

Проектування та
реалізація методів
взаємодії на основі жестів
у мультимедійних
програмах, таких як
безконтактні інтерфейси,
для підвищення взаємодії
та контролю користувачів

Доповідач: Черніков А.А.

Керівник: Афонін А.О.





Мета роботи

- Розширення можливостей взаємодії з комп'ютером.
 - Розроблення інтерфейсу розпізнавання жестів для збільшення контролю користувачів над переглядом відеоконтенту.
-



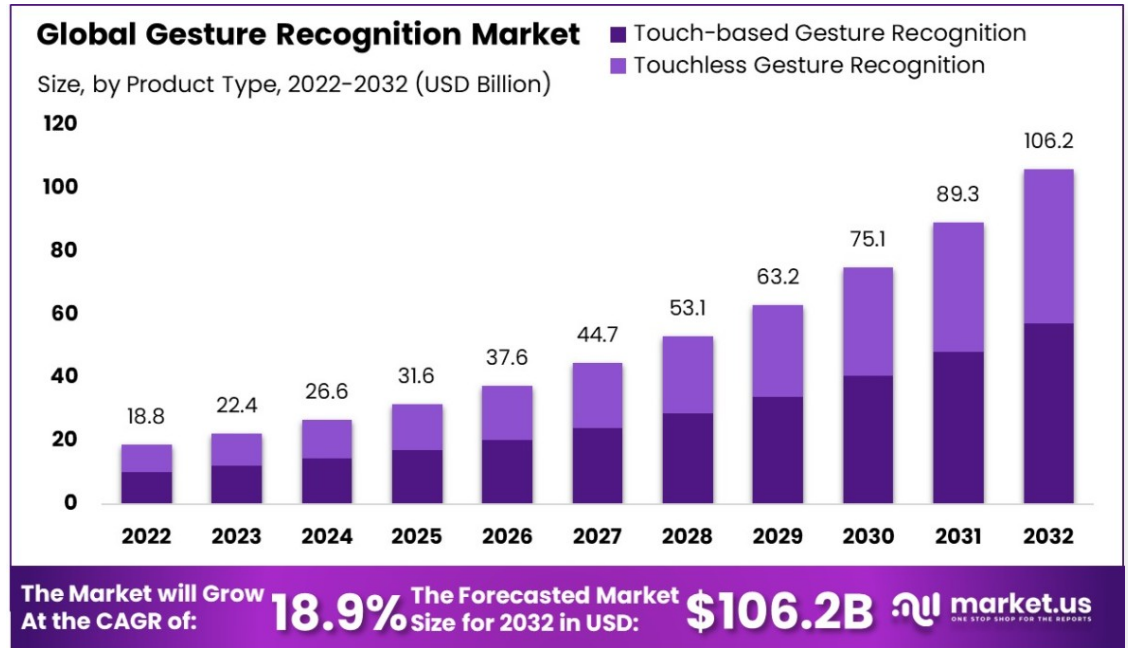
Завдання роботи

Розроблення методів жестової взаємодії між користувачем та комп'ютером, з використанням відеокамери для зчитування та виконання безконтактних команд.



Актуальність теми

- Щорічне збільшення розвитку ринку технології розпізнавання жестів.
- Зацікавленість провідних компаній у розробленні продуктів, які передбачають використання жестів.
- Визначення майбутніх трендів взаємодії людини із комп'ютером.





