мережа складається з трьох типів шарів: вхідного та вихідного, а також прихованого шару, рисунок 2.2. Метою цієї структури є апроксимація функції f^* , яка відображає вхідні дані у вихідні дані, тобто $y = f^*(x)$.

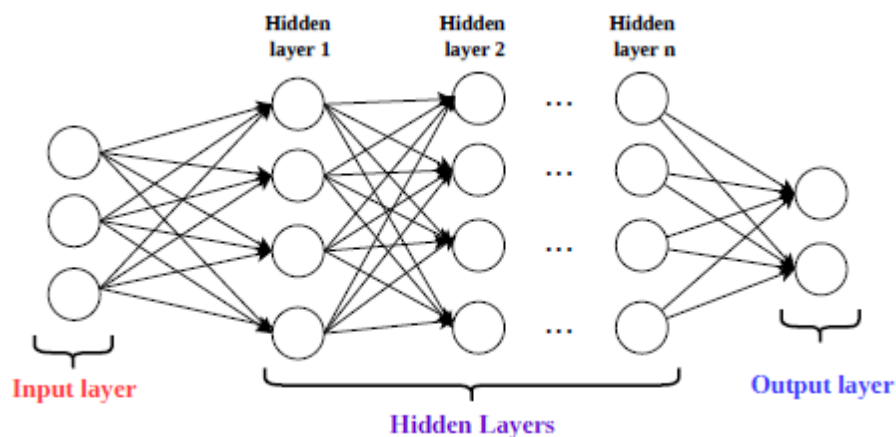


Рисунок 2.2 – Структура глибокої нейронної мережі

Поширення прямого зв'язку. Потік даних через вхідні рівні через приховані рівні до досягнення вихідного рівня називається прямим поширенням. Іншими словами, пряме розповсюдження є сукупністю всіх n рівнів, як показано нижче:

$$f^*(x) = (f^{(n)} \circ f^{(n-1)} \circ \dots \circ f^{(1)})(x)$$

Що робить цю архітектуру настільки потужною, так це її здатність використовувати універсальну теорему про наближення, тобто нейронна мережа може обчислити будь-яку довільну функцію [33]. За словами Гудфеллоу, «мережі прямого зв'язку з одним рівнем достатньо для представлення будь-якої функції, але рівень може бути нездійснено великим і не в змозі правильно навчитися та узагальнити» [28].

Зворотне розповсюдження. Коли інформація проходить через ШНМ, вона створює вихідні дані, які можна використовувати для перевірки функцією втрат або витрат. Функції втрат, такі як перехресна ентропія та квадратична середня помилка, направляють мережу на апроксимацію функції відображення. Отже, мінімізація функції втрат веде до кращої оцінки. Через нелінійність нейронної мережі функції вартості стають невикликами. Ось чому зазвичай використовуються ітераційні градієнтні методи.

Використовуючи зворотне розповсюдження, мережа може вивчати ваги та зміщення, позначені як $\Theta = (W, b)$, дотримуючись напрямку градієнта, який мінімізує втрати, які називають градієнтним спуском. Після цього алгоритм розподіляє зміну помилки в зворотному напрямку через мережу, ефективно змінюючи θ за рахунок невиклимих значень з великим впливом. Зворотне поширення безперервно повторюється, поки функція вартості не досягне мінімуму [28].

Незважаючи на те, що концепція НМ існувала деякий час, інтерес різко зріс в останні роки завдяки величезному прогресу в обробці. Сьогодні нейронні мережі є ключовим елементом багатьох широко використовуваних

парадигм навчання, таких як контрольоване навчання, неконтрольоване навчання та навчання з підкріпленням. Крім того, концепція нейронних мереж розвинулась у різні варіації, які краще використовувати для вирішення конкретних завдань, таких як розпізнавання мови та виявлення об'єктів.

2.3 Згорткові нейронні мережі

Згорткова мережа [34], також відома як згорткова нейронна мережа та скорочено CNN або ConvNet, може бути коротко визначена як нейронна мережа, яка замінює загальне множення матриць згорткою принаймні в одному зі своїх шарів [28]. CNN є потужним інструментом для обробки зображень. Тому вони зазвичай використовуються для виявлення особливостей на зображеннях.

2.3.1 Згортка

Неперервна згортка визначається інтегралом:

$$s(t) = \int x(a)w(t - a)da.$$

де перший аргумент x називається входом, w є ядром, а вихід іноді називають картою функцій. Операція згортки зазвичай позначається зірочкою $s(t)=(x * w)(t)$. Щоб краще зрозуміти згортку, важливо спочатку розпізнати кореляцію.

Кореляція є результатом зміщення центру ядра над входом і обчислення суми добутків у кожному місці. Згортка виконується за тією ж процедурою, що й кореляція, але вона відрізняється своїм ядром, яке повертається на 180° . Згортка відіграє важливу роль у обробці цифрових зображень, де кожне ядро призводить до певного ефекту на кінцевому зображенні, наприклад

згладжування або підвищення різкості. Для двовимірного цифрового зображення I та ядра K дискретизована форма згортки стає:

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n).$$

заміна інтеграла сумою.

2.3.2 Згортка в нейронних мережах

Кодування зображень для ШНМ вимагатиме стискання двовимірного зображення в одновимірний вектор. Однак при інтеграції згортки в нейронну мережу мережа набуває здатності використовувати просторову інформацію між пікселями, ефективно стаючи кращим у вивченні особливостей зображення [35].

Зокрема, включення згортки в нейронну мережу має три ключові переваги. Перший називається розрідженим зв'язком, що означає, що шари не пов'язані один з одним, таким чином різко зменшуючи кількість введених пікселів лише до значущих функцій. Це має величезний вплив на ефективність, оскільки потрібно менше пам'яті та виконувати менше операцій. Друга властивість – спільний доступ до параметрів, що також зменшує вимоги до пам'яті. Останньою властивістю є еківаріантні представлення, що по суті означає, що всі зміни вхідних даних будуть відображені у вихідних даних [28].

2.3.3 Згортковий шар

Загалом рівень CNN можна розділити на кілька етапів [28]:

1. Етап згортки: кілька згорток виконуються паралельно, ефективно вилучаючи із зображення високорівневі характеристики, такі як краї, і

створюючи набір лінійних активацій. Зазвичай для процесу згортання можна надати два параметри: крок і нульове заповнення. Stride визначає кількість кроків, на які ядро зміщується вздовж зображення, тоді як нульове доповнення дозволяє результуючому згорнутому об'єкту мати той самий розмір, що й вхідні дані щодо кроку. Вихід згортки можна отримати за допомогою формули $(WK + 2P)/S + 1$, де W – ширина, K – розмір ядра, P – заповнення, а S – крок [36].

2. Етап детектора: кожна лінійна активація обробляється функцією нелінійної активації. Широко використовуваною функцією активації в CNN є випрямлена лінійна функція активації (ReLU), яка відображає всі негативні значення на нуль, $\max(0;x)$. ReLU та його варіанти є кращими, оскільки вони допомагають подолати проблему зникнення градієнта.

3. Етап об'єднання: за допомогою функції об'єднання вихід мережі в певній позиції може бути реконструйований у підсумкову статистику найближчих виходів. Цей етап включено для подальшого зниження дискретизації просторових розмірів згорнутої функції, що також зменшує необхідну обчислювальну потужність. Прикладами операцій об'єднання є середнє об'єднання та об'єднання максимальних значень, які повертають середнє відповідне максимальне значення з частини зображення, охопленої прямокутною околицею.

2.3.4 Структура CNN

Структуру шару CNN можна просто описати як 3D-об'єм нейронів, який перетворює 3D-вхідний об'єм у 3D-вихідний об'єм. CNN зазвичай приймає зображення як вхідні дані та обробляє його за допомогою згорткових шарів. Для розрахунку балів класу потрібен повністю підключений вихідний рівень. Для n отриманих класів повністю зв'язаний шар можна звести до $[1 \times 1 \times n]$ об'єму, що означає, що шар згортки можна безпосередньо перетворити на повністю зв'язаний шар. Рисунок 2.3 візуалізує весь процес.

