

Львівський національний університет імені Івана Франка  
Факультет прикладної математики та інформатики  
Кафедра дискретного аналізу та інтелектуальних систем

ДИПЛОМНА РОБОТА

На тему:

**ЗНАХОДЖЕННЯ ОБ'ЄКТІВ НА ЗОБРАЖЕННІ З ДОПОМОГОЮ  
ШТУЧНОГО ІНТЕЛЕКТУ**

Виконав: студент IV курсу, групи ПМі-45  
напряму підготовки комп'ютерні науки

Голько Л.А.

Керівник

Щербина Ю. М.

2023

<b>ВСТУП</b>	<b>6</b>
Завдання дипломної роботи	6
Актуальність теми	6
<b>РОЗДІЛ 1 ТЕОРЕТИЧНІ ВІДОМОСТІ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ТА КОМП'ЮТЕРНОГО ЗОРУ</b>	<b>7</b>
1.1 Що таке штучний інтелект?	7
1.2 Що таке комп'ютерний зір?	7
1.3 Яким чином працює комп'ютерний зір?	7
1.4 Для чого використовувати комп'ютерний зір?	8
1.5 Штучні нейронні мережі	9
1.6 Багатошарові мережі прямого поширення	10
1.7 Теоретичні відомості комп'ютерного зору	10
<b>РОЗДІЛ 2 КРОКИ ДЛЯ СТВОРЕННЯ ПРАКТИЧНОГО ЗАСТОСУНКУ</b>	<b>13</b>
2.1 Алгоритм розв'язування	13
2.2 Датасет для комп'ютерного зору	13
2.3 Практичні прийоми для обробки зображення	14
2.3.1 Розмивання Гауса	14
2.3.2 Виділення контурів об'єктів	14
2.3.3 Опорні точки	15
2.4 Тренування нейронної мережі	15
2.5 Існуючі моделі систем штучного інтелекту для розпізнавання об'єктів на зображенні	16
2.5.1 ResNet	16
2.5.2 Inception	16
2.5.3 Faster R-CNN	16
2.6 Покращення для розпізнавання вхідного відео файлу	17
2.7 Покращення для розпізнавання відео в режимі реального часу	18
2.8 Створення загальнодоступного ресурсу	18
<b>РОЗДІЛ 3 ПРИКЛАДИ ПРАКТИЧНОЇ РЕАЛІЗАЦІЇ</b>	<b>21</b>
3.1 Приклад з використанням зображень оброблених локально	21
3.1.1 Приклад №1	21
3.1.2 Приклад №2	21
3.1.3 Приклад №3	23
3.2 Приклад з використанням зображень оброблених на загальнодоступному ресурсі	24
3.2.1 Приклад №1:	24
3.2.2 Приклад №2:	25

<b>ВИСНОВКИ</b>	<b>27</b>
<b>ДОДАТОК</b>	<b>28</b>
<b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:</b>	<b>32</b>

## **ВСТУП**

Завдання дипломної роботи

Дослідити теоретичну частину систем штучного інтелекту та комп'ютерного зору. Окрім цього, створити програму, яка може розпізнавати об'єкти на зображенні та однозначно їх ідентифікувати з допомогою штучного інтелекту.

Актуальність теми

Із кожним днем розвиток систем штучного інтелекту зростає, обчислювальні ресурси стають більш доступними, а навчальні методи ефективнішими. У результаті дана галузь набуває все більшої популярності. У наші дні наявні технології можуть бути використані для створення власних систем чи їх удосконалення. Тема є назвичайно актуальною, оскільки програми створюються, щоб допомагати людям. Наприклад, для автоматизації багатьох рутинних задач і процесів, які раніше виконувалися людиною виключно вручну.

# РОЗДІЛ 1 ТЕОРЕТИЧНІ ВІДОМОСТІ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ ТА КОМП'ЮТЕРНОГО ЗОРУ

## 1.1 Що таке штучний інтелект?

Штучний інтелект — здатність інженерних систем опрацьовувати, застосовувати та вдосконалювати отримані знання та навички. Хоча немає точного стандарту те, що робить комп'ютер «розумним», було запропоновано кілька гіпотез, як-от тест Тюрінга чи гіпотеза Ньюелла-Саймона. Тест Тюрінга — це тест, створений у 1950 році для визначення здатності машини виявляти інтелектуально обумовлену поведінку, ідентичну поведінці людини і не відрізнити від неї. Основна теза цього тесту: "Людина взаємодіє з комп'ютером та людиною. Відповідаючи на запитання, вона має визначити, з ким вона розмовляє: з людиною або з комп'ютерною програмою. Мета комп'ютерної програми - обдурити людину, змусивши її зробити неправильний вибір. У той же час, користувач повинен дослідити з ким спілкується». Ідея була запропонована Аланом Тюрінгом у британському науковому журналі "Computing Machinery and Intelligence".

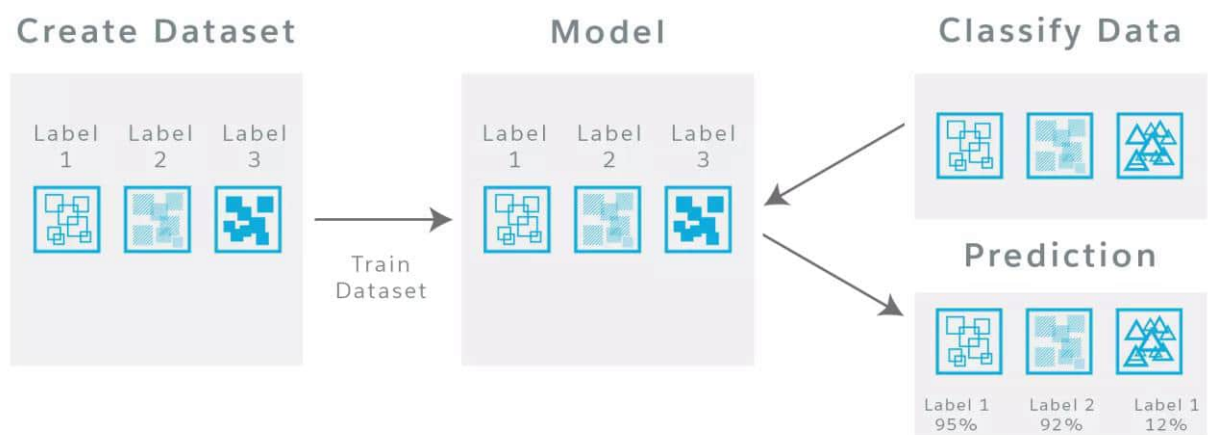
## 1.2 Що таке комп'ютерний зір?

Комп'ютерний зір – це один з напрямів штучного інтелекту, який спрямований на класифікацію і ідентифікацію предметів, на фото чи відео, які володіють певним набором властивостей та ознак. До завдань комп'ютерного зору відносяться отримання, обробка та аналіз вхідного зображення. Як наукова дисципліна комп'ютерний зір відноситься до теорії та технології створення штучних систем, які отримують інформацію за допомогою зображень. Крім того, вхідні дані можуть бути представлені в багатьох формах, наприклад, відео послідовність, зображення з різних камер або 3D-дані з медичного сканера.