Огляд технології

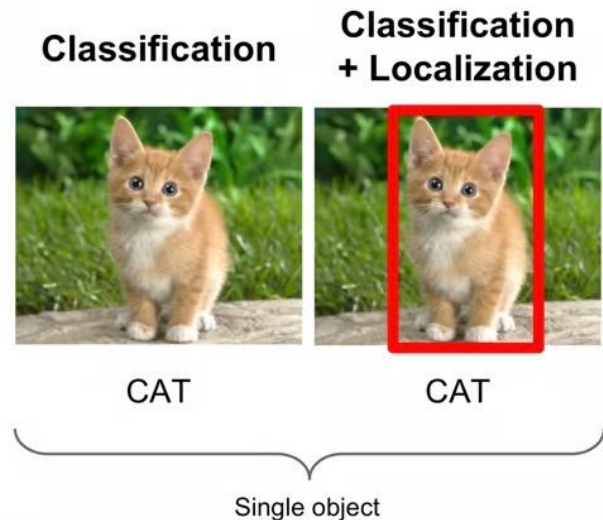
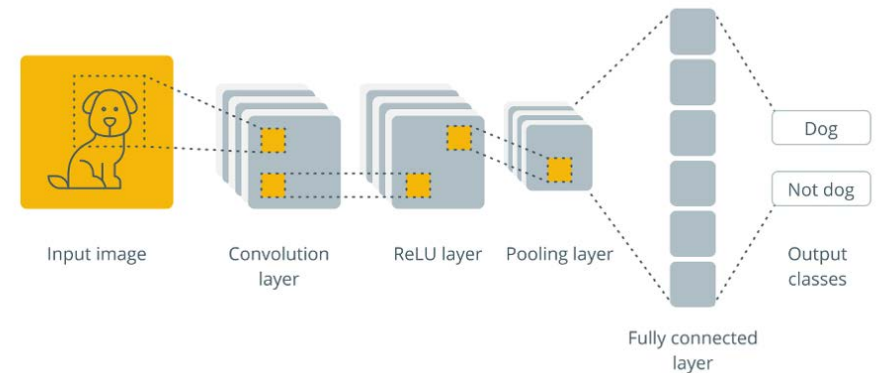
- Розпізнавання жестів – підгалузь комп'ютерного зору.
 - Дозволяє комп'ютерам «бачити» об'єкти.
 - Використання нейронних мереж (моделей) та пристрою зчитування інформації (відеокамери).
 - Жестова взаємодія – безконтактний спосіб комунікації людини з комп'ютером.
 - Інтерпретація жестів для виконання команд.
-