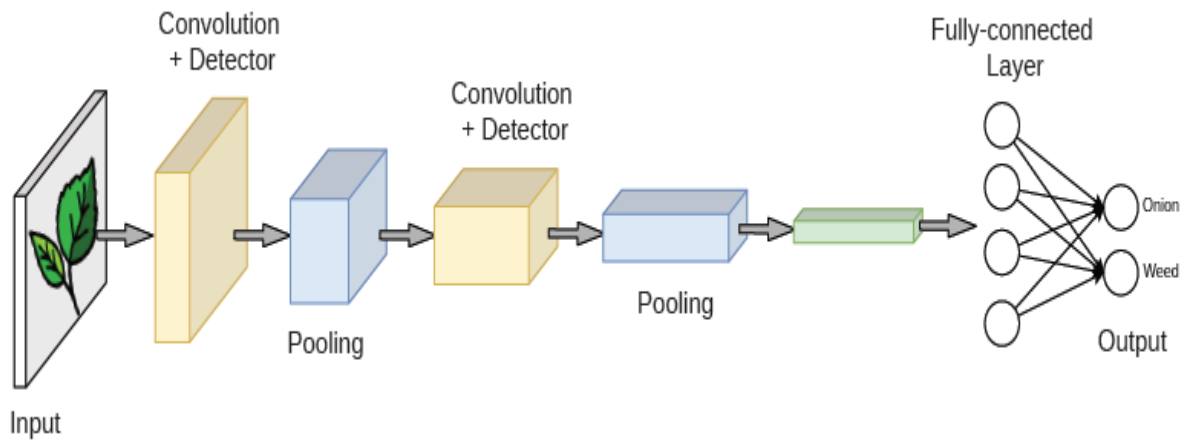


Рисунок 2.3 –Спрощений приклад того, як виглядає CNN.

CNN приймає зображення як вхідні дані, а потім витягує функції із зображення за допомогою згорткових шарів (жовтий + синій об'єми). Нарешті, вихідний обсяг вирівнюється та перетворюється на повністю пов'язаний рівень, який виводить результат мережі.

Завдяки модифікації вихідного рівня CNN на згортковий рівень структура модифікується так, щоб представляти те, що називається повністю згортковою мережею (FCN). Ця модифікація робить FCN корисними для виконання щільних прогнозів для піксельних завдань, таких як семантична сегментація [37].

2.3.5 Архітектури CNN

Точність мережі на основі CNN природно пов'язана з базовою архітектурою мережі. Таким чином, архітектура CNN постійно розвивається. Прикладами популярних мереж ConvNet є AlexNet [38], MobileNet [39] і ResNet [40].

Оскільки детектори об'єктів заснують свої виявлення на таких екстракторах ознак, використання швидкої/точної базової архітектури CNN забезпечує швидше та точніше виявлення об'єктів.

2.4 Виявлення об'єктів

З появою CNN системи зору стали надзвичайно здатними вирішувати проблеми, пов'язані із зором. Однією із захоплюючих областей комп'ютерного зору є виявлення об'єктів. Здобуття здатності розпізнавати та виявляти об'єкти надає безмежні можливості, такі як відстеження об'єктів, створення підписів до зображень та сегментація екземплярів, які можуть використовуватися різними програмами, такими як безпілотні автомобілі та медична візуалізація [41]. Процес виявлення об'єкта можна розділити на дві підпроблеми [41]:

1. Класифікація зображень: за допомогою патча зображення знайдіть клас, який найкраще відповідає патчу.

2. Локалізація та виявлення: на складному зображенні вирішіть, чи містить воно цікавий об'єкт, і спробуйте знайти цей об'єкт. Зазвичай об'єкт розташовується за допомогою обмежувальної рамки.

Завдання локалізації та виявлення — це простіша форма виявлення об'єктів, де метою є виявлення лише одного об'єкта. Виявлення об'єктів стає складнішим, оскільки воно спрямоване на виявлення кількох об'єктів на одному зображенні. На додаток до представлених проблем, варто згадати останню можливу задачу, яка називається сегментацією екземплярів. На відміну від виявлення об'єктів, сегментація екземпляра полягає не лише в пошуку об'єкта, але й у визначенні його меж аж до піксельного рівня. Усі чотири проблеми добре зображено на рисунку 2.4.

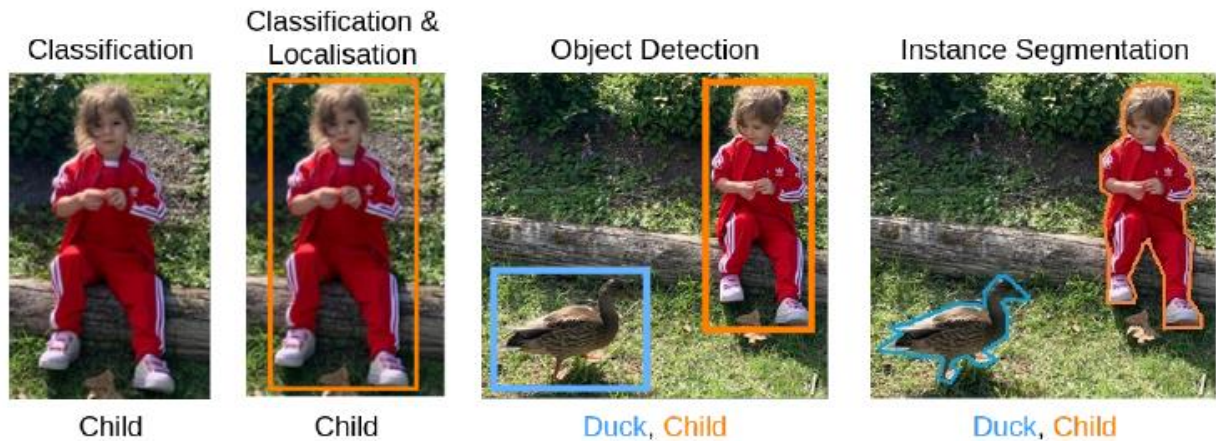


Рисунок 2.4 – Порівняння між класифікацією зображень, локалізацією, виявленням об’єктів і сегментацією примірників.

Аотрібно зауважити що два завдання ліворуч спрямовані на один об’єкт, тоді як завдання праворуч складніші, оскільки вони вражають кілька об’єктів.

Методи виявлення. Структура детектора об’єктів складається з мережі CNN, яка виділяє ознаки, які також називають магістраллю, і головки, яка передбачає класи та обмежувальні рамки об’єкта. Сучасні детектори об’єктів можуть також включати шию, яка є шарами, вставленими між хребтом і головою для збору характеристик на різних етапах [6].

Загалом методи виявлення об’єктів можна розділити на дві групи, а саме двоетапні та одноетапні методи. Двоетапний метод спочатку генерує пропозицію об’єкта-кандидата, а потім класифікує його у відповідний клас. Деякі з популярних двоетапних методів R-CNN [42], SPPnet [43] і R-FCN [44].

Однак одноетапні методи виконують два етапи одночасно, витягуючи пропозиції об’єктів і класифікуючи їх одночасно. Одними з одноетапних методів є YOLO [45], SSD [3] і RetinaNet [46].

Двома властивостями, які зазвичай відрізняють ці два методи, є швидкість і точність, тобто двоетапні методи порівняно повільні, але більш точні, тоді як одноетапні методи помітно швидші, але менш точні [47].

3 ПРАКТИЧНІ ВИМОГИ К ДОСЛІДЖЕННЮ

Вибіркове видалення бур'янів — це нова технологія, де боротьба з бур'янами виконується автономними роботами. Для правильного видалення бур'янів процес вимагає точного виявлення, яке можна отримати за допомогою контрольованих моделей виявлення.

Останніми роками спостерігався сплеск у контрольованих глибоких нейронних мережах, які досягають найсучаснішого рівня у виявленні об'єктів [49, 2, 3, 46, 45, 8], що є перспективним для вибіркового видалення. Однак контрольовані методи сильно залежать від трудомісткого процесу ручного маркування даних [61]. Отже, щоб масштабувати систему видалення бур'янів у різних середовищах, модель контрольованого виявлення вимагає більше даних, позначених вручну [17]. Таким чином, пошук процесів, які мінімізують потребу в більш маркованих даних або прискорюють загальний час анотації, забезпечать майбутню стабільність автономної системи видалення бур'янів.

Оскільки виявлення бур'янів/культур є важливою темою в цій роботі, у цьому розділі розглядаються відповідні методи виявлення бур'янів/культур. Крім того, представлені способи швидшої адаптації детектора.

3.1 Виявлення бур'янів і культур на основі навчання

Висока точність виявлення об'єктів є важливою для системи видалення бур'янів. Точний детектор найкращим чином використовує доступні дані та вивчає узагальнені характеристики, які можна масштабувати відповідно до випадків зміни середовища. Дослідження виявлення бур'янів і культур охоплюють широкий спектр різних методів. Останні дослідницькі роботи, які були знайдені, зосереджені переважно на проблемі сегментації екземплярів замість виявлення об'єктів. Оскільки сегментація екземплярів є складнішим завданням, ніж виявлення об'єктів, і обмежувальні рамки можна вивести з

сегментаційних масок, ці методи вважаються доречними для цієї роботи.

Сегментація кольорів. У [22] представлений метод виявлення бур'янів шляхом генерації маски сегментації та використання CNN для класифікації її пікселів на бур'ян або моркву. У статті докладно описано метод, який використовується для сегментації рослин від фону, який виконується шляхом перетворення зображень RGB у колірний простір HSV. Шляхом бінаризації H-каналу за допомогою зелених пікселів (G-канал) і S-каналу за допомогою вибраного порогу можна створити маску сегментації шляхом логічного поєднання отриманих масок H-каналу та S-каналу. CNN класифікує пікселі в масці сегментації як бур'ян або моркву, що забезпечує точність понад 98%.