## 1.3 Яким чином працює комп'ютерний зір?

Типовими функціями багатьох систем комп'ютерного зору є отримання зображень, попередня обробка, виокремлення деталей, детектування/сегментація, високорівнева обробка. Перше за все, розробляється модель, для цього потрібно

створити набір даних. Потім визначити мітки або категорії у наборі даних. Нарешті, надаються приклади даних для кожної мітки. Отже, процес навчання відбувається наступним чином: на вхід системи подається набір даних, система визначає спільні риси та відмінності між різними мітками та узагальнює характеристики, які визначають кожну мітку. Коли навчання успішно завершилось, можна використовувати модель, для цього надсилається зображення в модель, яка передбачає, до якої мітки потрапляє це зображення, на основі міток, указаних у навчальному наборі даних. На основі цього система виводить результат.



#### 1.4 Для чого використовувати комп'ютерний зір?

У сучасному світі існує безліч прикладів використання комп'ютерного зору. У промисловості інформацію отримують для підтримки виробничого процесу, наприклад, контроль якості, коли деталі чи кінцевий продукт автоматично перевіряють на наявність дефектів. Якщо це автомобільна промисловість - це машини з автопілотом, які визначають пішоходів, машини, та будь-які інші перешкоди.

У збройних силах очевидним прикладом є виявлення солдатів і транспортних засобів противника та управління ракетами. Найбільш досконалі системи управління ракетами направляють ракету в заданий район замість конкретної мети, а визначення цілей відбувається при вході ракети в заданий район на основі відео, що надходять.

## 1.5 Штучні нейронні мережі

Штучні нейронні мережі – це технологія, яка тісно пов'язана з багатьма дисциплінами: нейрофізіологією, математикою, статистикою, фізикою, комп'ютерними науками та технікою. Вони знаходять своє застосування у таких різноманітних областях як моделювання, аналіз часових рядів, розпізнавання образів, обробка сигналів та керування завдяки одній важливій властивості - здатності навчатися на основі даних за участю вчителя або без його втручання.

### Математична модель штучного нейрона

Нейрон - це одиниця обробки інформації нейронної мережі. На схемі наведено модель нейрона, яка лежить в основі штучних нейронних мереж.

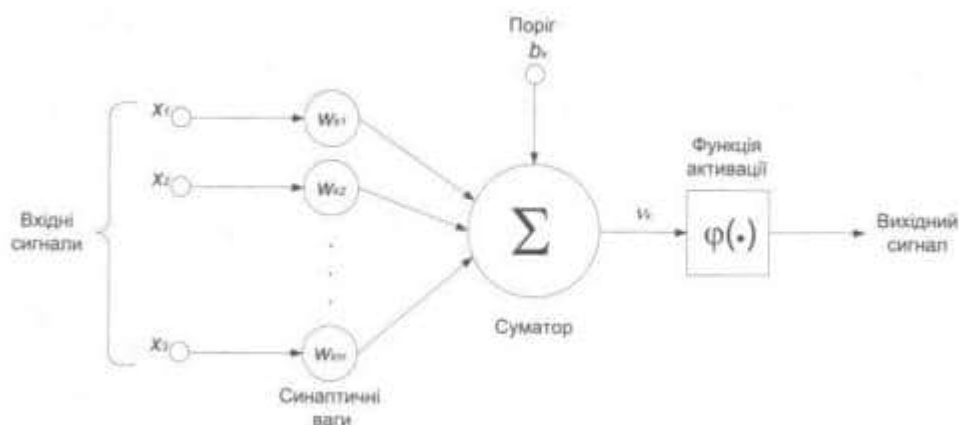


Рис. 1.1 Модель штучного нейрона

У цій моделі можна виділити три основні елементи.

1. Набір синапсів або зв'язків, кожен з яких характеризується своєю вагою або силою. Зокрема, сигнал  $x_j$  на вході синапсу  $j$ , пов'язаного з нейроном множиться на вагу  $w_{ki}$ . Важливо звернути увагу на те, в якому порядку зазначені індекси синаптичної ваги  $w_{ji}$ . Перший індекс відноситься до розглядуваного нейрона, а другий - вхідного закінчення синапсу, з яким пов'язана дана вага. На відміну від синапсів мозку синаптична вага штучного нейрона може мати як позитивні, так і негативні значення.

2. Суматор складає вхідні сигнали, зважені щодо відповідних синапсів нейрона. Цю операцію можна описати як лінійну комбінацію.

3. Функція активації обмежує амплітуду вихідного сигналу нейрона. Ця функція також називається функцією стиснення. Зазвичай нормальний діапазон амплітуд виходу нейрона належить інтервалу  $[0;1]$  або  $[-1;1]$ .

Модель нейрона містить пороговий елемент, який позначений символом  $b_k$ . Ця величина характеризує збільшення або зменшення вихідного сигналу, що подається на функцію активації.

### 1.6 Багатошарові мережі прямого поширення

Нейронні мережі прямого поширення характеризується наявністю одного або декількох прихованих шарів, вузли яких називають прихованими нейронами, або прихованими елементами. Функція останніх полягає в посередництві між зовнішнім вхідним сигналом та виходом нейронної мережі. Долучаючи один або кілька прихованих шарів, ми можемо виділити статистики високого порядку. Така мережа дає змогу виділяти глобальні властивості даних за допомогою локальних з'єднань завдяки наявності додаткових прихованих нейронів виділяти статистичні залежності високого порядку особливо суттєва, коли розмір вхідного шару досить великий.

### 1.7 Теоретичні відомості комп'ютерного зору

Методи комп'ютерного зору використовуються для вирішення завдань, які умовно можна розділити на прості і складні. Складні завдання відповідають на питання, який об'єкт зображений на картинці, до якого класу він належить. Для вирішення цих завдань найчастіше використовуються методи машинного навчання. При вирішенні найпростіших завдань виробляються маніпуляції безпосередньо з пікселями, використовуються евристики, а методи машинного навчання, як правило, не застосовуються. Завдяки досягненням у галузі штучного інтелекту та інноваціям у глибинному навчанні та нейронних мережах, ця галузь змогла зробити великі стрибки за останні роки та змогла перевершити людей у деяких завданнях, пов'язаних із виявленням та маркуванням об'єктів.