Етапи розпізнавання жестів

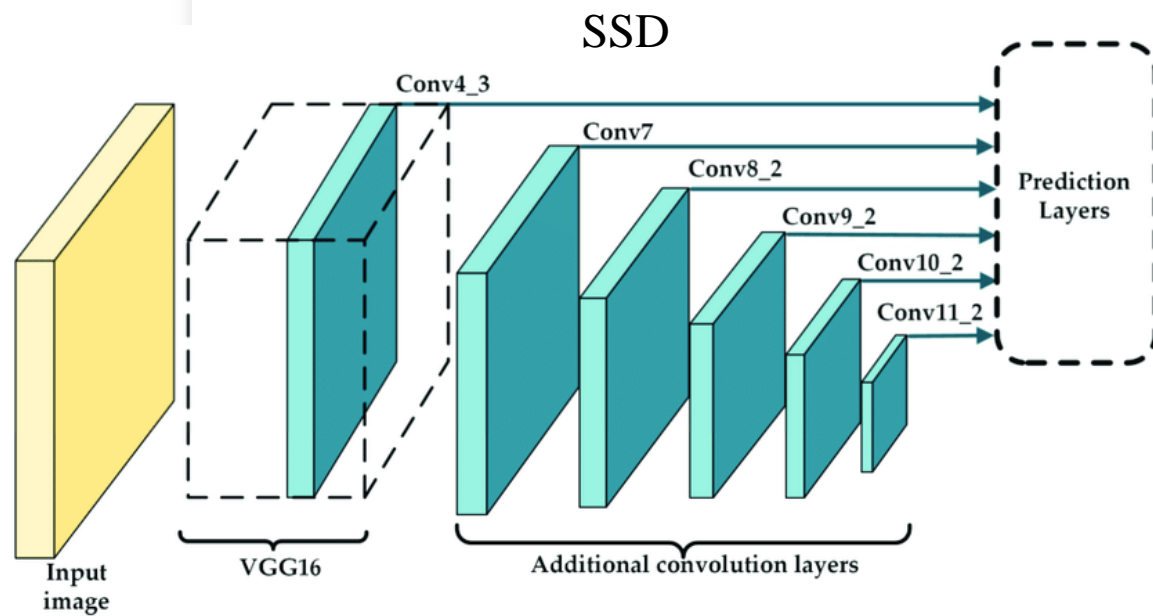
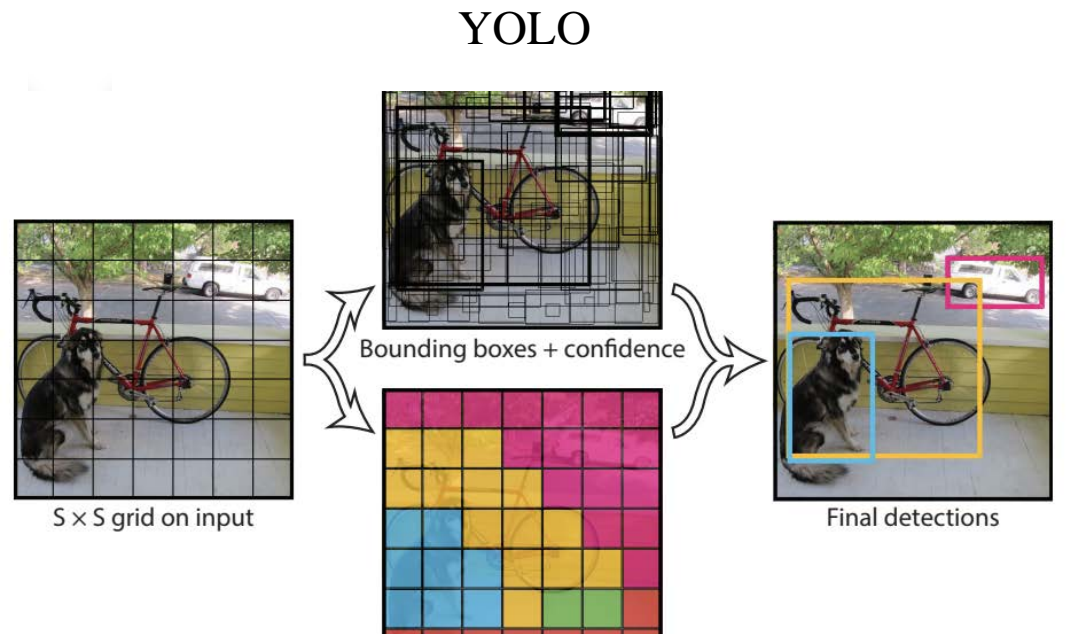
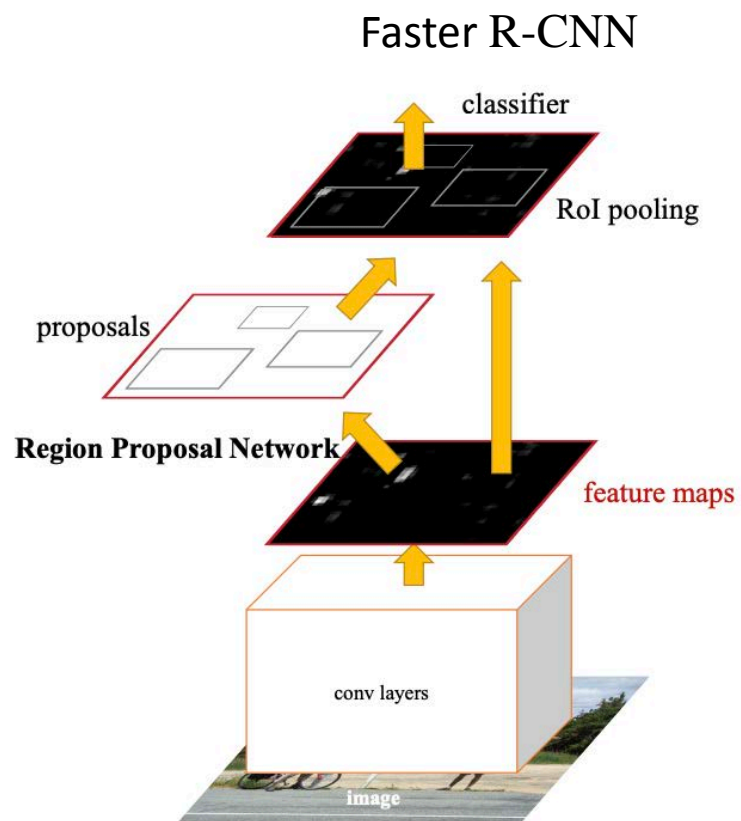
- Класифікація зображення.
 - Визначення положення об'єкту (руки).
 - Розпізнавання представленого жесту.
 - Використання ключових точок для збільшення гнучкості моделі.
 - Перехід до тривимірних ключових точок.
-

Класифікація зображення

- Перший етап для визначення жесту.
- Використання згорткових нейронних мереж (Convolutional Neural Network) у якості класифікатора.
- Згортання зображення до мінімальної кількості значущих елементів.
- Використання шарів, як-от: згортковий (convolutional), об'єднуючий (pooling) та повністю зв'язуючий (fully connected)
- Дає змогу отримати відповідь на питання чи є об'єкт на зображенні, без локалізації конкретного місця.