Сегментація на основі текстури. Підхід на основі текстури до сегментування бур'янів від посівів на основі вейвлет-текстури та ШНМ представлено в [18]. Витягуючи інформацію з кількох просторових орієнтацій, вейвлет-перетворення може розрізняти текстури на даному зображенні [62]. Щоб зменшити кількість характеристик текстури, що вводяться в ШНМ, також використовується аналіз головних компонентів (PCA). Запропонований підхід здатний відрізнити бур'ян від цукрових буряків навіть за наявності значної оклюзії та перекриття листя. Це досягається при досягненні високої середньої точності бур'янів та цукрових буряків, що відповідає $95,3 \pm 2,7$ % відповідно $91,2 \pm 3,4$ %.

Сегментація стебла. Причиною виконання виявлення об'єктів у цій роботі є усунення бур'янів шляхом виведення їх стебла з виявлення бур'янів. Таким чином, внесок [63] є цікавим, оскільки він використовує повністю пов'язану згортку (FCN) для спільного виконання екземплярів сегментації та визначення положення стебла культури та бур'янів. Цей метод кодує введені зображення у високоінформативну карту ознак, яка потім декодується як стеблом, так і рослинним декодером, ефективно витягуючи особливості зображення спеціально для кожного завдання. Це призводить до екземпляра сегментованої маски для рослин і для стебел, які додатково аналізуються, щоб визначити розташування стебла. Результатом цього методу є $mAP = 79,2$ для

виявлення стебла та $mAP = 83,8$ для сегментації об'єкта на одному з досліджуваних наборів даних.

3.2 Швидша адаптація

У цьому розділі представлено широкий розділ методів, спрямованих на швидшу адаптацію моделі за рахунок кращого використання даних і стратегій анотації. Він починається з представлення відповідних досліджень машинного навчання, які можна застосувати для вирішення цієї проблеми. Розділ завершується дослідженнями, більш пов'язаними із застосуванням точного землеробства.

3.2.1 Машинне навчання

Альтернативи обмежувальної рамки Одним із методів швидшої адаптації моделі є швидше анотування даних. У [64] Papadopoulos et al. пропонує метод для швидшого маркування даних. Замість використання обмежувальних рамок анотатор позначає об'єкти за допомогою 4 точок. Цей метод забезпечує таку ж якість етикетки, але в 5 разів швидше, ніж традиційне анотування.

Команда авторів [64] досліджувала подібну тему, запропонувавши метод під назвою екстремальне клацання. Метод навчає детектори об'єктів за допомогою анотації клацання по центру до двох клацань. Експериментальні результати на PASCAL VOC 2007 і 2012 років показують, що навчені детектори отримують точність у діапазоні, навченому за допомогою обмежувальних рамок. Крім того, метод анотації надзвичайно швидкий, оскільки він скорочує загальний час анотації в 9-18 разів.

Альтернативним підходом є лише перевірка обмежувальних рамок без їх фактичного анотування. Пападопулос та ін. надає метод [65], який генерує мітки шляхом перевірки. Навчання на мітках, створених за допомогою

перевірки людиною, призводить до подібної продуктивності до міток обмежувальної рамки, при цьому загальний час анотації скорочується в $6 \times - 9 \times$.

Людина в циклі. Автори [66] вважають, що об'єднання зусиль людини та машини може призвести до ефективного та точного виявлення всіх об'єктів на зображенні. Таким чином, у статті пропонується метод із використанням марковського процесу прийняття рішень, який інтегрує кілька моделей комп'ютерного зору з кількома джерелами людського введення для точного маркування даних. Оцінка системи показує, що комп'ютерний зір і людський вхід є взаємовигідними, оскільки комп'ютерний зір може майже втричі збільшити точність людського маркування менш ніж за 30 секунд.

Анотація відео. В [67], пропонує метод, який локалізує об'єкти в наборі відео цільового класу. Роблячи це, метод може запропонувати просторово-часові обмежувальні рамки, які можна використовувати як щотижневі контрольовані мітки. Оскільки нерухомі зображення мають вищу якість, ніж відеокадри, стаття поєднує слабко анотовані відео та повністю анотовані нерухомі зображення для навчання. Це призводить до покращення продуктивності детектора порівняно з тренуванням лише на нерухомих зображеннях.

Подібна проблема називається косегментацією. Косегментація — це проблема одночасного сегментування загальної категорії об'єктів із кількох відео або зображень [68]. Ця проблема розглядається в [69], пропонуючи метод спільного семантичного зіставлення та косегментації об'єктів, який можна навчити на слабко контрольованих мітках у формі пар зображень, що містять спільні об'єкти.

3.2.2 Точне землеробство

Активне навчання. В [70] можливість використання активного навчання для скорочення часу маркування зернових культур. Запропонований метод доповнює стандартні методи активного навчання, використовуючи щотижневі

контрольовані анотації у формі клацання в центрі об'єкта, щоб визначити, які зображення є більш ефективними для навчання. Виведені зображення потім позначаються анотаціями обмежувальної рамки. У статті зроблено висновок, що запропонований метод скорочує час маркування до 55% порівняно зі звичайним маркуванням.

Трансфер навчання. Один із підходів полягає в тому, щоб передати знання з більш узагальнених попередньо підготовлених моделей до моделей, що стосуються конкретного завдання. В [71], що перенесення навчання між культурами можливе, і це може скоротити час навчання до 80% порівняно з навчанням для нуля. Стаття також демонструє, що з мінімальними зусиллями щодо маркування класифікатор CNN можна точно налаштувати на неідеально позначених зображеннях. Іншим підходом є розширення трансферного навчання за допомогою напівконтрольованого навчання.

Дослідження, опубліковане в 2018 році, показує, що можна поєднати напівконтрольоване навчання з трансферним навчанням зі значно вищими перевагами, отриманими, коли домен джерела даних значно відрізняється від домену цільового завдання. Крім того, деякі напівконтрольовані методи, такі як Pseudo-Label [73], можуть просунути повністю контрольовані базові лінії.

Навчання без нагляду. Оскільки зображення польових культур/бур'янів можна дешево отримати за допомогою роботів-прополювачів або безіменних літальних апаратів, пошук автономного методу для маркування даних є корисним. На відміну від навчання під наглядом, моделі без нагляду не потребують позначених даних.

Дослідження, опубліковане в 2019 році [61] розглядає два нещодавні алгоритми глибокої кластеризації без нагляду, Спільне неконтрольоване навчання глибоких представлень і кластерів зображень (JULE) і Глибоке кластеризування для неконтрольованого вивчення візуальних функцій (DeepCluster), використовуючи два загальнодоступні набори даних бур'янів. Результати статті показують, що можна досягти точності 87% на одному з наборів даних за допомогою DeepCluster. Крім того, поєднання збільшення

даних і навчання передачі (попереднє навчання на ImageNet) з методами кластеризації підвищує найкращу точність до 92%. Останнім внеском статті є пропозиція напівавтоматичного алгоритму маркування даних, де невелика кількість кластерів вручну позначається людиною для прискорення маркування. Таким чином можна досягти 97% точності класифікації при зменшенні кількості ручних анотацій у 100 разів. Робота авторів разом із нещодавніми розробками в галузі самостійного/неконтрольованого навчання [74], де деяким з них вдалося перевершити контрольоване попереднє навчання [75, 76], передає багатообіцяючі можливості для виявлення бур'янів.

Адаптація онлайн-детектора. Розширення роботи [77] в [17] обговорює ту саму проблему, що й цей проект. У статті пропонується напівконтрольований метод для адаптації детекторів бур'янів і цукрових буряків між різними середовищами, оскільки навчальний розподіл характеристик змінюється від розподілу спостережень. Онлайн-візуальний класифікатор доповнюється незалежним геометричним класифікатором, який моделює просторову інформацію поля. Це робиться для того, щоб використовувати той факт, що багато культур висаджуються в рівновіддалені ряди. Геометричний класифікатор заснований на двох різних імовірнісних моделях, що представляють розташування цукрових буряків і бур'янів. Особливістю геометричного класифікатора є те, що він навчається на відносно невеликому наборі даних, що вимагає менше однієї хвилини зусиль для маркування. Візуальний і геометричний класифікатори можна комбінувати для кращої класифікації, онлайн-адаптації візуального класифікатора та для напівконтрольованого навчання. Метою напівконтрольованого підходу є генерація нових навчальних даних шляхом зберігання міток, коли геометричний класифікатор є впевненим, а система зору – ні. Метод дозволяє ідентифікувати цукрові буряки та бур'яни з точністю понад 95%.

