



БАГАТОРІВНЕВИЙ АНАЛІЗ ВІДЕОЗОБРАЖЕННЯ НА БАЗІ RASPBERRY PI ТА CHATGPT-4 VISION

Роботько С. П.

Аспірант

Національний університет кораблебудування імені адмірала Макарова

54007, вул. Героїв України, 9, м. Миколаїв, Україна

<https://orcid.org/0000-0002-9203-8385>

e-mail: robotkos@gmail.com

Анотація. У сучасному світі технології комп'ютерного зору та штучного інтелекту швидко розвиваються, відкриваючи нові горизонти для аналізу та інтерпретації відеозображень. Ця стаття представляє багаторівневий підхід до аналізу відеозображень, використовуючи платформу Raspberry Pi та передову модель ChatGPT-4 Vision. Основна увага приділяється методам покадрового аналізу зображень та комбінуванню різних підходів для підвищення точності та ефективності системи. Досліджено можливості Raspberry Pi 5 у виконанні складних обчислень для розпізнавання об'єктів та аналізу зображень у реальному часі. Розглянуто різні методи аналізу зображень, включаючи традиційні алгоритми обробки, методи машинного навчання та глибокого навчання з використанням згорткових нейронних мереж. Особливу увагу приділено інтеграції з моделлю ChatGPT-4 Vision, яка надає потужні можливості для розпізнавання та семантичного розуміння зображень. Запропоновано методики комбінування результатів різних рівнів аналізу для досягнення більш точних та надійних результатів. Показано, як поєднання апаратних можливостей Raspberry Pi з передовими алгоритмами штучного інтелекту може бути ефективним рішенням для широкого спектру застосувань, таких як системи безпеки, автоматизація, медична діагностика та наукові дослідження. У висновках підкреслюється важливість багаторівневого підходу до аналізу зображень та потенціал використання доступного апаратного забезпечення для реалізації складних задач комп'ютерного зору. Результати цього дослідження можуть стати основою для подальших розробок та вдосконалення систем аналізу відеозображень, сприяючи розвитку галузі загалом.

Ключові слова: аналіз відео; Raspberry Pi; камера; ChatGPT.

Вступ

У сучасну епоху цифрових технологій обробка та аналіз зображень стали невід'ємною частиною багатьох галузей, включаючи безпеку, медицину, промисловість та наукові дослідження. Зростаюча доступність потужних апаратних засобів та передових алгоритмів штучного інтелекту відкриває нові можливості для розробки ефективних систем комп'ютерного зору.

Raspberry Pi, як компактний та доступний одноплатний комп'ютер, здобув широку популярність серед дослідників та ентузіастів завдяки своїй гнучкості та можливостям інтеграції з різними сенсорами та периферією. Випуск Raspberry Pi 5 з покращеними апаратними характеристиками ще більше розширив його потенціал у задачах обробки зображень та розпізнавання об'єктів.

Запропоноване посилання: Роботько, С. (2025). Багаторівневий аналіз відео зображення на базі Raspberry Pi та Chat-GPT-4 Vvision. Методи та прилади контролю якості, 1(54), 78-88. doi: 10.31471/1993-9981-2025-1(54)-78-88

* Відповідальний автор



Copyright © The Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (<https://creativecommons.org/licenses/by/4.0/>)

Одночасно з цим, розвиток моделей штучного інтелекту, таких як ChatGPT-4 Vision, надає нові інструменти для глибокого аналізу та інтерпретації зображень. Поєднання потужності цих моделей з апаратними можливостями Raspberry Pi відкриває шлях до створення багаторівневих систем аналізу зображень, які можуть працювати в реальному часі та на місці збору даних.

Мета дослідження полягає у розробці та оцінці багаторівневого підходу до аналізу відеозображень, який об'єднує можливості Raspberry Pi 5 та моделі ChatGPT-4 Vision. Такий підхід дозволяє поєднати переваги різних методів аналізу, забезпечуючи високу точність та ефективність системи.

У цій статті розглядаються наступні аспекти:

- **Покадровий аналіз зображень:** оцінка різних методів обробки та аналізу кожного кадру відеопотоку, включаючи традиційні алгоритми та підходи машинного навчання.
- **Можливості Raspberry Pi 5:** [1] аналіз апаратних покращень нової версії Raspberry Pi та їх вплив на продуктивність системи в задачах комп'ютерного зору.
- **Інтеграція з ChatGPT-4 Vision:** дослідження способів поєднання моделі ChatGPT-4 Vision з Raspberry Pi для покращення розпізнавання та інтерпретації зображень.
- **Комбінування результатів:** розробка методик об'єднання даних з різних рівнів аналізу для підвищення загальної ефективності системи.