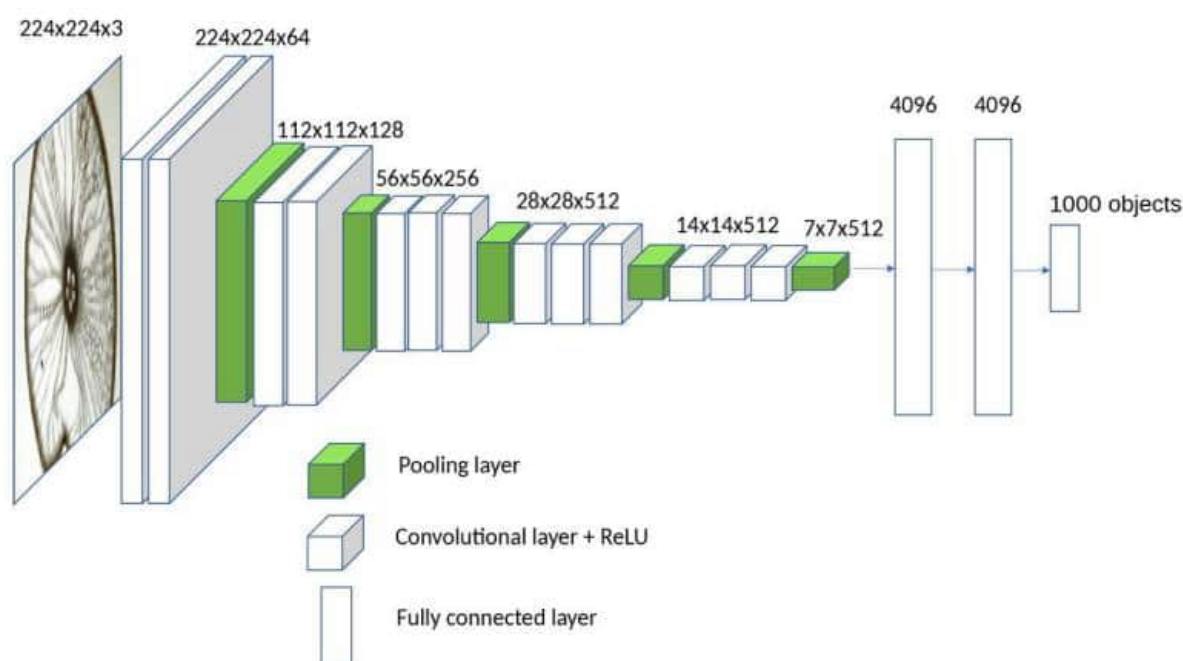


Для розуміння вмісту зображення людина використовує весь свій життєвий досвід, освіту, а іноді і інтуїцію. На сьогодні, «сильного» штучного інтелекту, що за своїми можливостями співмірний до людини поки не створено, тому, розробники систем комп'ютерного зору вирішують тільки певні задачі.

Зорове сприйняття людини є досконалим надбанням природи. Людський мозок - це унікальна система розуміння і обробки тієї картинки, яку бачить людина. Вона схильна бачити те, що хоче бачити, і цьому не можна навчити комп'ютер. Мозок часто шукає знайомі образи, «добудовує» картинку і додає семантику.

Людська зорова система самонавчається. Європейцеві складно розрізнити обличчя азіатів, оскільки, зазвичай, в житті він рідко їх зустрічає. Його зорова система навчилася вловлювати відмінності в європейських особах, а азіати, яких він бачив мало, здаються йому «на одне обличчя» і навпаки. Але достатньо провести довший час в Азії, європеєць помічає відмінності обличчя і без ускладнень розпізнає осіб. Ключовою проблемою у складності розпізнавання вмісту зображення є різниця у семантичному сприйнятті картини людиною і комп'ютером. Людина легко виділяє на зображенні важливе і неважливе, наповнює її змістом, за не явними ознаками розуміє багато деталей.

Однією з найкращих моделей, на сьогодні є нейромережа VGG16, яка розпізнає 1000 різних об'єктів. Нижче наведено загальний опис її архітектури



### *Рис. 1.2 Архітектура нейронної мережі VGG16*

Шари нейронної мережі складаються з набору різних згорткових фільтрів. Кожен з фільтрів відповідає за пошук певного візерунка, і коли він виявляє деяку область зображення, яка має цей візерунок, фільтр посилає сигнал наступному шару. У свою чергу, сигнали попереднього шару становлять нове зображення для наступного шару. На кресленні архітектури VGG16 видно, що спочатку було кольорове зображення RGB розміром  $224 \times 224$  пікселів з 3 каналами (червоний, зелений, синій). Після проходження першого шару згортки було отримано зображення розміром  $224 \times 224$  пікселів з 64 каналами. Ці канали більше не представляють кольори, а результати кожного з 64 фільтрів згортки. І так далі, до зображення  $7 \times 7$  пікселів з 512 каналами. Після стека згорткових шарів (мають різну глибину в різних архітектурах) є три повністю пов'язані шари: перші два мають по 4096 каналів, третій – 1000 каналів (за кількістю категорій 1000 категорій – один канал відповідає одному класу). Шляхом побудови каскадів згорткових шарів і навчання моделі формуються шари, які містять абстракції зображення. Перші шари можуть містити найдрібніші деталі: лінії. Далі йдуть комбінації деталей – фігур. Наступні шари вже міститять фігури, а в кінці – цілі об'єкти.

## РОЗДІЛ 2 КРОКИ ДЛЯ СТВОРЕННЯ ПРАКТИЧНОГО ЗАСТОСУНКУ

### 2.1 Алгоритм розв'язування

Як описано вище, для початку нам потрібні вхідні дані – це може бути будь-яке зображення. Далі робимо первинну обробку зображення, наприклад, змінюємо розмір зображення, виділяємо контури між переходом об'єктів, розмиття для зменшення зашумленості та зниження деталізації, переведення кольорової гами в чорно-білу чи інші необхідні маніпуляції з зображенням. Після зібраного датасету необхідно буде створити нейронну мережу, яка може приймати на вхід наші зображення і на них вчитись. Коли датасет і нейронна мережа готові починається процес тренування, який є часозатратний і вимагає потужних обчислювальних ресурсів. У результаті навчання буде отримано модель, на якій вже можна буде протестувати наскільки добре вона виконує свою функцію, тобто знаходить та розпізнає об'єкт на зображенні. Якщо модель добре натренована, то результати будуть отримуватись з високою точністю, у іншому випадку потрібно буде запустити процес навчання додатково декілька разів.