Подібна робота представлена в [24]. Оскільки дані, отримані та зібрані в цьому проекті, базуються на відеозаписах, [24] є тісно пов'язаною роботою, оскільки вона зосереджена на адаптації детектора за допомогою

послідовностей зображень. Основним внеском статті є послідовний модуль, який використовує геометричну структуру культур, представлених у послідовності зображень, для онлайн-адаптації візуального класифікатора. Це стало можливим завдяки представленню FCN, який інтегрує просторово-часовий декодер разом із візуальним декодером для цукрових буряків і сегментації бур'янів. Експериментальні результати показують, що запропонована система добре узагальнює на раніше невидимі поля за різних умов навколишнього середовища та різних стадій росту рослин без повторного навчання моделі.

4 МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ

Метою цієї роботи є адаптація онлайн-системи виявлення бур'янів/культур до нових умов, вимагаючи якомога менше ручної роботи. Цей розділ починається зі знайомства з напівконтрольованим адаптивним навчанням, яке є внеском цього проекту. Потім він глибше розглядає компоненти методу, тобто інструмент анотації та виявлення об'єктів.

У компоненті виявлення об'єктів моделі виявлення об'єктів навчаються для досягнення вищої точності. Ці моделі вищого рівня виявлення краще адаптуються за допомогою підобрених міток, створених інструментом анотацій.

4.1 Напівконтрольоване адаптивне навчання

Щоб реалізувати відповідну архітектуру системи для проблеми швидшої адаптації моделі, необхідно окреслити можливі рішення. Розглянуті рішення: швидша анотація даних і підвищення точності моделі виявлення об'єкта. Ці дві можливості можна використати наступним чином:

1. Швидша анотація даних: важко досягти швидкої адаптації, маючи обмежувальні рамки вручну. Альтернативою є автоматичне створення міток. Такі мітки не можна вважати повністю правдивими, оскільки вони не контролюються людиною. Таким чином, неконтрольовані комп'ютерні мітки будуть віднесені до псевдоміток.

2. Підвищення точності виявлення: покращуючи точність виявлення моделі, модель вивчає кращі функції, тобто отримує більше узагальнених знань щодо цільових класів. Отже, вдосконалені знання моделі більш масштабовані для нових середовищ, що мінімізує втрату точності.

Поєднання двох підходів разом створює взаємну залежність, де кожен підхід прагне посилити інший. Приклад того, як може виглядати ця

залежність, зображено на рисунку 4.1

У цій архітектурі передбачається, що онлайн-модель має найвищу можливу точність з урахуванням набору даних. Щоб створити псевдомітки, модель виявляє на зібраних зображеннях із нового середовища. Завдяки цьому інструмент анотації, який обробляє виявлення для їх покращення, може генерувати псевдомітки. Обробка виявлень можлива за допомогою додаткових даних датчика, які зберігаються під час захоплення зображень. Нарешті, експортовані псевдомітки використовуються для повторного навчання онлайн-моделі.

Ця архітектура включала існуючі мітки та псевдомітки. Таким чином, це вважається формою напівконтрольованого навчання.

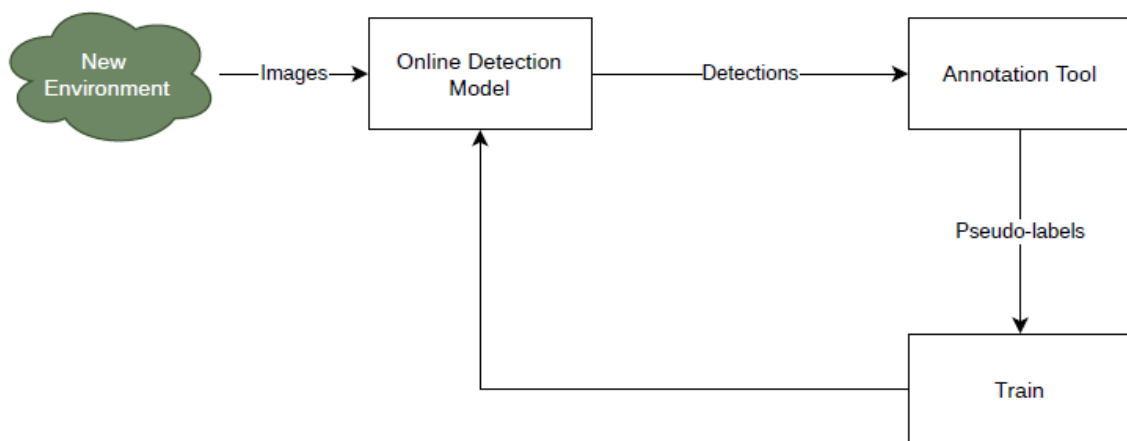


Рисунок 4.1 – Адаптація моделі за допомогою псевдоміток.

Однак важливо пам'ятати, що створення псевдоміток не обов'язково відбувається в реальному часі. Таким чином, пропонується модифікація попередньої архітектури, що призвело до головного внеску цього проекту під назвою Напівконтрольоване адаптивне навчання (SAL).

SAL навчає дві окремі моделі виявлення об'єктів, зокрема модель виявлення в автономному режимі, яка забезпечує точне виявлення для

точної офлайн-моделі в онлайн-модель, що ефективно прискорює навчання.

4.2 Апаратне та програмне забезпечення

Розробка інструменту анотації можлива за допомогою набору бібліотек програмного забезпечення та інструментів під назвою Robot Operating System (ROS), яка зазвичай використовується для роботів. У ROS виконується програма називається вузлом. Для роботи вузлів ROS потрібен ROS Master, який використовується як середовище зв'язку між вузлами. Базове спілкування між вузлами відбувається шляхом обміну повідомленнями на іменованих шинах, які називаються темами, які доступні на запущеному ROS Master. ROS також реалізує сумки, які використовуються для збереження та відтворення даних повідомлень ROS [78]. У цій роботі активно використовуються ROS-мішки, оскільки вони використовуються для зберігання даних, записаних роботом для прополки, що полегшує виконання дистанційних тестів і експериментів.

Обробка зображень і хмари точок в інструменті анотації виконується за допомогою OpenCV [79], відповідної бібліотеки хмари точок (PCL) [80]. Однак аналіз зображень виконується за допомогою MatLab [81].

Мітки набору даних візуалізуються за допомогою CVAT [82], інструменту анотації, який використовується для маркування зображень і надійного експорту їх у різні формати анотацій.

Аналіз даних і попередня обробка виконуються за допомогою скриптів Python і Bash. Навчання виконується за допомогою PyTorch [83] або Darknet [84] у хмарі за допомогою Google Colab [85] або в контейнері докерів, який працює віддалено на одному з доступних GPU в AASS в університеті Еребру. Усі графічні процесори, які використовуються в університеті, є ідентичними, тобто NVIDIA GeForce RTX 2080 Ti з 11 ГБ оперативної пам'яті. Google Colab пропонує різноманітні графічні процесори, які випадково призначаються під час кожного нового сеансу. Апаратне забезпечення, яке використовується для

навчання, недоступне для визначення на місці. Замість цього виявлення бур'янів і посівів буде розгорнуто за допомогою Jetson Xavier AGX від Nvidia, який є енергоефективним обчислювальним пристроєм, який здатний виконувати високошвидкісні висновки [86].

4.3 Інструмент анотації

Інструмент анотацій скорочує час, необхідний для позначення зображень, використовуючи датчики робота разом із системою виявлення об'єктів на основі глибокого навчання, щоб пропонувати нові мітки. Оскільки мітки, створені інструментом, не можна вважати повністю точним відображенням основної істини, їх називають псевдомітками. Отже, псевдомітки можна прирівняти до міток лише після того, як людина підтвердить їхню істинність. Таким чином, інструмент анотації призначений для створення максимально точних псевдоміток, щоб мінімізувати можливе втручання людини в процес маркування.

Для розробки інструменту анотації використовується сумка ROS, що містить виявлення та вимірювання робота. Аналізуючи доступні виявлення, виявлено три загальні проблеми: модель виявлення часто здатна виявляти об'єкти, але це не узгоджується з цим, вона може групувати кілька об'єктів як один і має тенденцію генерувати повторювані виявлення для кожного об'єкта, рисунок 4.3. Основна мета — спробувати покращити узгодженість, щоб можна було генерувати додаткові мітки. Решта дві проблеми, описані в 4.3(a) і 4.3(c), є вторинними.