Визначення положення об'єкта



Визначення положення об'єкта (порівняльна характеристика)

Назва	Принцип роботи	Недоліки	Переваги	Сценарії використання
Faster R-CNN (Region-based CNN)	Поділ зображення на регіони	Потребує багато апаратних ресурсів	Висока точність та універсальність використання	Використовується в системах, де точність важливіша за швидкість
YOLO (You Only Look Once)	Поділ зображення на сітку	Проблеми з виявленням маленьких об'єктів; Вимагає значного набору даних.	Висока точність виявлення об'єктів в реальному часі	Використовується в системах з обробкою в реальному часі
SSD (Single-shot object detection)	Набір обмежувальних рамок і прогнозів класів	Має труднощі з виявленням дрібних деталей; Потребує великого датасету	Виявлення об'єктів в реальному часі; Ідентифікація об'єктів різного розміру	Використовується в системах, де потрібен баланс між швидкістю та точністю

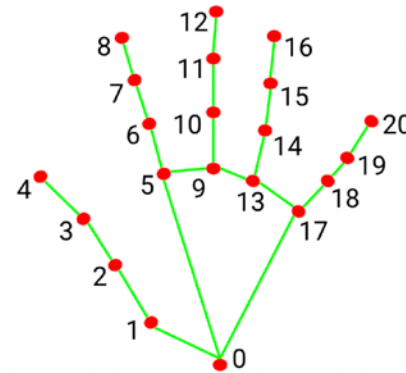
Розпізнавання представленого жесту

- Наступний крок після класифікації та локалізації.
- На цьому етапі комп'ютер ще не розуміє, що рука може перебувати в різних положеннях.
- Приклад тренування моделі на грі «камінь, ножиці, папір».
- Має бути наданий великий набір зображень.
- Підхід є не гнучким оскільки за зміни правил гри – потрібно проводити повторне тренування.



Використання ключових точок для збільшення гнучкості моделі

- Уникнути неоптимального підходу можна через використання ключових точок.
- Рука представляється не як єдиний об'єкт, а як набір елементів.
- Складається з 21 точки (4 для кожного пальця та 1 для зап'ястя).
- Підхід дає змогу проектувати більш гнучку систему, із ширшою бібліотекою жестів.