Актуальність дослідження зумовлена потребою у доступних та ефективних рішеннях для аналізу відеозображень, які можуть бути впроваджені в різних сферах без значних фінансових та ресурсних витрат. Використання Raspberry Pi як платформи дозволяє створювати портативні та енергоефективні системи, а інтеграція з передовими моделями штучного інтелекту розширює їх функціональні можливості.

Структура статті побудована таким чином, щоб послідовно розглянути кожен із зазначених аспектів. У розділі 2 коротко описані переваги Raspberry Pi High Quality Camera, яка використовується для збору високоякісних зображень. Розділ 3 присвячений покадровому аналізу зображень та огляду основних підходів у цій сфері. У розділі 4 детально розглядаються можливості Raspberry Pi 5 в контексті задач комп'ютерного зору. Розділ 5 фокусується на можливостях ChatGPT-4 Vision у розпізнаванні зображень та способах його інтеграції з апаратною платформою. У розділі 6 описані методи комбінування результатів багаторівневого аналізу для досягнення оптимальної продуктивності системи. Нарешті, у висновку підведені підсумки дослідження та окреслені напрямки для майбутніх робіт.

Переваги Raspberry Pi High Quality Camera

Raspberry Pi High Quality Camera (HQ Camera) є значним кроком вперед у розвитку камер для платформи Raspberry Pi. Випущена у 2020 році, ця камера надає користувачам можливість отримувати високоякісні зображення та відео, що є критично важливим для задач комп'ютерного зору та аналізу зображень. Основні переваги HQ Camera полягають у її високій роздільній здатності, гнучкості у виборі об'єктивів та покращеній чутливості при низькому освітленні.

HQ Camera оснащена 12,3-мегапіксельним сенсором Sony IMX477, який забезпечує високу роздільну здатність та деталізацію зображення. Це суттєве покращення порівняно з попередніми модулями камер для Raspberry Pi, які зазвичай мали 5 або 8 мегапікселів. Висока роздільна здатність дозволяє отримувати чіткі та детальні зображення, що є важливим для розпізнавання дрібних об'єктів або аналізу складних сцен.

У таблиці 1 наведено порівняльну таблицю ключових характеристик трьох камер для Raspberry Pi.

Таблиця 1 – Порівняльна характеристика камер

Характеристика	Raspberry Pi High Quality Camera	Arducam 16MP Autofocus Camera	Raspberry Pi Camera Module v2
Image			
Сенсор (Мпікс)	Sony IMX477 (12,3 Мпікс)	Sony IMX519 (16 Мпікс)	Sony IMX219 (8 Мпікс)
Розмір пікселя	Великі (1,55 мкм)	Менші (1,12 мкм)	Менші (1,12 мкм)
Фокусування	Ручне (C/CS mount)	Автофокус (M12 mount)	Фіксований фокус (вбудований)
Можливість заміни об'єктивів	Так (C/CS mount)	Так (M12 mount)	Ні
Якість зображення при слабкому світлі	Висока	Добра, але менші пікселі	Обмежена
Максимальна роздільна здатність фото/відео	Фото ~12 Мпікс, до 1080p@50/60fps	Фото ~16 Мпікс, до 1080p@30fps	Фото ~8 Мпікс, до 1080p@30fps
Підтримка та документація	Широка (офіційна)	Наявна, але менш поширена	Найширша (офіційна, давня)
Ціна (приблизно)	65\$	35\$	20\$
Сенсор (Мпікс)	Sony IMX477 (12,3 Мпікс)	Sony IMX519 (16 Мпікс)	Sony IMX219 (8 Мпікс)

Однією з ключових особливостей HQ Camera є підтримка змінних об'єктивів з байонетом C та CS mount. Це дає можливість користувачам обирати з широкого спектру об'єктивів з різними фокусними відстанями, апертурами та оптичними характеристиками. Така гнучкість дозволяє налаштовувати систему під конкретні задачі, чи то ширококутна зйомка для панорамних зображень, чи телефотозйомка для деталізації віддалених об'єктів. Можливість використовувати професійні об'єктиви значно розширює потенціал камери в різних застосуваннях.