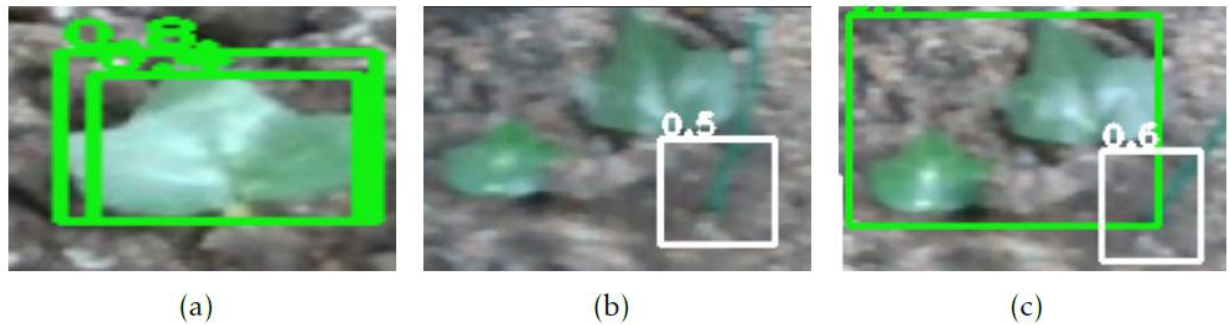


Рисунок 4.3 – Приклади поширених проблем: (а) повторювані виявлення, (б) непослідовність виявлення та (с) групування об’єктів як одного.

Таким чином, у наступних розділах буде обговорено, як інструмент анотації поступово еволюціонував у наступні методи висновку:

1. Наївний висновок: який приймає виявлення та експортує їх без подальшої обробки.

2. Висновок на основі вокселів: який поєднує датчики робота для побудови складної тривимірної карти навколишнього середовища для висновку псевдоміток.

3. Висновок на основі сегментації: який представляє методи сегментації для оптимізації локалізації та розміру міток.

4. виправлений висновок на основі вокселів: який реалізує висновок на основі вокселів після виконання уточнення пози.

4.3.1 Наївний висновок

Перша ітерація інструмента анотації розроблена з урахуванням простоти. Наївний висновок — це форма автоматичного маркування, яка вимагає мінімальної обробки. По суті, система приймає всі виявлення, фільтрує їх, використовуючи їх достовірність, а потім експортує їх як псевдомітки.

Враховуючи підхід наївного висновку, система стає сильно залежною

від точності та якості моделі виявлення об'єкта. У результаті всі загальні проблеми, зображені на малюнку 4.3, поширюються на виявлення. Крім того, навіть якщо моделі вдається знайти більшість об'єктів, немає гарантії, що модель буде краще узагальнювати, коли навчатиметься на тому, що вона вже вивчила.

Наступна ітерація суттєво змінить підхід логічного висновку, використовуючи всі доступні дані для створення складної карти світу.

4.3.2 Воксельний висновок

Правильно налаштований наївний висновок пропонує простий і точний підхід до висновку. У такому випадку виведені мітки представляють лише підмножину знань, з якою модель впевнена. На додаток до цього, наївний висновок пропонує погане використання доступних даних сенсора. Таким чином, наївний висновок покращується за допомогою воксельного висновку.

Висновок на основі вокселів використовує повний набір даних для створення тривимірного представлення світу об'єктів (корінь цибулі, бур'ян), яке називається воксельною картою зайнятості. Воксельна карта зайнятості створюється шляхом агрегування всіх зображень для створення 3D-воксельної карти, яка описує розташування кожного 2D-виявлення. Карта також призначає ймовірність для кожного вокселя, яка збільшується для кожної точки, обмеженої обсягом вокселя. Таким чином, карту можна використовувати для обробки виявлень, фільтруючи їх за допомогою ймовірностей вокселів або виводячи більше міток, також використовуючи розташування вокселів.

Процес створення воксельної карти зайнятості можна побачити на рисунку 4.4. Спочатку створюється хмара точок з використанням даних RGB і глибини. Після цього хмара точок розбивається на більш дрібні хмари точок об'єктів за допомогою виявлень. Кожна хмара точок об'єкта може представляти або корінь цибулі, або бур'ян. Для створення воксельної сітки

кожна хмара точок об'єкта перетворюється на світовий кадр за допомогою трансформації камери. Нарешті, хмара точок об'єкта воксалізується та вставляється у воксельну сітку зайнятості.

У наступних частинах буде пояснено, як працює логічний висновок на основі вокселів з посиланням на компоненти та процеси, показані на рисунку 4.4.

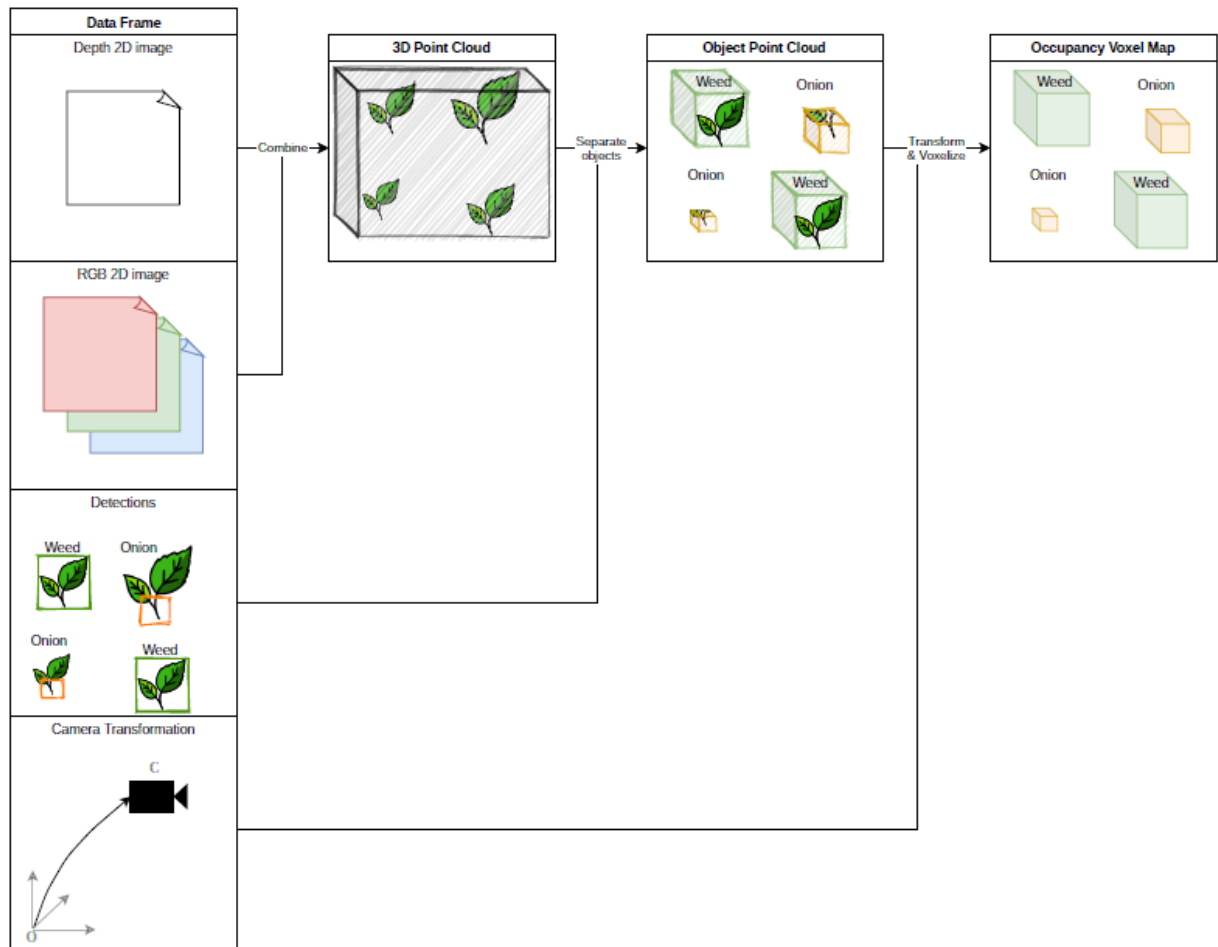


Рисунок 4.4 – Процес створення воксельної карти зайнятості.

Створення хмари точок об'єктів. Створення хмари точок об'єднує зображення RGB і глибини разом із внутрішніми параметрами камери. Використовуючи модель камери з отворами, зображену на рисунку 4.5, пікселі зображення проєктуються з площини зображення в 3D-світ. Це встановлюється основним зв'язком між координатами площини зображення та

фізичним світом.

$$\frac{\tilde{p}_x}{p_x} = \frac{\tilde{p}_y}{p_y} = \frac{p_z}{f},$$

где $(\tilde{p}_x, \tilde{p}_y) = (\tilde{x}, \tilde{y}) - (c_x, c_y)$.

Зазвичай неможливо знайти проєкцію пікселя на реальний світ, оскільки глибина, p_z , недоступна. На щастя, у цьому випадку є зображення глибини, яке дає змогу виконати перетворення. Щоб дізнатися більше про те, як працює цей процес, можна ознайомитися з алгоритмом 1.