### 2.2 Датасет для комп'ютерного зору

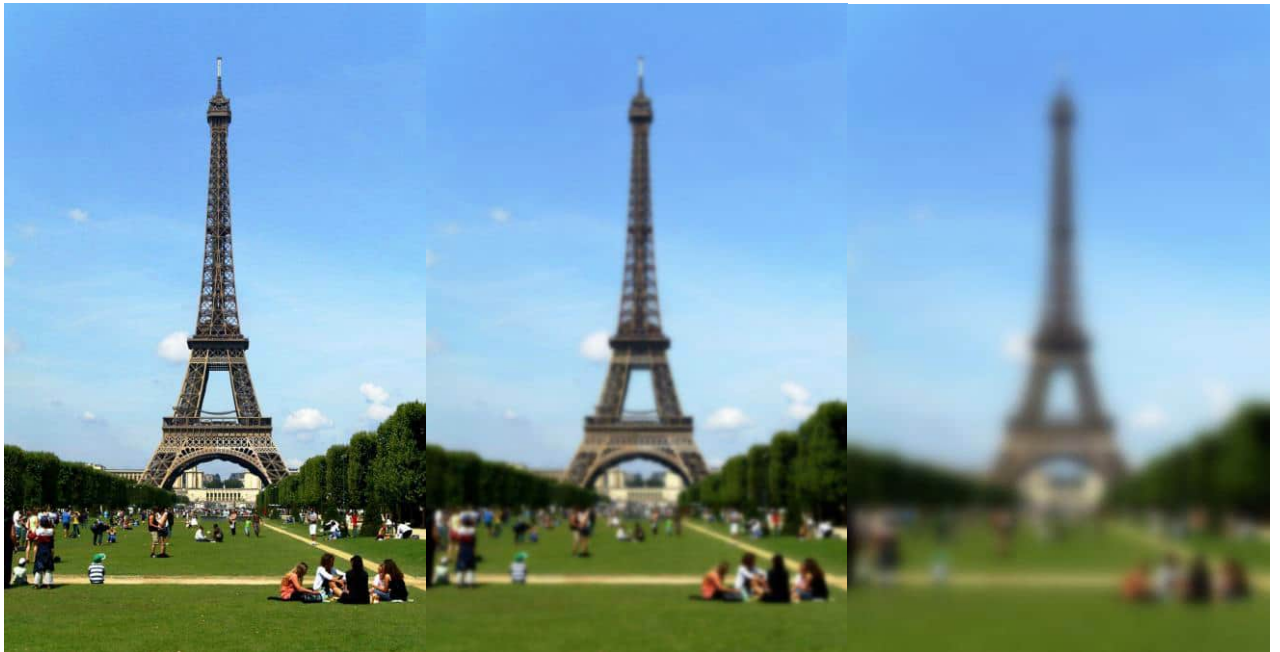
Датасет – це набір зображень, який за обсягом зазвичай дуже великий, і містить лише конкретний об'єкт. У Інтернеті можна знайти багато датасетів у відкритому доступі, як правило, вони будуть підготовлені для конкретної задачі. Наприклад, якщо це датасет формату PASCAL VOC, і він містить фотографію людини, то її тіло буде виділено в прямокутну рамку і добавлено спеціальний файл з інформацією про точки рамки. Таким чином, датасети такого формату можна використовувати для тренування нейронної мережі без підготовки власноруч.

Якщо ж готовий датасет не використовується, а створюється власний, то для цього потрібно буде скачати зображення з певною тематикою, мінімальна кількість для прийнятних результуючих даних – 500 екземплярів. Якщо такої кількості зображень певного об'єкту не існує, то використовують обробку зображень для того, щоб збільшити кількість фотографій, при цьому, не зменшуючи якість результату.

## 2.3 Практичні прийоми для обробки зображення

### 2.3.1 Розмивання Гауса

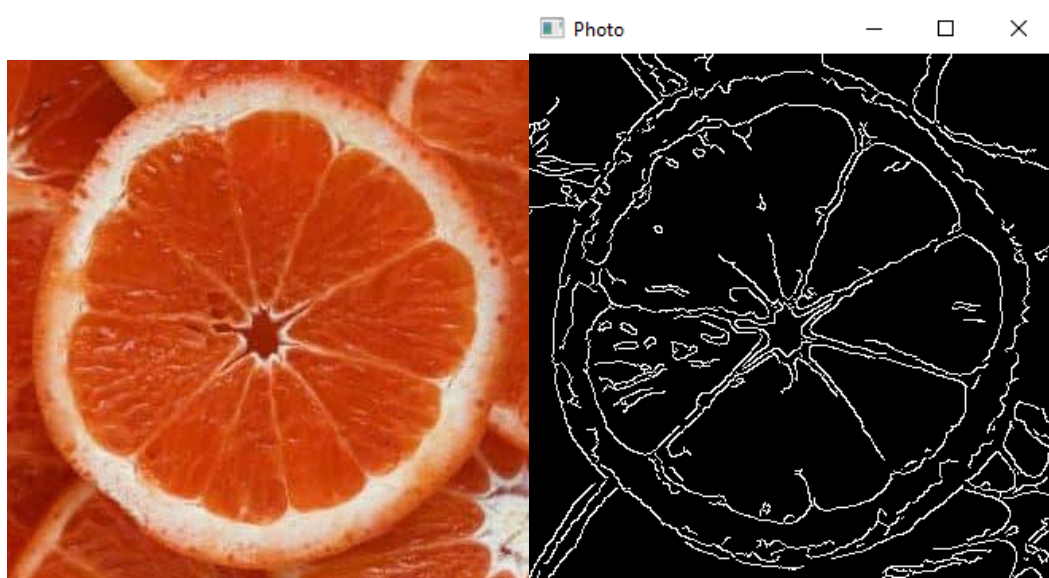
Найбільш простим і часто вживаним прикладом є розмивання Гауса. Якщо провести паралель до людської системи зору, розмивання уособлює собою короткозорість. Людина не може побачити чіткі риси, але має здатність виділяти певні критерії, такі як: хода, ріст, комплекція тіла, татуювання, окуляри, колір волосся. Результат розпізнавання зменшується, бо на відміну від унікальних рис обличчя інші критерії можуть пересікатись у деяких людей, але якщо взяти узагальнений випадок, то ймовірність правильно ідентифікувати людину залишається високою.



*Рис. 2.1 Зображення розмите по Гаусу з різним радіусом*

### 2.3.2 Виділення контурів об'єктів

Виділення контурів об'єктів - відбувається шляхом порівняння сусідніх пікселів за яскравістю. Чим більша різниця, тим більший шанс, що алгоритм порівняння виділить контуром. У кожного об'єкта є схожий набір контурів, проте не можна спиратись лише на цей критерій, бо, наприклад, апельсин і мяч мають однакові контури, але це абсолютно різні об'єкти, які крім своєї форми не мають нічого спільного.



*Рис. 2.2 Зображення з виділеними контурами об'єкту*

### 2.3.3 Опорні точки

На зображенні вибирається певна множина точок (наприклад, самі контрастні кольори на зображенні). Далі обчислюються взаємне розташування опорних точок як характеристики картинки. В результаті складається рівняння, що описує розташування точок. Алгоритм передбачає порівняння результатів всіх наявних картинок і знаходить схожу. Ця група методів досить хороша для виявлення деформованих копій зображення, однак для визначення подібності двох принципово різних, нехай і візуально схожих картинок, метод непридатний (наприклад, дві фотографії одного і того ж котика в різних позах матимуть мало спільного для таких методів). Тому, такий підхід може лише визначати, чи є в базі модифікації шуканого зображення, але не може знаходити схожі в будь-якому іншому сенсі.

### 2.4 Тренування нейронної мережі

Коли зібрано досить великий датасет, проведено обробку і аналіз зображень, то можна переходити до безпосереднього навчання мережі. Потрібно розділити дані на тренувальні та тестувальні. Доволі часто ділять на 80 відсотків для тренування і 20 відсотків для тестування. Процес навчання дуже ресурсозатратний як і в часовому плані, так і в технічному. Для того, щоб пришвидшити навчання використовуються відеокарти замість процесора. У результаті навчання буде отримано так звану модель, тобто файл який містить дані для розпізнавання

об'єктів. Модель використовується для того, щоб лише один раз запустити процес навчання нейронної мережі, а після цього використовувати її результат для виявлення. Якщо процес навчання розбивається на довгий термін з паузами, то потрібно просто “запам'ятати” місце в якому закінчилось тренування і на наступний раз не починати з самого початку, а просто дописувати результати до моделі з місця попередньої зупинки.