Покращена чутливість при низькому освітленні є ще однією важливою перевагою HQ Camera. Сенсор Sony IMX477 має більші пікселі розміром 1,55 мкм, що дозволяє вловлювати більше світла. Це забезпечує кращу якість зображення в умовах слабого освітлення, знижує рівень шуму та підвищує загальну чутливість системи. Для задач, які передбачають

роботу в нічний час або в умовах недостатнього освітлення, це є критичним фактором.

Камера також підтримує ручне налаштування параметрів зйомки, таких як діафрагма та фокус, за умови використання відповідних об'єктивів. Це дозволяє контролювати глибину різкості, експозицію та інші важливі параметри, що є корисним для професійних застосувань та експериментальних проєктів. Регульована діафрагма дає можливість змінювати кількість світла, що потрапляє на сенсор, а ручний фокус забезпечує точне налаштування на об'єкт зйомки.

Висока якість виготовлення та надійність HQ Camera забезпечуються міцним металевим корпусом та надійним кріпленням для об'єктивів. Камера підключається до Raspberry Pi через інтерфейс CSI (Camera Serial Interface), що забезпечує високу швидкість передачі даних та стабільне з'єднання. Це особливо

важливо для застосувань, де камера може піддаватися вібраціям або іншим механічним впливам.

Сумісність з різними моделями Raspberry Pi та підтримка популярного програмного забезпечення робить HQ Camera зручною у використанні. Вона легко інтегрується з бібліотеками для обробки зображень, такими як OpenCV, що спрощує розробку програмного забезпечення для аналізу зображень та комп'ютерного зору. Крім того, активна спільнота Raspberry Pi надає широкий спектр ресурсів, інструкцій та прикладів для роботи з камерою.

Економічна ефективність HQ Camera також є значущою перевагою. При відносно низькій вартості вона пропонує функціональність та якість, які зазвичай доступні лише в професійних камерах вищого цінового сегмента. Це робить її привабливим вибором для освітніх проєктів, досліджень та стартапів з обмеженим бюджетом.

У контексті задач комп'ютерного зору та аналізу зображень HQ Camera відкриває широкі можливості. Висока роздільна здатність та якість зображення покращують точність алгоритмів розпізнавання об'єктів та класифікації. Гнучкість у виборі об'єктивів дозволяє адаптувати систему до різних сценаріїв, а покращена чутливість при низькому освітленні розширює умови, в яких може працювати система.

Таким чином, Raspberry Pi High Quality Camera є потужним інструментом для розробки систем комп'ютерного зору на базі Raspberry Pi. Її переваги в поєднанні з можливостями сучасних алгоритмів штучного інтелекту роблять її ідеальним вибором для різноманітних проєктів – від простих освітніх експериментів до складних наукових досліджень та промислових застосувань.

Покадровий аналіз зображень та огляд підходів

Покадровий аналіз зображень є фундаментальною методикою комп'ютерного зору [2], яка полягає в обробці та аналізі

кожного окремого кадру відеопотоку. Цей підхід дозволяє отримувати детальну інформацію про об'єкти та сцени в кожному кадрі, що є критично важливим для розпізнавання об'єктів, відстеження руху та інших задач аналізу зображень.

Існує декілька основних підходів до покадрового аналізу зображень, які можна розділити на традиційні методи обробки зображень, методи машинного навчання та методи глибокого навчання.

Традиційні методи обробки зображень базуються на використанні алгоритмів, які безпосередньо оперують пікселями зображення. До них належать фільтрація зображень для видалення шуму, детекція країв за допомогою операторів Собеля або Кенні, сегментація зображень для виділення об'єктів та інші техніки. Ці методи є швидкими та ефективними для простих задач, але їх точність та здатність до узагальнення обмежені, особливо в складних або варіативних умовах зйомки.

Методи машинного навчання використовують статистичні підходи для аналізу зображень. Алгоритми, такі як метод опорних векторів (SVM), k-найближчих сусідів або рішення дерев, можуть бути навчені на попередньо оброблених ознаках зображень для класифікації або регресії. Ці методи дозволяють враховувати більш складні взаємозв'язки між ознаками, але їх ефективність сильно залежить від якості вибраних ознак та обсягу навчальних даних.

Глибоке навчання та, зокрема, згорткові нейронні мережі (CNN) стали домінуючими в галузі комп'ютерного зору [3] завдяки їх здатності автоматично виявляти релевантні ознаки зображень. CNN [4] можуть обробляти «сирі» піксельні дані та вчитися розпізнавати складні візуальні патерни через багаторівневі згорткові шари. Це дозволяє досягати високої точності в задачах класифікації зображень, детекції об'єктів та семантичної сегментації.