Алгоритм показує процес генерації хмари точок \mathbf{P} із заданим RGB-зображенням I_{rgb} , ідентифікатором зображення глибини та внутрішніми параметрами камери c_x , c_y , f_x і f_y .

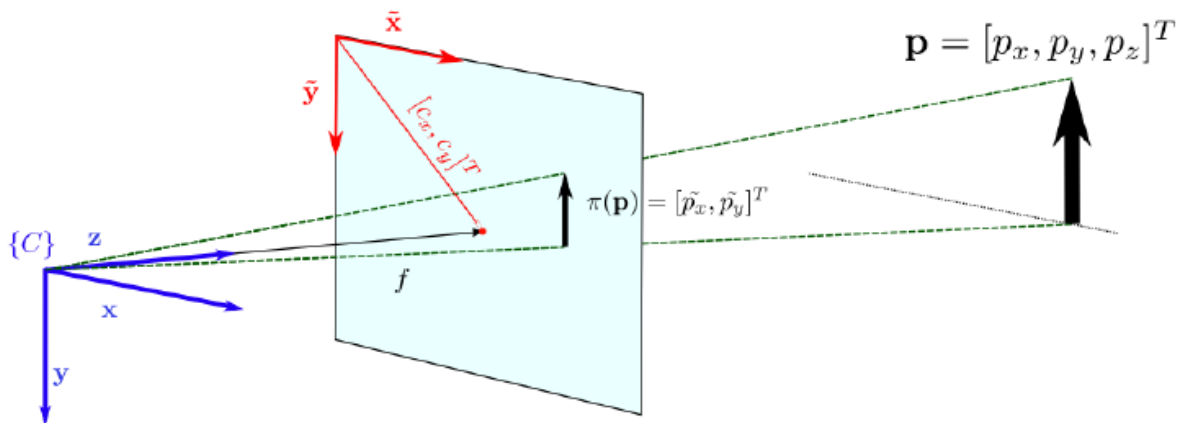


Рисунок 4.5 – Модель камери Pin-hole.

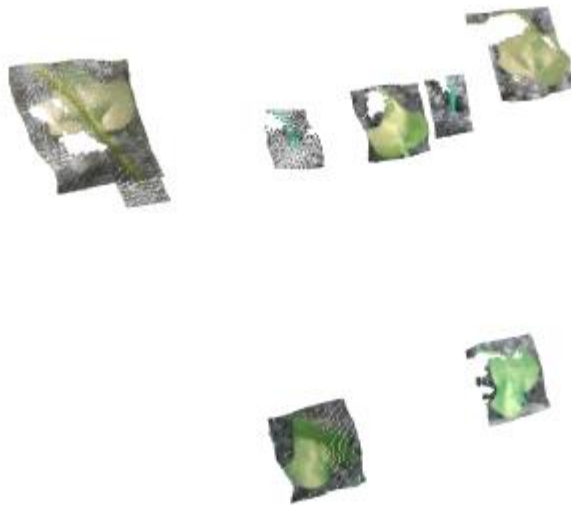
Хмара точок візуалізується за допомогою rviz і PCLVisualizer, щоб підтвердити її правильність. Дивіться рисунок 4.6(a) для візуалізації.

Незважаючи на те, що алгоритм 1 використовується для генерації хмари точок для всієї сцени, його можна легко модифікувати для генерації хмари точок об'єкта, тобто замість ітерації по всьому зображенню можна генерувати

лише хмари точок, які відповідають виявленням, дивіться рисунок 4.6(b). При цьому нерелевантна фонові інформація відкидається. Хмару точок об'єкта не обов'язково використовувати повністю, тобто достатньо використовувати лише центр хмари точок. Тим не менш, добре мати його легкодоступним для можливих майбутніх ітерацій.



(a)



(b)

Рисунок 4.6 –Згенерована хмара точок, нанесена за допомогою PCLVisualizer (a); хмари точок об'єкта бур'янів і коренів цибулі (b).

Воксельна карта зайнятості. Воксельна карта зайнятості представляє потужний метод для інкапсуляції доступних даних в одному місці. Його структура натхненна картами сітки зайнятості, які використовуються в

програмах картографування [87]. Замість двовимірної сітки воксельна карта представляє інформацію про зайнятість за допомогою 3D-вокселів із використанням хмар точок об'єктів. Це означає, що кожен воксель має попередньо визначений розмір одиниці, визначений роздільною здатністю карти вокселів.

Щоб реалізувати воксельну карту зайнятості, кожен воксель підтримує імовірнісну модель, яка описує, наскільки ймовірно, що певний воксель представлятиме об'єкт реального світу або буде незайнятим. Отже, воксельну карту зайнятості можна використовувати для маркування шляхом перетворення й проектування ймовірних вокселів назад на площину зображення. Крім того, його можна використовувати для відхилення помилкових спрацьовувань.

Вокселізація. Щоб полегшити аналіз карти світу, карта світу спрощена за допомогою попередньо визначеної роздільної здатності до однакових 3D-клітин. Це робить можливим кластеризувати найближчі трансформовані точки разом у так звані вокселі. Використання вокселів забезпечує спосіб керування вхідними даними та прийняття рішень на їх основі. Оскільки існують підкарти, кожна з яких представляє клас, немає необхідності, щоб воксель мав власний клас. Тим не менш, кожен воксель ві використовує імовірнісну модель спостереження, щоб визначити його ймовірність бути зайнятим об'єктом, а саме $p(v_i)$.

Модель спостереження. Щоб мати можливість змоделювати зайнятість вокселя, необхідно зробити два припущення: (i) кожен воксель приймає двійковий стан, який позначає порожній або зайнятий воксель, тобто $p(v_i) = 0$ і $p(v_i) = 1$ і середовище статичне. Ці припущення допомагають дискретизувати проблему та забезпечують хорошу оцінку, враховуючи, що роздільна здатність карти є прийнятною для пристосування до припущення (i).

Імовірнісне відображення вокселів. Об'єднавши камеру для трансформації світу, воксельну абстракцію та модель спостереження, можна зібрати карту. Весь процес створення карти зводиться до алгоритму 2.

Алгоритм оновлює воксельну карту з урахуванням нового розташування та спостережень. Це виконується шляхом ітерації доступних хмар точок об'єктів, знайдених у t , і оцінювання того, чи вже існують їхні відповідні точки у воксельному обсязі v_i . Якщо так, модель спостереження для відповідного вокселя оновлюється відповідно. В іншому випадку створюється новий об'єм вокселя, який містить запитувану точку, і додається до карти вокселів, рисунок 4.7.

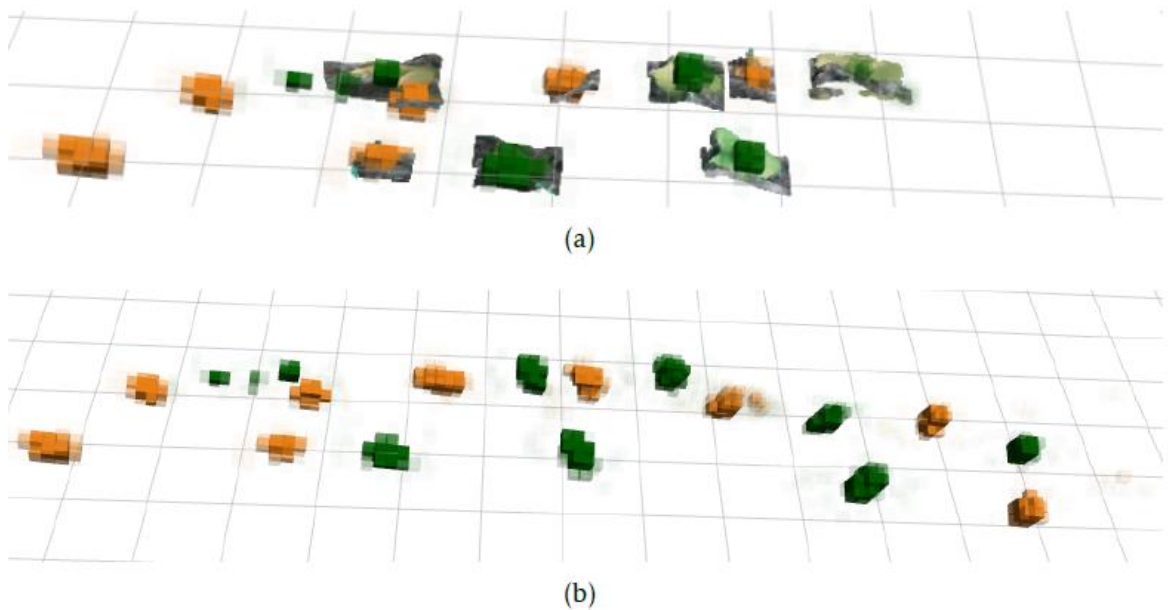


Рисунок 4.7 – Візуалізація воксельної карти зайнятості, створеної за допомогою алгоритму 2.

