ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
КАФЕДРА ІНФОРМАЦІЙНИХ СИСТЕМ ТА ТЕХНОЛОГІЙ

КВАЛІФІКАЦІЙНА РОБОТА

на тему:
«МОДЕЛЬ СИСТЕМИ ВІДЕОСПОСТЕРЕЖЕННЯ ЗІ ШТУЧНИМ ІНТЕЛЕКТОМ»

на здобуття освітнього ступеня магістр
за спеціальності 126 Інформаційні системи та технології
(код,найменування спеціальності)
освітньо-професійної програми Інформаційні системи та технології
(назва)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Олексій ДЕГТЯР
(підпис) *(ім'я, ПРІЗВИЩЕ здобувача)*

Виконав: здобувач вищої освіти гр.ІСДМ-63
Олексій ДЕГТЯР
(ім'я, ПРІЗВИЩЕ)
Керівник: Ольга ПОЛОНЕВИЧ
к.т.н,доцент. *(ім'я, ПРІЗВИЩЕ)*
Рецензент: _____
, *(ім'я, ПРІЗВИЩЕ)*

Київ 2024

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут Інформаційних технологій

Кафедра Кафедра Інженерії систем та технологій
Ступінь вищої освіти магістр
Спеціальність 126 Інформаційні системи та технології
Освітньо-професійна прорама Інформаційні системи та технології

ЗАТВЕРДЖУЮ
Завідувач кафедрою ІСТ
_____ Каміла СТОРЧАК

"____" _____2024 року

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Дегтяру Олексію Михайловичу
(прізвище, ім'я, по батькові здобувача)

1.Тема кваліфікаційної роботи: Модель системи відеоспостереження зі штучним інтелектом
керівник кваліфікаційної роботи Ольга ПОЛОНЕВИЧ, к.т.н.,доцент
(ім'я, ПРІЗВИЩЕ, науковий ступінь, вчене звання)
затверджені наказом Державного університету інформаційно-комунікаційних технологій від "15" 10.2024 р. № 320
2. Строк подання кваліфікаційної роботи «27» грудня 2024 р.
3. Вихідні дані кваліфікаційної роботи:
1. Система відеоспостереження.
2. Штучний Інтелект.
3. Комп'ютерний зір.
4. Науково-технічна література.

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити):
1. Огляд тенденцій розвитку та особливостей застосування штучного інтелекту у відеоспостереженні.
2. Аналіз архітектури глибоких нейронних мереж та особливості їх Застосування.
3. Впровадження моделі системи відеоспостереження зі штучним інтелектом.

5. Перелік ілюстраційного матеріалу: *презентація*

6. Дата видачі завдання «15» жовтня 2024р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Підбір технічної літератури	16.10.24-27.10.24	Виконано
2.	Аналіз тенденцій розвитку та особливостей застосування штучного інтелекту у відеоспостереженні	27.10.24-05.11.24	Виконано
3.	Дослідження архітектури глибоких нейронних мереж та особливості їх застосування	05.11.24-14.11.24	Виконано
4.	Практичне впровадження моделі системи відеоспостереження зі штучним інтелектом	14.11.24-22.11.24	Виконано
5.	Висновки по роботі	22.11.24-04.12.24	Виконано
6.	Розробка демонстраційних матеріалів, доповідь.	04.12.24-16.12.24	Виконано
7.	Оформлення роботи	16.12.24-25.12.24	Виконано

Здобувач вищої освіти _____ Олексій ДЕГТЯР
(підпис) *(ім'я, ПРІЗВИЩЕ)*

Керівник кваліфікаційної роботи _____ Ольга ПОЛОНЕВИЧ
(підпис) *(ім'я, ПРІЗВИЩЕ)*

РЕФЕРАТ

Текстова частина кваліфікаційної роботи на здобуття ступня магістр: 82 стор., 22 рис., 5 табл., 30 джерел.

Мета роботи – реалізація інтелектуальної системи виявлення, відстеження та підрахунку об’єктів із застосуванням ШІ.

Об’єкт дослідження – процес виявлення об’єктів.

Предмет дослідження – інтелектуальні системи виявлення об’єктів.