## 2.5 Існуючі моделі систем штучного інтелекту для розпізнавання об'єктів на зображенні

### 2.5.1 ResNet

ResNet (Residual Network) — це архітектура глибокої згорткової нейронної мережі, яка була представлена в 2015 році дослідниками Microsoft. ResNet використовує залишкові блоки, щоб дозволити мережі вивчати складніші функції. Він досяг найсучаснішої продуктивності на наборі даних ImageNet, який зазвичай використовується для порівняння моделей розпізнавання зображень.

### 2.5.2 Inception

Inception — ще одна архітектура глибокої згорткової нейронної мережі, розроблена дослідниками компанії Google у 2014 році. Вона використовує кілька згорткових шарів із різними розмірами ядра для захоплення функцій у різних масштабах. Inception також досягла найсучаснішої продуктивності на наборі даних ImageNet.

### 2.5.3 Faster R-CNN

Faster R-CNN — це двоетапна модель виявлення об'єктів, яку представили дослідники Microsoft у 2015 році. Вона використовує мережу регіональних пропозицій (RPN) для створення пропозицій об'єктів, а потім використовує CNN для класифікації та вдосконалення пропозицій. Faster R-CNN досяг найвищої продуктивності в кількох тестах виявлення об'єктів, включаючи PASCAL VOC і MS COCO.

Варто зазначити, що існує багато інших чудових моделей для розпізнавання об'єктів на зображеннях, і вибір моделі залежить від таких факторів, як конкретне завдання, розмір набору даних і доступні обчислювальні ресурси.

## 2.6 Покращення для розпізнавання вхідного відео файлу

Відео файл та зображення відрізняються за декількома основними аспектами:

1. **Формат даних:** Відео файл - це послідовність кадрів, збережених у вигляді послідовності зображень, що відтворюють рух, звук і можуть містити інші дані (наприклад, аудіо доріжку, субтитри і т. д.). Зображення - це одна статична картинка, збережена у певному форматі, такому як JPEG, PNG, BMP тощо.
2. **Рух:** Відео може містити рухомі зображення, тоді як зображення є статичним і не містить руху. Відео може бути відтворене в реальному часі зі змінними кадрами, тоді як зображення відображає один момент часу.
3. **Розмір файлу:** Відео файл зазвичай має більший розмір порівняно зі зображенням, оскільки він містить більше даних через послідовність кадрів. Зображення, як правило, має менший розмір файлу, оскільки воно містить лише одну статичну картинку.
4. **Часовий аспект:** Відео може мати тривалість та відтворюватися впродовж певного часового проміжку, тоді як зображення не має тривалості і відображається одразу.
5. **Застосування:** Відео використовується для передачі рухомих зображень, відео кліпів, фільмів, телепередач, відеоігор та інших мультимедійних контентів. Зображення використовується для фотографій, малюнків, ілюстрацій, логотипів та інших статичних зображень.

Отже, відео та зображення відрізняються за форматом даних, наявністю руху, розміром файлу, часовим аспектом та застосуванням, в яких вони використовуються. Однак в розрізі розпізнавання об'єктів на зображенні та відео процес має багато спільних кроків. Головною відмінністю все ж таки залишається те, що відео це набір декількох зображень в секунду. Для того, щоб розпізнати об'єкти на відео потрібно:

- Уміти розпізнавати об'єкти на зображенні
- Видалити аудіодоріжку з відеофайлу
- Розділити відеопослідовність на кадри
- Запустити виконання пошуку об'єктів на зображенні для кожного кадру, або для певної послідовності кадрів, наприклад, кожного третього для оптимізації часу роботи програми
- З'єднати зображення назад у послідовність, яка у результаті створить відео
- За бажанням, приєднати аудіодоріжку знову

## 2.7 Покращення для розпізнання відео в режимі реального часу

Припустимо, що у нас уже є готова програма для того, щоб розпізнавати вхідний відеофайл. Що потрібно змінити для того, щоб мати можливість працювати з файлом в режимі реального часу. Перш за все, потрібно володіти потужними обчислювальними ресурсами. У випадку вхідного зображення чи відео час знаходження та розпізнання об'єктів не мало ключової ролі, на відміну від вхідних файлів в режимі реального часу, так як в даному випадку суттєва затримка може призвести до того, що результат буде надходити коли у ньому нема сенсу і актуальності. Очевидним прикладом є система автопілота, камера відеоспостереження у розумних будинках чи ракети самонаведення. У випадку автопілота – необхідно якнайшвидше розпізнати перешкоду, для камери відеоспостереження виявити посторонніх осіб в приміщенні, щодо ракети, то розпізнати ціль швидше ніж ракета долетить та здетонує. Однак, глобально програма нічим не відрізняється від тої, яка розпізнає об'єкти без режиму реального часу. Аналогічно потрібно отримати кадр, розпізнати об'єкт, повернути дані з інформацією про файл.

## 2.8 Створення загальнодоступного ресурсу

Створення програми для загального користування, яка може бути використана широким колом користувачів, має кілька важливих переваг порівняно з локальною програмою, яка призначена лише для використання на одному комп'ютері або в обмеженому контексті:



1. **Доступність:** Програма для загального користування може бути доступна багатьом користувачам з різних місць і на різних пристроях, таких як комп'ютери, планшети, мобільні телефони тощо. Це дає можливість використовувати програму в будь-якому зручному для користувача місці та часі.
2. **Розширені можливості:** Програма для загального користування може мати більше можливостей та функцій, оскільки вона призначена для використання різними користувачами з різними потребами. Це може включати можливості збереження та обробки даних в хмарі, взаємодії з іншими користувачами, можливості налаштування та розширення функціональності.
3. **Оновлення та підтримка:** Програма для загального користування може бути оновлювана та підтримувана розробниками, що дозволяє виправляти помилки, вдосконалювати функціональність, додавати нові можливості та забезпечувати безпеку програми. Користувачі можуть отримувати оновлення автоматично та без зусиль з свого боку.
4. **Масштабованість:** Програма для загального користування може бути розроблена з врахуванням можливості масштабування, що дозволяє забезпечувати використання програми більшим числом користувачів без втрати продуктивності та функціональності.
5. **Зворотній зв'язок:** Зворотній зв'язок від користувачів може допомогти виявити слабкі місця або помилки в програмі, а також зрозуміти потреби та очікування користувачів. Це може дозволити розробникам вдосконалювати продукт, вносити в нього зміни, що поліпшують його функціональність та відповідають вимогам ринку, що в свою чергу може призвести до більшої популярності та використання програми. Також, це дозволяє користувачам висловлювати свої думки, враження та відгуки щодо програми. Це допомагає розробникам забезпечити високий рівень задоволеності користувачів, що може призвести до вірної клієнтської бази, рекомендацій від задоволених користувачів, та підвищення лояльності до продукту.

Для цього потрібно створити інсталятор або використати хмарні технології для забезпечення можливості користувачам взаємодіяти із застосунком.