Однак, використання глибокого навчання часто пов'язане з високими обчислювальними витратами та потребою в великому обсязі навчальних даних. Для систем на базі Raspberry Pi це може бути викликом через обмежені апаратні ресурси. Проте з появою Raspberry Pi 5 з покращеною продуктивністю стає можливим використовувати оптимізовані моделі глибокого навчання або застосовувати методи компресії моделей, такі як прунг та квантизація, для зменшення їх розміру та обчислювальної складності.

Крім того, існують гібридні підходи, які поєднують традиційні методи обробки зображень з методами машинного або глибокого навчання. Наприклад, попередня обробка зображень для виділення областей інтересу може зменшити обсяг даних, які потребують подальшого аналізу, що підвищує ефективність системи.

Важливим аспектом покадрового аналізу є також управління часом обробки. Для систем, які працюють в реальному часі, критично важливо забезпечити достатню швидкість обробки кадрів. Оптимізація алгоритмів, використання апаратних прискорювачів та паралельна обробка є ключовими факторами в досягненні необхідної продуктивності.

У контексті цієї роботи, покадровий аналіз зображень стає основою для побудови багаторівневої системи, яка комбінує різні підходи для досягнення оптимальної точності та швидкості. Використання Raspberry Pi 5 дозволяє реалізувати частину обчислень на місці, зменшуючи потребу в передачі великих обсягів даних на віддалені сервери. Інтеграція з моделлю [5] ChatGPT-4 Vision надає додаткові можливості для глибокого розуміння та інтерпретації зображень, що виходить за рамки простого розпізнавання об'єктів.

Таким чином, огляд підходів до покадрового аналізу зображень демонструє широкий спектр методів та технологій, які можуть бути використані для різних задач. Вибір конкретного підходу залежить від вимог до точності, швидкості, ресурсів та

специфіки застосування. Комбінування цих підходів у багаторівневій системі дозволяє скористатися їх перевагами та компенсувати недоліки, що є ключовим для створення ефективних та надійних систем аналізу зображень.

Можливості Raspberry Pi 5 в аналізі зображень та розпізнаванні об'єктів

Випуск Raspberry Pi 5 став значущою подією в спільноті розробників та ентузіастів одноплатних комп'ютерів. Ця нова версія приносить суттєві покращення в апаратній частині, що розширює можливості використання Raspberry Pi в задачах аналізу зображень та розпізнавання об'єктів. У цьому розділі розглянемо основні апаратні та програмні вдосконалення Raspberry Pi 5 та їх вплив на продуктивність систем комп'ютерного зору.

Покращений процесор та продуктивність

Raspberry Pi 5 оснащений потужним чотириядерним процесором ARM Cortex-A76 з тактовою частотою до 2,4 ГГц. Це значний крок вперед порівняно з попередніми моделями, які використовували Cortex-A72. Нова архітектура процесора забезпечує підвищену продуктивність та ефективність енергоспоживання. Для задач аналізу зображень, які потребують значних обчислювальних ресурсів, це означає швидшу обробку даних та можливість виконання складніших алгоритмів у реальному часі.

Покращена графіка та апаратне прискорення

Відеоядро VideoCore VII в Raspberry Pi 5 підтримує сучасні графічні API, такі як OpenGL ES 3.1 та Vulkan 1.2. Це відкриває можливості для використання апаратного прискорення графіки та обробки зображень. Апаратне прискорення дозволяє виконувати операції, такі як згортки та матричні множення, більш ефективно, що є критичним для глибоких нейронних мереж. Використання GPU для

обчислень також звільняє ресурси CPU для інших задач.

Збільшений обсяг оперативної пам'яті

Raspberry Pi 5 доступний у конфігураціях з оперативною пам'яттю до 8 ГБ LPDDR4X-4266 RAM. Більший обсяг та швидкість оперативної пам'яті дозволяють завантажувати та обробляти більші моделі нейронних мереж та обробляти більш високоякісні зображення. Це особливо важливо для застосувань, де потрібно одночасно обробляти кілька потоків даних або працювати з відео високої роздільної здатності.

Покращений інтерфейс для камер

Raspberry Pi 5 має два роз'єми для камер CSI (Camera Serial Interface), що дозволяє підключати одночасно дві камери. Це відкриває можливості для стереозйомки та більш складних застосувань комп'ютерного зору, таких як 3D-реконструкція та відстеження руху в просторі. Крім того, покращена пропускна здатність інтерфейсу дозволяє працювати з камерами високої роздільної здатності та частотою кадрів.