Рівень прозорості кожного вокселя прямо корелює з його ймовірністю. (a) алгоритм у дії: хмари точок об'єктів перетворюються на вокселі. (b) показує завершену воксельну карту зайнятості.

Серіалізація

Щоб зробити воксельну карту доступною для подальшого аналізу, відповідні дані в пакетному файлі ROS разом із воксельною картою серіалізуються та зберігаються на диску. Це дозволяє легко отримати карту для обробки незалежно від методу висновку.

4.2.2.5 Висновок

Представлене тут поняття воксельної карти зайнятості є основою, яку можна формувати в різні методи, як це буде продемонстровано в наступних розділах. Тим не менш, для визначення мітки обмежувальної рамки потрібно отримати як центр виявлення, так і розміри (ширину та висоту) обмежувальної рамки. У цьому сценарії виведення воксельна карта використовується для фільтрації результатів іншого методу.

Метод проектує центри виявлення у світовий кадр за допомогою трансформації камери. Через те, що в показаннях датчика присутній шум, спроектовані 2D-центри виявлення не вирівнюються ідеально, коли вони потрапляють у світову рамку, рисунок 4.8(b). Таким чином, отримані центри виявлення 3D-об'єктів кластеризуються для визначення центру, який представляє кожен об'єкт. Це пов'язано з тим, що поточна карта світу описує лише певний клас, тобто є дві підкарти, одна для кореня цибулі (помаранчева), а інша для бур'янів (зелена), рисунок 4.8(a). Таким чином, карта не описує окремі базові об'єкти в кожній підкарті. При цьому кластеризація найближчих центрів виявлення призводить до уніфікованого центру виявлення, що позначає один конкретний об'єкт, рисунок 4.8(c). Для досягнення цього використовується метод кластеризації під назвою DBSCAN [88]. Перевага використання DBSCAN полягає в тому, що він не вимагає введення кількості кластерів, що робить його придатним для кластеризації невідомої кількості центрів виявлення. Реалізація DBSCAN заснована на [89].

Нарешті, розміри обмежувальної рамки визначаються шляхом усереднення розмірів згрупованих центральних точок.

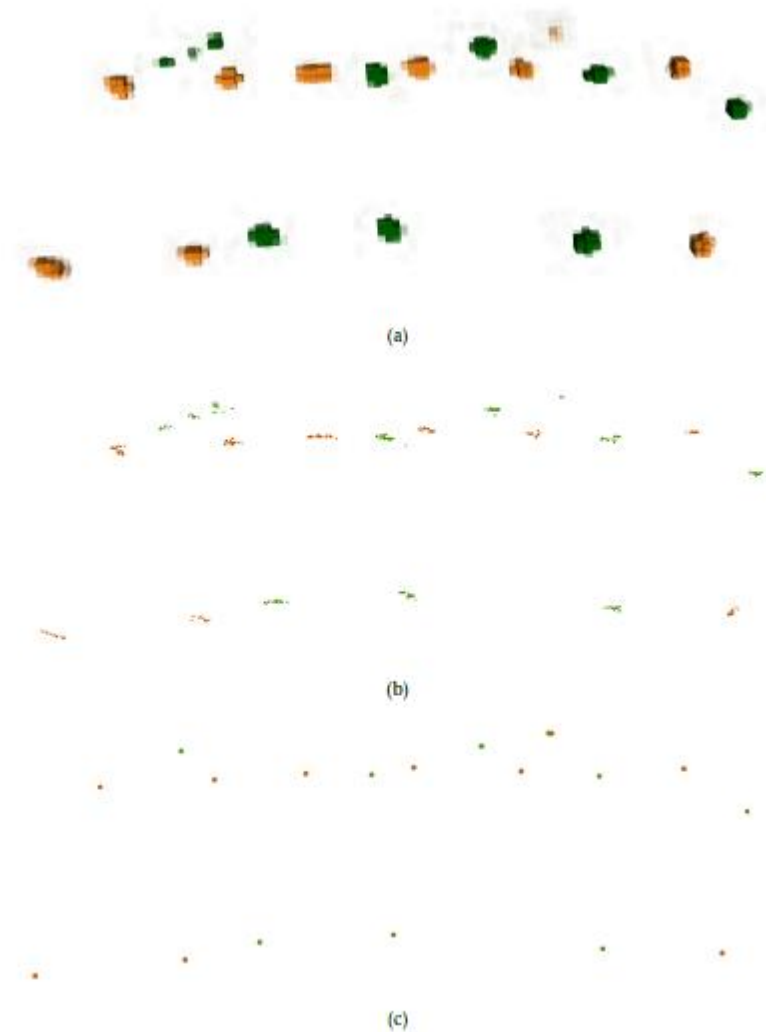


Рисунок 4.8 – Кластеризація за допомогою DBSCAN: а) воксельна карта зайнятості; б) центри виявлення, спроектовані та трансформовані у світову рамку; с) масштабовані кластерні центри виявлення.

4.3.3 Висновок на основі сегментації

Хоча воксельна карта зайнятості є потужним методом виведення інформації з навколишнього середовища, її точність сильно залежить від правильної локалізації. Таким чином, помилки одометрії коліс поширюються через мітки, що призводить до погано локалізованих міток обмежувальної рамки. У спробі мінімізувати помилки локалізації вводиться сегментація на основі кольорів, щоб мінімізувати помилки в кадрі зображення. Це навіяно

візуальним обслуговуванням [90].

Ефективним методом сегментації рослин є використання камери RGB+NIR (ближнього інфрачервоного діапазону), оскільки рослини мають високу відбивну здатність у спектрі NIR [23]. Оскільки цей тип апаратного забезпечення недоступний, натомість досліджуються альтернативні методи, які потребують лише даних RGB. Аналіз ґрунтується на загальноживаних індексах рослинності, див. [91, 21], а саме ExG (надлишок зеленого), ExG-ExR (надлишок зеленого – надлишок червоного), NDI (індекс нормалізованої різниці рослинності) та CIVE (індекс кольору Vegetation Extraction), а також альтернативний колірний простір HSV.

Усі індекси поєднують кілька каналів зображення з використанням різних методів для створення зображення в градаціях сірого, яке виділяє рослини. Таким чином, при оцінці індексів враховуються два чинники, а саме те, чи можна легко розрізнити рослини на отриманому зображенні та чи можна постійно обмежувати порогове значення отриманого зображення для створення бінарних масок. Бажано, щоб бінарна маска відображала точну сегментацію рослин. Крім того, якщо потрібно, маска повинна більше відображати рослини, ніж недостатньо. Усі доступні зображення сегментуються за допомогою 4 індексів і вибираються вручну, щоб отримати оцінку того, наскільки добре зображено рослини

ВИСНОВКИ

Основна мета цієї роботи полягає в тому, щоб адаптувати детектор бур'янів/культур у режимі реального часу до нових умов з мінімізацією ручних зусиль. Рішення, яке має намір вирішити цю проблему, має відповідати трьом вимогам: надійність, швидкість та ефективність.

У цій роботі запропоновано, реалізовано та оцінено систему напівконтрольованого адаптивного навчання (SAL), яка складається з двох компонентів, а саме інструменту анотації та виявлення об'єктів. За допомогою експериментальної оцінки показано, що система відповідає всім вимогам:

1. Надійність: надійність системи можна визначити лише шляхом перевірки надійності інструменту анотації. Це тому, що інструмент анотації є єдиним компонентом, який обробляє введені дані. Тим не менш, для інструменту анотації були запропоновані різні методи висновку, серед яких найпростіший і воксельний висновок. Ці два методи вважаються надійними, оскільки вони забезпечують загальні рішення, які вимагають мінімального втручання людини незалежно від середовища. Тому SAL вважається надійним методом.

2. Швидкість: SAL надзвичайно швидкий, оскільки він виконує адаптацію за величину хвилин, тоді як ручна анотація займала б величину тижнів. Під час експерименту SAL вдалося адаптувати онлайн-модель менш ніж за 10 хвилин, тоді як анотація вручну тривала майже 2 тижні.

3. Ефективність: експериментальні результати показують, що моделі, адаптовані за допомогою SAL, досягають подібної точності до адаптації з використанням базових міток істинності. Точніше кажучи, точність mAP50, отримана за допомогою SAL, лише на 2% відрізняється від точності, отриманої під час навчання на наземних правдивих даних.

Таким чином, SAL відповідає всім вимогам і пропонує чудову адаптацію моделі щодо необхідного часу виконання.

Робота в основному обговорювався з технічної точки зору. Однак важливо пролити світло на інші економічні, соціальні та етичні аспекти. Екологічний аспект не обговорюється, оскільки він уже був представлений у вступі.