## РОЗДІЛ 3 ПРИКЛАДИ ПРАКТИЧНОЇ РЕАЛІЗАЦІЇ

### 3.1 Приклад з використанням зображень оброблених локально

#### 3.1.1 Приклад №1

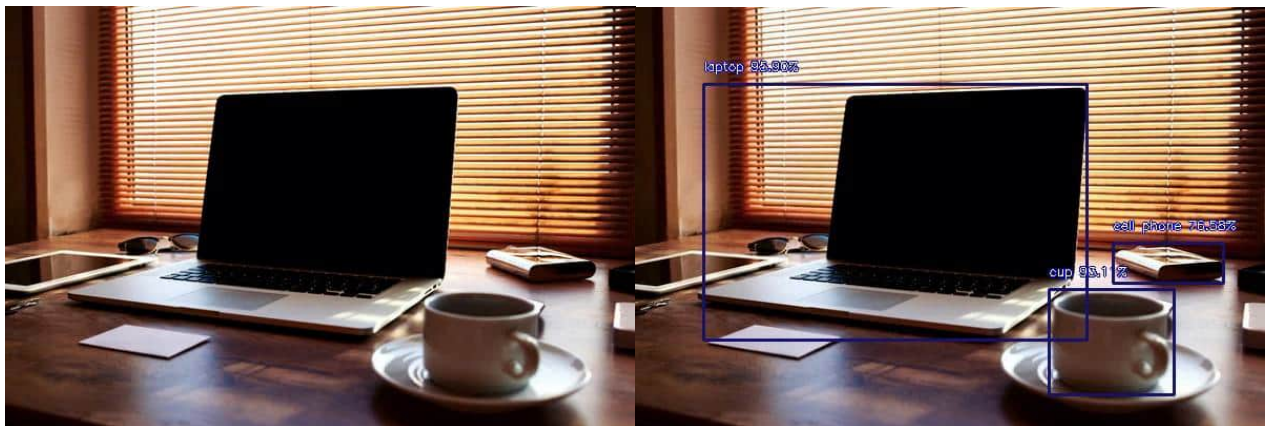


Рис. 3.1 Приклад №1

На даному прикладі можемо спостерігати, що ноутбук та чашку мережа розпізнає з точністю більше ніж 93%, але, наприклад, мережа вважає, що фляга для води – це мобільний телефон.

#### 3.1.2 Приклад №2

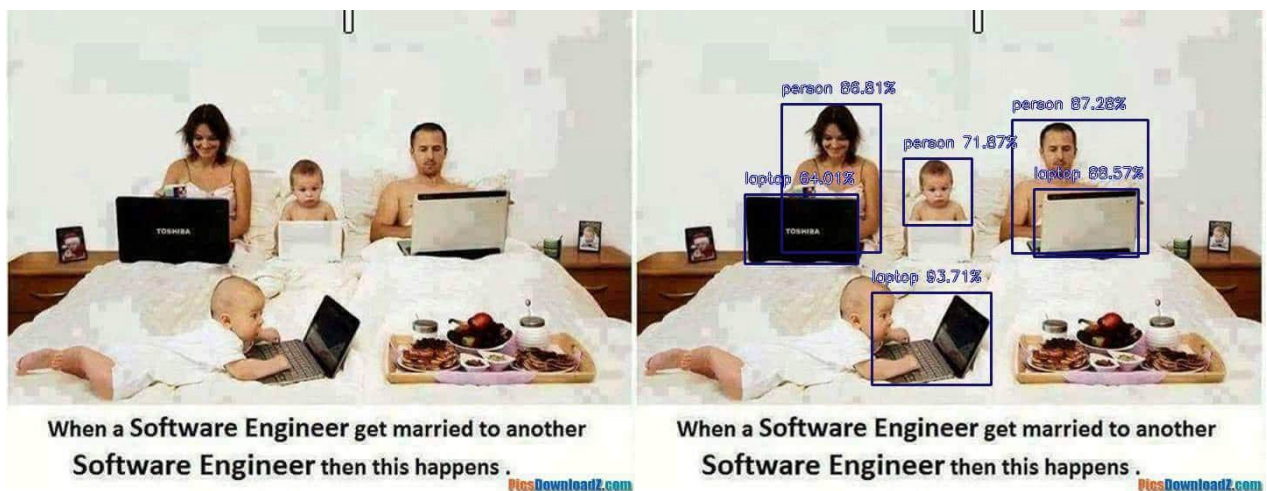
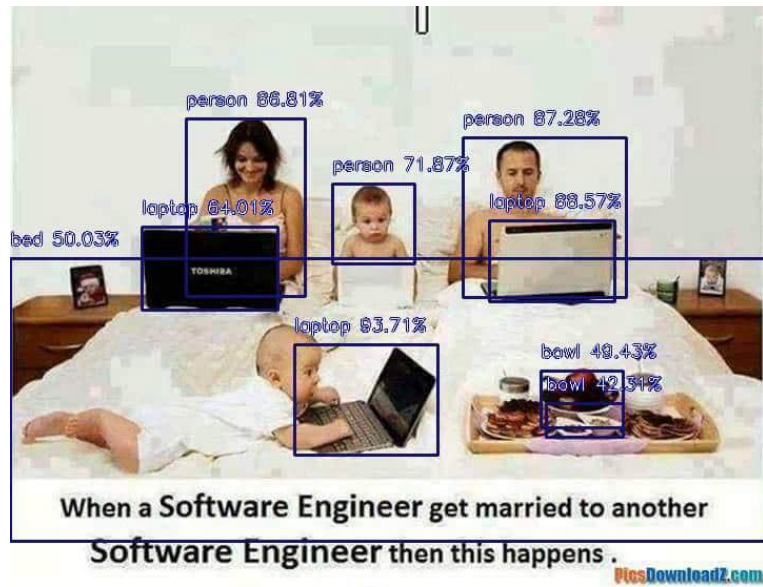


Рис. 3.2 Приклад №2 з порогом прийняття 0.6

На даному прикладі, поріг прийняття дорівнює 0.6, тобто якщо мережа видає передбачення з нижчою точністю ніж 60%, то об'єкт не буде враховано до результуючого зображення. У цьому випадку розпізнає всі ноутбуки і 75% людей. Для того, щоб виділити більше об'єктів зменшимо поріг прийняття до 0.4.



*Рис. 3.3 Приклад №2 з порогом прийняття 0.4*

Додатково розпізнало ліжко і дві тарілки.

При великій кількості об'єктів стає складно спостерігати результат, бо рамки накладаються одна на одну. Удосконалимо нашу програму, щоб для кожного об'єкту утворювалось нове зображення і створювало файл лише з цим об'єктом.



*Рис. 3.4 Фото ноутбуків з прикладу №2*



*Рис. 3.5 Фото людей з прикладу №2*



*Рис. 3.6 Фото тарілок з прикладу №2*



*Рис. 3.7 Фото ліжка з прикладу №2*

### 3.1.3 Приклад №3

Також, можна налаштувати на конкретну вибірку об'єктів. Це збільшить швидкодію програми і зменшить кількість об'єктів для подальшого аналізу результатів. Наприклад, ми наперед знаємо, що з вхідних даних, нам потрібно виділити лише м'ячі. Налаштовуємо програму так, щоб брало до уваги, лише м'ячі.