Підтримка інтерфейсу PCIe

Нова модель оснащена роз'ємом PCIe 2.0x1, що дозволяє підключати високошвидкісні периферійні пристрої, такі як твердотільні накопичувачі (SSD) або спеціалізовані прискорювачі штучного інтелекту (наприклад, Google Coral TPU). Це відкриває можливості для подальшого підвищення продуктивності системи та розширення її функціональності.

Покращена енергетична ефективність та тепловідведення

Розробники Raspberry Pi 5 приділили увагу оптимізації енергоспоживання та тепловідведення. Ефективна система охолодження забезпечує стабільну роботу при високих навантаженнях, що є важливим для тривалого виконання обчислювально інтенсивних задач, таких як аналіз зображень у реальному часі.

Програмна підтримка та сумісність

Завдяки широкій спільноті розробників, Raspberry Pi 5 підтримує різноманітні операційні системи та програмне забезпечення, оптимізоване для ARM-архітектури. Для задач комп'ютерного зору доступні бібліотеки та фреймворки, такі як:

- **OpenCV:** [6] широко використовувана бібліотека з відкритим кодом для обробки зображень та комп'ютерного зору.

- **TensorFlow Lite:** [7] полегшена версія фреймворку TensorFlow, оптимізована для виконання моделей машинного навчання на пристроях з обмеженими ресурсами.

- **PyTorch Mobile:** підтримка виконання моделей PyTorch на мобільних та вбудованих пристроях.

Можливості для глибокого навчання на пристрої

З покращеними апаратними характеристиками, Raspberry Pi 5 (рис. 1) здатний виконувати складні моделі глибокого навчання безпосередньо на пристрої. Це знижує залежність від хмарних сервісів та дозволяє обробляти дані локально, що є важливим для застосувань з обмеженнями на передачу даних або вимогами до конфіденційності.

NCNN 20230517

Target: Vulkan GPU-v2-v2 - Model: mobilenet-v2



Рисунок 1 – Порівняння графічних обчислень

Реалізація алгоритмів розпізнавання об'єктів

Покращена продуктивність Raspberry Pi 5 дозволяє реалізовувати сучасні алгоритми розпізнавання об'єктів, такі як:

- **YOLO (You Only Look Once):** модель для швидкого та точного виявлення об'єктів у реальному часі.

- **SSD (Single Shot MultiBox Detector):** ефективна модель для виявлення об'єктів, оптимізована для мобільних пристроїв.

- **MobileNet:** архітектура нейронних мереж, розроблена для роботи на пристроях з обмеженими ресурсами.

Приклади застосувань

- **Системи відеоспостереження та безпеки:** виявлення та розпізнавання осіб, відстеження руху, виявлення підозрілої поведінки.

- **Робототехніка:** навігація на основі зору, розпізнавання перешкод, взаємодія з навколишнім середовищем.

- **Промисловий контроль якості:** автоматичний огляд продуктів, виявлення дефектів, сортування.

Огляд можливостей розпізнавання зображень від ChatGPT-4 Vision

Сучасні досягнення в галузі штучного інтелекту призвели до появи потужних моделей, здатних до глибокого аналізу зображень. ChatGPT-4 Vision поєднує можливості мовної моделі з комп'ютерним зором, надаючи розширені функції розпізнавання та інтерпретації візуальної інформації. Порівняємо її можливості з локальними моделями глибокого навчання, що виконуються безпосередньо на Raspberry Pi 5.

Демонстрація можливостей ChatGPT-4 Vision. Розглянемо практичний приклад використання ChatGPT-4 Vision для розпізнавання складного об'єкта. Маємо зображення металевого предмета, що лежить у густій траві. Об'єкт частково прихований рослинністю та має відблиски, що ускладнює його ідентифікацію за допомогою локальних моделей на пристрої.

ChatGPT-4 Vision здатна не лише розпізнати об'єкт [8], але й надати детальний опис сцени (рис. 2).

Додавши до статті зображення цього металевого об'єкта та результати його аналізу за допомогою ChatGPT-4 Vision [9], можна наочно продемонструвати переваги моделі в складних умовах

розпізнавання. Результатом аналізу зображення є текстовий опис: «Об'єкт на землі, ймовірно, є нерозірваним боєприпасом (UXO), можливо, суббоєприпасом або частиною касетної бомби. Його циліндрична форма з внутрішніми компонентами є характерною рисою таких пристроїв. Вони є надзвичайно небезпечними і не повинні бути потривожені за жодних обставин».