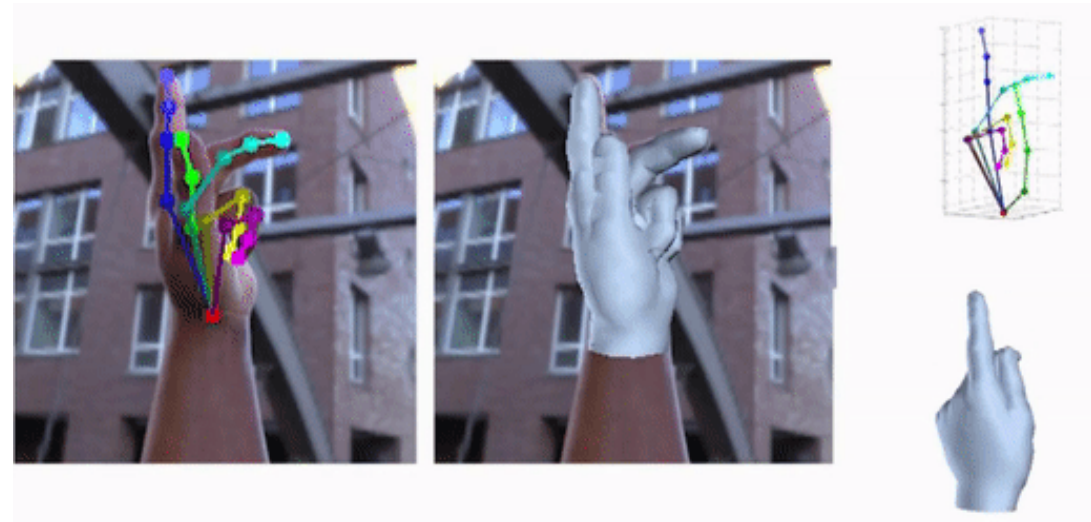


0. WRIST
 1. THUMB_CMC
 2. THUMB_MCP
 3. THUMB_IP
 4. THUMB_TIP
 5. INDEX_FINGER_MCP
 6. INDEX_FINGER_PIP
 7. INDEX_FINGER_DIP
 8. INDEX_FINGER_TIP
 9. MIDDLE_FINGER_MCP
 10. MIDDLE_FINGER_PIP

11. MIDDLE_FINGER_DIP
 12. MIDDLE_FINGER_TIP
 13. RING_FINGER_MCP
 14. RING_FINGER_PIP
 15. RING_FINGER_DIP
 16. RING_FINGER_TIP
 17. PINKY_MCP
 18. PINKY_PIP
 19. PINKY_DIP
 20. PINKY_TIP

Перехід до тривимірних ключових точок

- Останній етап розвитку технології розпізнавання жестів.
- Додавання параметру глибини.
- Дає змогу збільшити гнучкість системи.
- Можна проектувати більш складні рухи.



Додаток для Google Chrome

- Вирішено використати готове рішення відео плеєра.
 - Через брак доступу до змінних YouTube напямую, прийнято рішення відтворювати взаємодію через DOM (Document Object Model) вебсторінки.
 - Розроблену програму інтегровано в розширення Google Chrome.
-



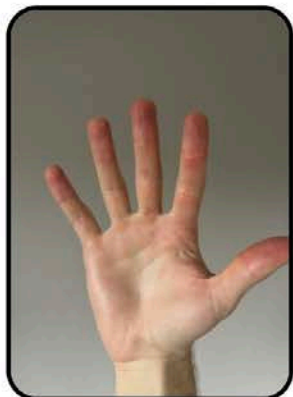
Tensorflow handpose

- Визначення положення руки за запуском програми досягається з допомогою використання моделі handpose.
 - Належить до бібліотеки Tensorflow.
 - Платформа надає готову нейронну мережу.
 - Керування процесом за схемою з ключовими елементами на руці.
-

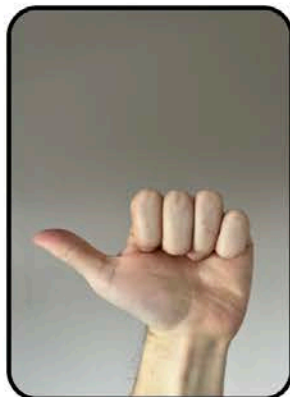
Проектування методів для взаємодії з YouTube плеєром на основі жестів

- Функції для обробки ідентифікованих пальців.
- Функції для доступу до контрольної панелі YouTube.
- Зміна значень вебсторінки.
- Використання динамічних порогових значень під час роботи з ключовими точками.
- Система підлаштовується під користувача.