*Рис. 3.8 Вхідне зображення прикладу №3*



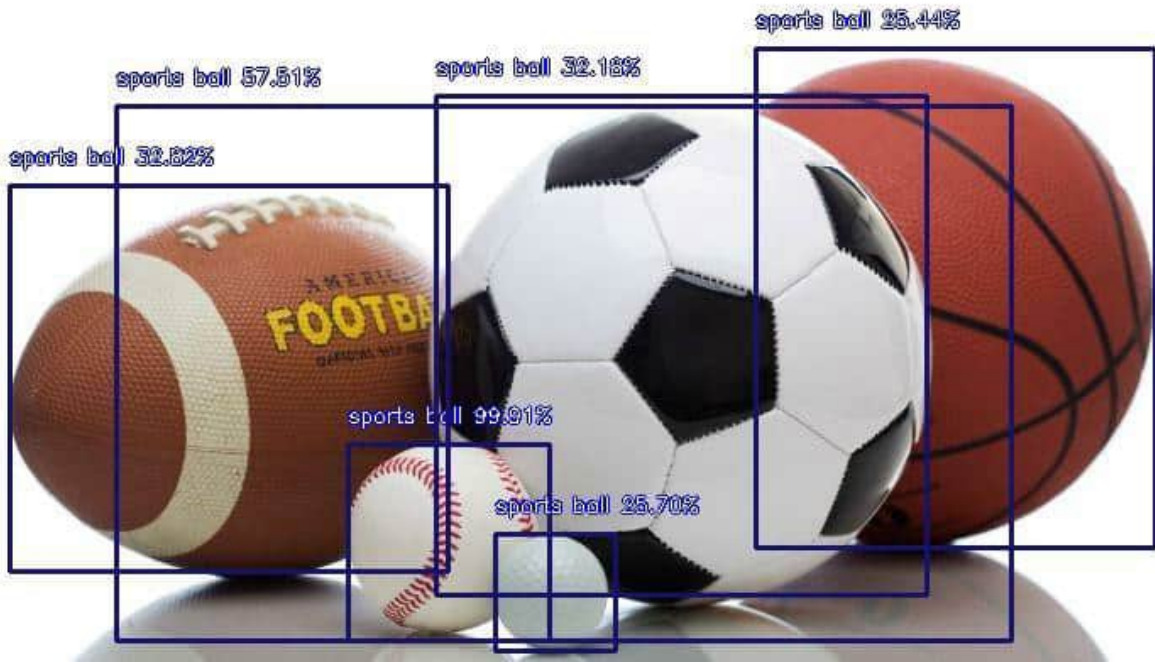
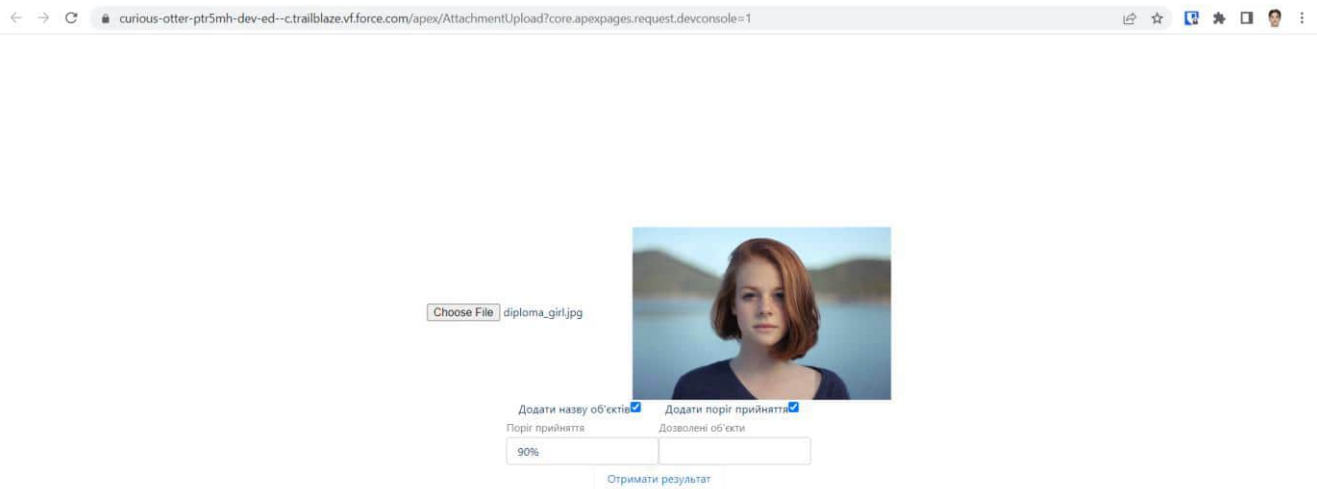
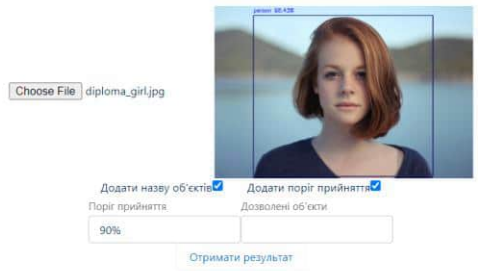


Рис. 3.9 Вихідне зображення прикладу №3

### 3.2 Приклад з використанням зображень оброблених на загальнодоступному ресурсі

#### 3.2.1 Приклад №1:





Результат отримано відповідно до вхідних параметрів, а саме: виведено назву об'єкту, виведено поріг прийняття, а також вказано, що отриманий результат повинен мати поріг прийняття не менше ніж 90%. У цьому випадку нейронна мережа розпінала людину на зображенні з впевненістю у 98.43%.



## **ВИСНОВКИ**

У своїй роботі я детально описав як відбувається розпізнавання об'єктів на зображенні та створив програму для практичного застосування. Розпізнавання об'єктів на зображенні може бути використано в багатьох напрямках буденного життя. Наприклад, розпізнавання перешкод на дорозі, для запобігання порушень правил паркування, для обходження аутентифікації на сайті, де людина повинна вибрати певний тип зображень, для того, щоб система перевірила чи справді людині надає доступ до сайту.

Оскільки даний напрямок стрімко зростає за останні роки і з розвитком обчислювальних ресурсів можна досягти точнішого навчання моделей, продуктивності, і меншого часу на виявлення. Висока ймовірність визначення типів об'єктів досягається за рахунок ефективних моделей згорткових нейромереж. На точність розпізнавання типів об'єктів не впливають погодні умови, зміна часу доби, освітленість тощо. Технології навчання нейромереж дозволяють створювати комплекси для ідентифікації будь-яких типів об'єктів.



## ДОДАТОК

Для реалізації практичного прикладу я використовую:

- Python - це високорівнева інтерпретована мова програмування загального призначення, яка була створена в 1991 році Гвідо ван Россумом. Python має простий і зрозумілий синтаксис, що робить його дуже легким для вивчення і використання. Ця мова програмування може бути використана для розробки різноманітних програм, включаючи веб-сайти, наукові обчислення, штучний інтелект, графічні ігри та інші проекти. Python є вільним програмним забезпеченням і може бути використаний на різних платформах, таких як Windows, Mac та Linux. Він також має велику кількість сторонніх бібліотек, що роблять його ще більш потужним і універсальним.

- TensorFlow - це відкрите програмне забезпечення для машинного навчання, розроблене Google. Воно дозволяє розробляти, навчати і застосовувати моделі машинного навчання, включаючи нейронні мережі, для розв'язання різноманітних завдань. TensorFlow використовує високорівневе API, що дозволяє розробникам легко створювати та навчати моделі з високою точністю.