Рисунок 2 – Зображення, оброблене ChatGPT-4 Vision

Аналіз переваг та недоліків: ChatGPT-4 Vision забезпечує високу точність і детальність розпізнавання, здатна розуміти контекст та взаємодію між об'єктами, а також відповідає на різноманітні запитання щодо зображення. Однак вона залежить від стабільного підключення до Інтернету, що може бути обмеженням у деяких застосуваннях. Питання конфіденційності також виникають через передачу даних у хмару, а використання сервісу може потребувати додаткових витрат.

Локальні моделі на пристрої працюють автономно, забезпечуючи контроль над даними та можливість адаптації під конкретні задачі. Проте їх продуктивність і точність можуть бути обмежені апаратними ресурсами Raspberry Pi 5, а налаштування та оптимізація можуть вимагати значних зусиль. У таблиці 2 ми порівнюємо характеристики локальних моделей та можливості ChatGPT-4 Vision.

Таблиця 2 – Порівняльна характеристика моделей

Характеристика	Локальні моделі на пристрої	ChatGPT-4 Vision
Обчислювальні ресурси	Виконуються на Raspberry Pi 5 з обмеженими ресурсами	Використовує хмарні сервери з високою продуктивністю
Точність розпізнавання	Може бути обмеженою через компроміс між швидкістю та точністю	Висока точність завдяки масштабному навчанню
Швидкість обробки	Оптимізовані моделі працюють у реальному часі	Висока швидкість завдяки потужним серверам
Семантичне розуміння	Обмежене розуміння; зазвичай зводиться до класифікації об'єктів	Глибоке розуміння контексту та взаємодії об'єктів
Потреба в Інтернеті	Не потребує інтернету; працює автономно	Потребує стабільного підключення до мережі
Конфіденційність даних	Дані залишаються на пристрої	Дані передаються в хмару; можливі питання конфіденційності
Налаштування та адаптація	Можна адаптувати під конкретні задачі	Обмежена можливість налаштування користувачем
Вартість використання	Одноразові витрати на апаратне забезпечення	Може включати постійні витрати за доступ до сервісу
Енергоспоживання	Низьке енергоспоживання; підходить для автономних систем	Енергоспоживання на стороні сервера
Масштабованість	Обмежена можливостями пристрою	Висока; масштабування на рівні хмарних сервісів

Оптимальним рішенням може стати комбінований підхід, де локальні моделі виконують базові завдання розпізнавання на пристрої, а ChatGPT-4 Vision залучається для глибшого аналізу за необхідності. Це дозволяє:

- **Знизити навантаження на мережу:** передавати в хмару лише ті зображення, які потребують детального аналізу.

- **Підвищити ефективність:** використовувати переваги обох підходів для досягнення кращих результатів.

- **Збалансувати швидкість і точність:** локальні моделі забезпечують швидку обробку, а ChatGPT-4 Vision додає семантичної глибини.

Комбінування результатів багаторівневого аналізу

Багаторівневий аналіз зображень передбачає використання різних методів та інструментів для отримання максимально повної та точної інформації про вміст зображення. У контексті цієї роботи

поєднання локальних моделей глибокого навчання на пристрої з можливостями ChatGPT-4 Vision дозволяє створити ефективну систему аналізу, що використовує переваги кожного з підходів.

Комбінування результатів багаторівневого аналізу полягає в інтеграції даних, отриманих на різних етапах та рівнях обробки, для досягнення більш точного та повного розуміння зображення. Це може включати декілька аспектів:

Поєднання локальних і хмарних обчислень: локальні моделі на Raspberry Pi 5 виконують початкову обробку та аналіз зображень. Вони можуть швидко визначати основні об'єкти та проводити базову класифікацію. У випадках, коли локальна модель не може з високою точністю ідентифікувати об'єкт або потрібен глибший аналіз, зображення передається до ChatGPT-4 Vision. Такий підхід дозволяє зменшити обсяг даних, що передаються в мережу, та оптимізувати використання ресурсів.

Інтеграція результатів різних моделей: результати локальних моделей та ChatGPT-4 Vision можуть бути об'єднані для покращення загальної точності системи. Наприклад, якщо обидві моделі ідентифікують об'єкт однаково, це підвищує впевненість у результаті. У разі розбіжностей, система може використовувати додаткові критерії або залучати користувача для уточнення. Такий підхід дозволяє враховувати різні перспективи та компенсувати можливі помилки окремих моделей.