Соціальні та етичні наслідки. Цікаво обговорити суперечливу тему про те, як технології замінюють майбутні робочих місць. Ця тема дуже пов'язана з цим проектом із соціальної точки зору через такі причини: (i) проект має на меті покращити зручність використання автономних роботів для прополки, що призводить до меншої залежності від людей. (ii) Запропоноване рішення для (i) зменшує залежність від людей-анотаторів. Однак, коли цю тему далі обмежувати сферою цього проекту, тобто сільським господарством, можна виявити, що її характеристики кардинально змінюються.

Національна робоча сила у сільському господарстві постійно скорочується, оскільки в період з 2003 (13,1 мільйона) до 2018 (9,1 мільйона) робоча сила скоротилася на тривожні 30% [101]. Це призвело до того, що ЄС залежить від 1 мільйона сезонних робітників щорічно, щоб заповнити цю прогалину [102]. Тому розробка таких технологій, як сільськогосподарські роботи, необхідна для забезпечення продовольчої безпеки. Пандемія COVID-19 продемонструвала, як ця проблема може загостритися, коли сезонна робоча сила не могла подорожувати у 2020 році через обмеження на поїздки [102].

Незрозуміло, як резонувати щодо етичного аспекту проблеми, особливо щодо сезонних працівників. Це можна розглядати як етичну дилему, оскільки використання автономних роботів для прополки є вигідним для ЄС, але це вплине на засоби до існування сезонних працівників. Проблема ускладнюється, якщо врахувати, що реальна кількість сезонних працівників насправді ще більша. Це пояснюється тим, що величезна частина сезонних працівників не має документів, тому вони не представлені в статистиці [102].

Тим не менш, приємно прочитати, що нещодавній звіт [103] передбачає, що до 2025 року завдяки ІІІ буде створено 97 мільйонів нових робочих місць, що більше, ніж 85 мільйонів робочих місць, які будуть втрачені через

автоматизацію.

Економічні наслідки. Ця робота має дуже позитивні економічні наслідки. Швидша адаптація призводить до менших витрат і часу на адаптацію, що дає змогу ефективніше використовувати автономні роботи-прополювачі для: зниження витрат завдяки кращому використанню ресурсів і автоматизації та покращення якості для отримання більш цінної сільськогосподарської продукції. Це можливо завдяки точному землеробству та його здатності забезпечувати більший контроль над сільськогосподарськими угіддями.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Jonathan Huang, Vivek Rathod, Chen Sun, Menglong Zhu, Anoop Korattikara, Alireza Fathi, Ian Fischer, Zbigniew Wojna, Yang Song, Sergio Guadarrama, and Kevin Murphy. Speed/accuracy trade-offs for modern convolutional object detectors, April 2017.
2. Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN, January 2018.
3. Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single Shot Multi-Box Detector, December 2016. doi: 10.1007/978-3-319-46448-0_2
4. Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection., April 2020.
5. Ultralytics/yolov5. Ultralytics, May 2021
6. Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Scaled-YOLOv4: Scaling Cross Stage Partial Network. arXiv:2011.08036 [cs], February 2021.
7. Food and Agriculture Organization of the United Nations, editor. The Future of Food and Agriculture: Trends and Challenges. Food and Agriculture Organization of the United Nations, Rome, 2017. ISBN 978-92-5-109551-5.
8. James E. Addicott. The Precision Farming Revolution. In James E. Addicott, editor, *The Precision Farming Revolution: Global Drivers of Local Agricultural Methods*, pages 1–35. Springer, Singapore, 2020. ISBN 9789811396861. doi: 10.1007/978-981-13-9686-1_1.
9. Robin Gebbers and Viacheslav Adamchuk. Precision Agriculture and Food Security. *Science* 327(5967), 828-831. *Science (New York, N.Y.)*, 327:828–31, February 2010. doi: 10.1126/science.1183899.
10. B. Blackmore. *The Role of Yield Maps in Precision Farming*. PhD thesis, January 2003.

11. Judith Tsouvalis, Susanne Seymour, and Charles Watkins. Exploring Knowledge-Cultures: Precision Farming, Yield Mapping, and the Expert-Farmer Interface. *Environment and Planning A: Economy and Space*, 32(5):909–924, May 2000. ISSN 0308-518X. doi: 10.1068/a32138.

12. Alessandro dos Santos Ferreira, Daniel Matte Freitas, Gercina Gonçalves da Silva, Hemerson Pistori, and Marcelo Theophilo Folhes. Weed detection in soybean crops using ConvNets. *Computers and Electronics in Agriculture*, 143:314–324, December 2017. ISSN 0168-1699. doi: 10.1016/j.compag.2017.10.027.

13. James H. Westwood, Raghavan Charudattan, Stephen O. Duke, Steven A. Fennimore, Pam Marrone, David C. Slaughter, Clarence Swanton, and Richard Zollinger. Weed Management in 2050: Perspectives on the Future of Weed Science. *Weed Science*, 66(3):275–285, May 2018. ISSN 1550-2759. doi: 10.1017/wsc.2017.78.

14. Philipp Lottes and Cyrill Stachniss. Semi-Supervised Online Visual Crop and Weed Classification in Precision Farming Exploiting Plant Arrangement. September 2017. doi: 10.1109/IROS.2017.8206403.

15. Adel Bakhshipour, Abdolabbas Jafari, Seyed Mehdi Nassiri, and Dariush Zare. Weed segmentation using texture features extracted from wavelet sub-images. *Biosystems Engineering*, 157:1–12, May 2017. ISSN 1537-5110. doi: 10.1016/j.biosystemseng.2017.02.002.

16. Petra Bosilj, Tom Duckett, and Grzegorz Cielniak. Connected attribute morphology for unified vegetation segmentation and classification in precision agriculture. *Computers in Industry*, 98:226–240, June 2018. ISSN 0166-3615. doi: 10.1016/j.compind.2018.02.003.

17. Esmael Hamuda, Brian Mc Ginley, Martin Glavin, and Edward Jones. Automatic crop detection under field conditions using the HSV colour space and morphological operations. *Computers and Electronics in Agriculture*, 133:97–107, February 2017. ISSN 0168-1699. doi: 10.1016/j.compag.2016.11.021.

18. Moacir Ponti. Segmentation of Low-Cost Remote Sensing Images

Combining Vegetation Indices and Mean Shift. *IEEE Geoscience and Remote Sensing Letters*, 10:67–70, January 2013. doi: 10.1109/LGRS.2012.2193113.

19. Florian J. Knoll, Vitali Czymmek, Sascha Poczihoski, Tim Holtorf, and Stephan Hussmann. Improving efficiency of organic farming by using a deep learning classification approach. *Computers and Electronics in Agriculture*, 153:347–356, October 2018. ISSN 0168-1699. doi: 10.1016/j.compag.2018.08.032.

20. Philipp Lottes, Markus Höferlin, S. Sander, M. Muter, Peter Schulze Lammers, and Cyrill Stachniss. An Effective Classification System for Separating Sugar Beets and Weeds for Precision Farming Applications. May 2016. doi: 10.1109/ICRA.2016.7487720.

21. Philipp Lottes, Jens Behley, Andres Milioto, and Cyrill Stachniss. Fully Convolutional Networks with Sequential Information for Robust Crop and Weed Detection in Precision Farming. *IEEE Robotics and Automation Letters*, 3(4):2870–2877, October 2018. ISSN 2377-3766, 2377-3774. doi: 10.1109/LRA.2018.2846289.

22. David Slaughter, D.K. Giles, and D. Downey. Autonomous robotic weed control systems: A review. *COMPUTERS AND ELECTRONICS IN AGRICULTURE*, 61:63–78, April 2008. doi: 10.1016/j.compag.2007.05.008.

23. Xiaolong Wu, Stéphanie Aravecchia, Philipp Lottes, Cyrill Stachniss, and Cédric Pradalier. Robotic weed control using automated weed and crop classification. *Journal of Field Robotics*, 37(2):322–340, 2020. ISSN 1556-4967. doi: 10.1002/rob.21938.

24. Tom M. Mitchell. *Machine Learning*. McGraw-Hill Series in Computer Science. McGraw-Hill, New York, 1997. ISBN 978-0-07-042807-2.

25. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

26. Lisa Torrey and Jude Shavlik. Transfer Learning. In *Handbook of Research on Machine Learning Applications*. IGI Global, 2009.

27. Burr Settles. Active Learning Literature Survey. Technical Report,

University of Wisconsin-Madison Department of Computer Sciences, 2009.

28. Mark Everingham, Luc Van Gool, Christopher K. I. Williams, John Winn, and Andrew Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-009-0275-4.

29. Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical bio-physics*, 5(4):115–133, December 1943. ISSN 1522-9602. doi: 10.1007/BF02478259.

30. Michael A. Nielsen. *Neural Networks and Deep Learning*. 2015.

31. Y. Le Cun and Yann Le Cun. Steels (Eds) 'Connectionism in Perspective', Elsevier 1989. *Generalization and Network Design Strategies*. 1989.