TensorFlow має велику популярність у галузі машинного навчання і використовується для різноманітних завдань, таких як розпізнавання образів, обробка мови, аналіз даних і багато іншого. Він має багато готових моделей, що дає можливість використовувати їх у власних проектах, а також дозволяє тренувати моделі на великих обсягах даних. TensorFlow підтримує роботу з багатьма мовами програмування, включаючи Python, C++, Java та інші, що дозволяє розробникам використовувати ті мови, з якими вони найкраще знайомі.

- ImageAI - це бібліотека Python, яка надає прості і потужні інструменти для розробки систем комп'ютерного зору. Вона дозволяє розробникам легко створювати програми, що здатні розпізнавати образи, виявляти об'єкти, класифікувати зображення, виконувати відеоаналіз та багато іншого.

ImageAI базується на Tensorflow, Keras і OpenCV, тому вона дозволяє легко і швидко створювати та навчати моделі машинного навчання для розв'язання різних задач комп'ютерного зору. ImageAI має простий та зрозумілий інтерфейс, який

дозволяє розробникам швидко створювати програми комп'ютерного зору без необхідності великого досвіду у машинному навчанні та розробці програм.

ImageAI може бути використаний для розв'язання різноманітних задач комп'ютерного зору, таких як детектування об'єктів, класифікація зображень, виявлення облич, аналіз відео та іншого. Вона також підтримує обробку зображень в реальному часі, що дозволяє використовувати її для розробки програм, що потребують високої швидкодії та точності.

- RetinaNet - це алгоритм об'єктного впізнавання (object detection) в машинному навчанні, який був запропонований в 2017 році командою дослідників з компанії Facebook AI Research (FAIR). Він є одним з найпопулярніших алгоритмів для розв'язання задачі об'єктного впізнавання, яка полягає в виявленні та класифікації об'єктів на зображеннях. RetinaNet вирішує проблему "класифікації з обмеженими даними" в задачі об'єктного впізнавання, де деякі класи об'єктів можуть зустрічатись дуже рідко, тобто мати невелику кількість тренувальних прикладів. У основі RetinaNet лежить архітектура з декількома шарами, яка використовує дві основні складові: Retina (Focal) Loss і Feature Pyramid Network (FPN). Retina (Focal) Loss - це спеціальна функція втрат, яка зводить до мінімуму вплив легко класифікованих об'єктів на процес навчання. Вона дозволяє зосередитися на важких прикладах, таких як малі об'єкти або об'єкти з великим відношенням розмірів сторін. Feature Pyramid Network (FPN) - це мережа, яка дозволяє використовувати зображення різних масштабів для виявлення об'єктів різних розмірів. Вона допомагає забезпечити точне виявлення як малих, так і великих об'єктів на зображеннях. RetinaNet є ефективним алгоритмом об'єктного впізнавання, який дозволяє досягати високої точності виявлення об'єктів на зображеннях, навіть в умовах обмежених даних. Він знайшов широке застосування в ряді застосувань, таких як автоматичне оброблення відео, автономні автомобілі, медичне зображення, відеоспостереження, розпізнавання об'єктів на зображеннях з дронів, робототехніка та багато інших сфер. RetinaNet може бути використаний для виявлення різних класів об'єктів, таких як люди, автомобілі, тварини, об'єкти

різних форм та розмірів, що робить його дуже універсальним та потужним рішенням в галузі об'єктного впізнавання.

- Ngrok – це веб-сервіс тунелювання, який дозволяє встановлювати безпечний зовнішній доступ до локальних веб-серверів та локальних розробок. Він працює як проксі-сервер між локальним сервером та Інтернетом, дозволяючи зовнішнім користувачам отримувати доступ до локального веб-додатку через Інтернет. Ngrok забезпечує публічний URL-адресу, яка відображається на зовнішньому Інтернеті та перенаправляє запити на локальний сервер, де заведений додаток. Це дозволяє легко демонструвати, тестувати та спілкуватися з веб-додатками, які виконуються локально на комп'ютері розробника або в локальній мережі. Деякі можливості Ngrok включають підтримку HTTP і HTTPS тунелювання, можливість налаштовувати піддомени, доступ до деталей запитів та відповідей, можливість використовувати автентифікацію для захисту доступу та підтримку різних протоколів, таких як TCP і UDP. Ngrok є корисним інструментом для розробників та тестування веб-додатків, особливо тих, які виконуються локально, і вимагають зовнішнього доступу. Він може бути використаний в ряді сценаріїв, таких як веб-розробка, веб-демонстрації, тестування веб-сервісів та багато іншого.

- Salesforce – це провідна у світі CRM (Customer Relationship Management) платформа, яка дозволяє компаніям керувати відносинами з клієнтами, автоматизувати продажі, маркетинг, обслуговування клієнтів та аналізувати дані для прийняття рішень. Salesforce надає широкий спектр хмарних рішень, включаючи різні модулі, такі як Sales Cloud, Service Cloud, Marketing Cloud, Commerce Cloud, Community Cloud, Analytics Cloud, та інші, які взаємодіють між собою, щоб надати повний спектр функціональностей для ефективного керування взаємодією з клієнтами та розвитку бізнесу. Salesforce надає інструменти для ведення бази клієнтів, відстеження продажів, управління маркетинговими кампаніями, надання обслуговування клієнтів, аналітики даних, автоматизації бізнес-процесів та багато інших функціональностей, які допомагають компаніям покращувати взаємодію з клієнтами, збільшувати ефективність роботи та досягати кращих бізнес-результатів. Salesforce використовується різними компаніями,

незалежно від розміру та галузі, включаючи продажі, маркетинг, обслуговування клієнтів, фінанси, медицину, технології та багато інших.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

- [1] Нікольський Ю.В. Системи штучного інтелекту. Навчальний посібник / Ю.В. Нікольський, В.В. Пасічник, Ю.М. Щербина – Львів, видавництво «Магнолія – 2006», 2013.
- [2] Ясницкий Л.Н. Введение в искусственный интеллект: учеб. пособие для студ. высш. учеб. заведений / Л.Н. Ясницкий – 2-е изд., испр. Издательский центр “Академия”, 2008.
- [3] Трушевський В.М. Метод скінченних елементів і штучні нейронні мережі: теоретичні аспекти та застосування / В.М. Трушевський, Г.А. Шинкаренко, Н.М. Щербина. – Львів: ЛНУ імені Івана Франка, 2014. – 396 с.
- [4] [Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence](#)
- [5] Річард Селіскі Комп'ютерний зір: алгоритми та застосування, 2010.
- [6] Ян Ерік Солем Програмування комп'ютерного зору за допомогою Python: інструменти та алгоритми для аналізу зображень автора, 2012. – 260 с.
- [7] Бенджамін Планш Практичний комп'ютерний зір із TensorFlow 2: використовуйте глибоке навчання для створення потужних програм для обробки зображень із TensorFlow 2.0 і Keras / Еліот Андрес – 2019. – 479 с.
- [8] Девід А. Форсайт Комп'ютерний зір: сучасний підхід / Жан Понсе – 2-е видання – 2011. – 790 с.