Використання контекстної інформації: локальні моделі можуть забезпечувати швидкий аналіз певних аспектів зображення, таких як кольори, форми або рух об'єктів. ChatGPT-4 Vision додає до цього семантичне розуміння та контекст, що дозволяє інтерпретувати взаємодію між об'єктами та робити висновки на основі сцени в цілому. Комбінування цих рівнів аналізу забезпечує більш глибоке та повне розуміння зображення.

Оптимізація алгоритмів обробки: поєднання результатів дозволяє оптимізувати алгоритми обробки шляхом використання найефективніших методів для кожного конкретного завдання. Локальні моделі можуть бути налаштовані на швидке виявлення певних об'єктів або подій [10], тоді як ChatGPT-4 Vision може використовуватися для складніших завдань, що потребують глибокого аналізу. Це дозволяє зменшити затримки та підвищити продуктивність системи.

Покращення надійності та стійкості системи: комбінування результатів від різних моделей підвищує надійність системи, оскільки знижує ймовірність помилок, пов'язаних з обмеженнями окремих методів. Якщо одна модель не може правильно розпізнати об'єкт через шум або інші перешкоди, інша модель може компенсувати це завдяки своїм особливостям. Такий підхід підвищує стійкість системи до зовнішніх впливів та варіацій у даних.

Забезпечення балансу між продуктивністю та ресурсами: Багаторівневий

аналіз дозволяє ефективно розподіляти обчислювальні ресурси. Локальні моделі використовують потужність Raspberry Pi 5 для виконання задач, які не потребують значних ресурсів, тоді як складніші задачі передаються до хмарної моделі ChatGPT-4 Vision. Це забезпечує оптимальний баланс між швидкістю обробки та використанням ресурсів.

Реалізація на практиці: для реалізації комбінованого підходу необхідно забезпечити ефективну інтеграцію між локальними моделями та ChatGPT-4 Vision. Це включає розробку алгоритмів для прийняття рішень про те, коли залучати хмарну модель, а також механізмів для об'єднання та узгодження результатів. Важливим аспектом є також забезпечення безпеки та конфіденційності даних при передачі інформації до хмари.

Приклад застосування: У системі відеоспостереження локальна модель може постійно моніторити сцену та виявляти рух або появу нових об'єктів. Якщо виявлено підозрілий об'єкт або подію, система може автоматично передати відповідне зображення до ChatGPT-4 Vision для детального аналізу та отримання розширеного опису. Це дозволяє оперативно реагувати на потенційні загрози та приймати обґрунтовані рішення.

Переваги комбінованого підходу:

- **Підвищена точність:** об'єднання результатів різних моделей дозволяє досягти більш точних та надійних результатів.

- **Ефективність:** оптимальне використання ресурсів та розподіл обчислювальних задач між локальними та хмарними компонентами.

- **Гнучкість:** система може бути налаштована під конкретні вимоги та умови експлуатації.

- **Масштабованість:** можливість додавання нових моделей або сервісів без значних змін у структурі системи.

Виклики

У цій статті було досліджено багаторівневий підхід до аналізу відеозображень з використанням платформи Raspberry Pi

та моделі ChatGPT-4 Vision. Розглянуто переваги Raspberry Pi High Quality Camera, яка надає високу якість зображення та гнучкість у налаштуваннях, що є критично важливим для задач комп'ютерного зору.

Проведено огляд методів покадрового аналізу зображень, включаючи традиційні алгоритми обробки, методи машинного навчання та глибокого навчання. Особлива увага приділена можливостям Raspberry Pi 5, який завдяки покращеним апаратним характеристикам дозволяє виконувати складні обчислення безпосередньо на пристрої, що розширює можливості автономних систем аналізу зображень.

Розглянуто можливості ChatGPT-4 Vision у розпізнаванні та інтерпретації зображень, а також її інтеграцію з локальними моделями на пристрої. Показано, що комбінування результатів багаторівневого аналізу дозволяє досягти більш високої точності та ефективності, поєднуючи переваги локальних обчислень та хмарних сервісів.

Основні висновки дослідження:

- підвищена точність розпізнавання: поєднання локальних моделей та ChatGPT-4 Vision дозволяє компенсувати недоліки кожного з підходів та досягти більш надійних результатів;
- ефективне використання ресурсів: оптимальний розподіл обчислювальних задач між пристроєм та хмарою підвищує загальну продуктивність системи;
- гнучкість та адаптивність: можливість налаштування системи під конкретні задачі та умови експлуатації забезпечує широке коло застосувань.