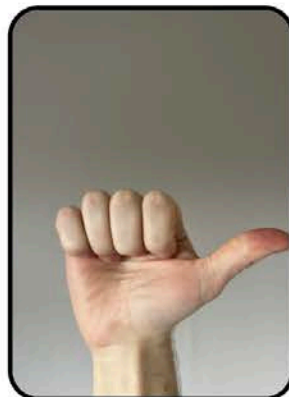
Перелік доступних команд



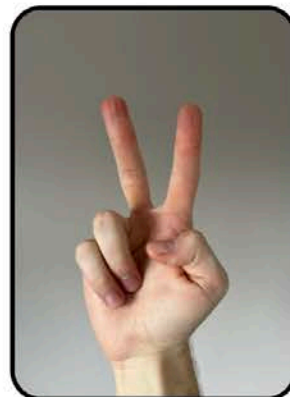
Play button



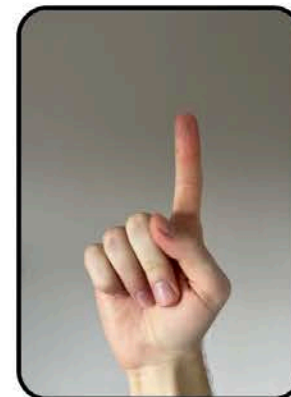
Rewind left



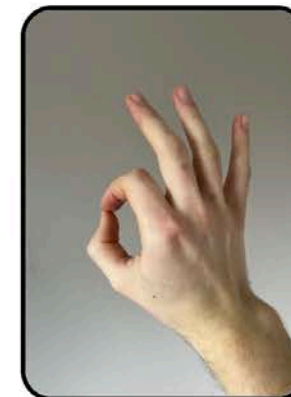
Rewind right



Volume up

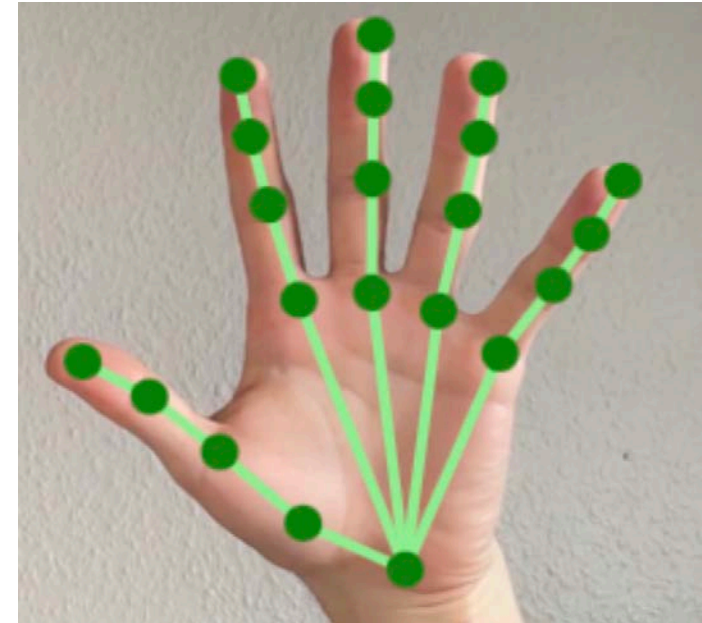
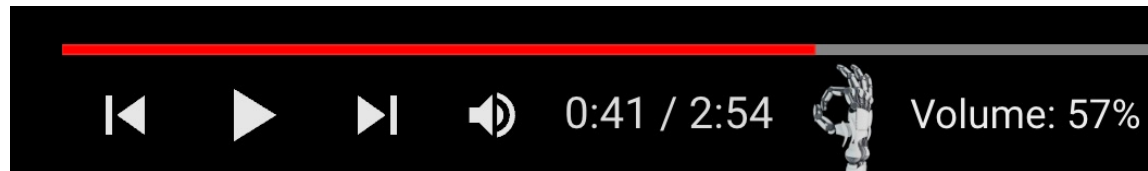
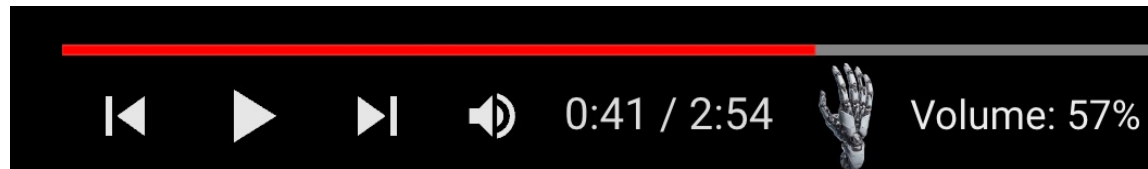


Volume down



Mute/Unmute

Демонстрація роботи застосунку



Висновки

- Розглянуто основні аспекти роботи технології розпізнавання жестів.
 - Зроблено порівняння підходів для виявлення об'єктів.
 - Використано систему Tensorflow handpose для взаємодії з ключовими точками.
 - Розроблено методи підрахунку та ідентифікації представлених пальців.
 - Вивантажено застосунок у розширення Google Chrome.
 - Отримано доступ до контрольної панелі YouTube відеоплеєра.
 - Інтегровано елементи користувацького інтерфейсу на вебсторінці.
 - Реалізовано безконтактний функціонал взаємодії з контрольною панеллю.
 - Надано застосунок у загальне користування через Chrome Web Store.
-

Дякую за увагу!