Перспективи подальших досліджень:

- покращення конфіденційності та безпеки: розробка методів захисту даних при передачі до хмарних сервісів та використання локальних обчислень для чутливих даних;
- оптимізація моделей для вбудованих систем: створення легковагових та ефективних моделей глибокого навчання, які можуть працювати на обмежених апаратних ресурсах Raspberry Pi;
- розширення функціональності системи: інтеграція додаткових сенсорів та

використання мультимодальних даних для покращення розпізнавання та інтерпретації.

Висновки

У даній статті було продемонстровано підхід до багаторівневого аналізу відеозображень на базі Raspberry Pi із залученням можливостей моделі ChatGPT-4 Vision. Розглянуто переваги використання Raspberry Pi High Quality Camera, методи покадрового аналізу зображень та проаналізовано можливості Raspberry Pi 5 у виконанні ресурсомістких завдань комп'ютерного зору. Було показано, що поєднання локальних обчислень з глибокими нейронними мережами та інтеграцією хмарних сервісів, таких як ChatGPT-4 Vision, дозволяє отримати високоточні та семантично глибокі результати розпізнавання.

Завдяки використанню Raspberry Pi 5 як апаратної основи вдалося досягти балансу між продуктивністю та економічною ефективністю. Комбінування результатів від різних рівнів аналізу — базових локальних моделей на пристрої та розширеного контекстного аналізу з боку ChatGPT-4 Vision — підвищує надійність, точність та швидкість обробки. Застосування такого підходу є перспективним у широкому спектрі завдань, включаючи системи моніторингу й безпеки, робототехнічні платформи, медичну діагностику, промисловий контроль якості та наукові дослідження.

Перспективи подальших робіт полягають в оптимізації моделей для вбудованих систем, підвищенні безпеки та конфіденційності даних, а також у використанні мультимодальних підходів для покращення інтерпретації та аналізу складних сцен.

Подяки

Відсутні.

Конфлікт інтересів

Відсутній.

Список використаних джерел / References

1. Raspberry Pi Foundation. Raspberry Pi Documentation URL: <https://www.raspberrypi.com/documentation/>
2. Howard A.G., Zhu M., Chen B. et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. arXiv:1704.04861, 2017.
3. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
4. Liu W., Anguelov D., Erhan D. et al. SSD: Single Shot MultiBox Detector. European Conference on Computer Vision (ECCV), 2016.
5. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks. NIPS, 2012.
6. OpenCV Library URL: <https://opencv.org/>
7. Abadi M., Barham P., Chen J. et al. TensorFlow: A system for large-scale machine learning. OSDI. 2016.
8. Paszke A., Gross S., Massa F. et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. NeurIPS, 2019.
9. OpenAI. ChatGPT & GPT-4 API Documentation URL: <https://platform.openai.com/docs/introduction>
10. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. NIPS, 2015. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)

MULTI-LEVEL ANALYSIS OF VIDEO IMAGES BASED ON RASPBERRY PI AND CHATGPT-4 VISION

Robotko S. P.

Postgraduate Student

Admiral Makarov National Shipbuilding University

54007, 9 Heroes of Ukraine St., Mykolaiv, Ukraine

<https://orcid.org/0000-0002-9203-8385>

e-mail: robotkos@gmail.com

Abstract. In the modern world, computer vision and artificial intelligence technologies are rapidly evolving, opening up new horizons for the analysis and interpretation of video images. This article presents a multi-level approach to video image analysis, using the Raspberry Pi platform and the advanced ChatGPT-4 Vision model. Particular attention is paid to frame-by-frame image analysis methods and the combination of various approaches to improve system accuracy and efficiency. The capabilities of the Raspberry Pi 5 for performing complex computations in object recognition and real-time image analysis have been investigated. Various image analysis methods are considered, including traditional image processing algorithms, machine learning approaches, and deep learning techniques employing convolutional neural networks. Special emphasis is placed on integration with the ChatGPT-4 Vision model, which provides powerful capabilities for image recognition and semantic understanding. Techniques for combining the results obtained at different levels of analysis are proposed to achieve more accurate and reliable outcomes. It is shown how the integration of Raspberry Pi's hardware capabilities with cutting-edge artificial intelligence algorithms can be an effective solution for a wide range of applications such as security systems, automation, medical diagnostics, and scientific research. The conclusions highlight the importance of a multi-level approach to image analysis and the potential of using affordable hardware to tackle complex computer vision tasks. The results of this study may form the basis for further development and enhancement of video image analysis systems, contributing to the overall progress of the field.

Keywords: video analysis; Raspberry Pi; camera; ChatGPT.