Короткий зміст роботи. Виконуючи перше поставлене завдання, у магістерській роботі, мною досліджено ринок розвитку систем відеоспостереження, особливості функціонування та перспективи розвитку.

Проаналізовано особливості впровадження ШІ в ситстеми відеоспостереження, виконано порівняння з традиційними ситстемами спостереження.

У другому розділі мною було розглянуто останні розробки глибоких нейронних мереж. Досліджено деякі широко використовувані архітектури глибокого навчання та виділено окремі програми для комп’ютерного зору, розпізнавання образів і розпізнавання мови.

У третьому розділі досліджено портативна система відеоспостереження з обробкою ШІ, яка може виявляти та відстежувати людей.

КЛЮЧОВІ СЛОВА: ШТУЧНИЙ ІНТЕЛЕКТ, ВІДЕОСПОСТЕРЕЖЕННЯ, КАМЕРА, МЕТОД, АРХІТЕКТУРА, НЕЙРОННА МЕРЕЖА, ГЛИБОКЕ НАВЧАННЯ, МОНІТОРИНГ, ЕНЕРГОЕФЕКТИВНІСТЬ, РЕСУРС.

ABSTRACT

The text part of the qualifying work for obtaining a bachelor's degree: 82 pp., 22 fig., 5 tables, 30 sources.

The purpose of the work is to implement an intelligent system for detecting, tracking and counting objects using AI.

The object of the study is the process of detecting objects.

The subject of the study is intelligent object detection systems.

Summary of the work. In fulfilling the first task, in my master's thesis, I studied the market for the development of video surveillance systems, the features of their functioning and development prospects.

The features of implementing AI in video surveillance systems were analyzed, and a comparison was made with traditional surveillance systems.

In the second section, I reviewed the latest developments in deep neural networks. Some widely used deep learning architectures were studied and separate programs for computer vision, pattern recognition and speech recognition were highlighted.

In the third section, a portable video surveillance system with AI processing that can detect and track people was studied.

KEYWORDS: ARTIFICIAL INTELLIGENCE, VIDEO SURVEILLANCE, CAMERA, METHOD, ARCHITECTURE, NEURAL NETWORK, DEEP LEARNING, MONITORING, ENERGY EFFICIENCY, RESOURCE.

ЗМІСТ

ВСТУП

Актуальність теми. Сьогодні глибоке навчання показало великі переваги в в багатьох галузях життєдіяльності людини. Зокрема, комп'ютерний зір був першою галуззю досліджень, у якій відбулося глибоке навчання. Нові методи обробки, засновані на глибокому навчанні, замінюють традиційні алгоритми комп'ютерного зору, що спираються на фізичні представлення, моделі, функції з певним рівнем значення глибокого навчання проти традиційного комп'ютерного зору. Досконаліші системи здатні обробляти величезні обсяги даних у великих обчислювальних системах. Поточні виклики пов'язані з навчанням системи машинного навчання достатньою кількістю інформації, яка потребує маркування цих даних.

Модернізація технологій як на програмному, так і на апаратному рівні дозволяє розробляти інтелектуальні відеосистеми, здатні не тільки керувати відеоподачами із замкнутого контуру камери, але й аналізувати та отримувати інформацію в режимі реального часу з відеопотоків.

Мета роботи – реалізація інтелектуальної системи виявлення, відстеження та підрахунку об'єктів із застосуванням ШІ.

Для досягнення мети, у магістерській роботі успішно виконано наступні завдання:

- огляд тенденцій розвитку та особливостей застосування штучного інтелекту у відеоспостереженні;

- аналіз архітектури глибоких нейронних мереж та особливості їх застосування;

- впровадження моделі системи відеоспостереження зі штучним інтелектом.

Об'єкт дослідження – процес виявлення об'єктів.

Предмет дослідження – інтелектуальні системи виявлення об'єктів.

Методи дослідження. Під час написання магістерської кваліфікаційної роботи були використані методи

Наукова новизна одержаних результатів. Реалізована нова, ефективна

система відеоспостереження, що поєднує переваги глибокого навчання та спеціалізованого апаратного забезпечення. Досягнуто високої швидкості обробки відеопотоків завдяки використанню спеціалізованого апаратного забезпечення (VPU) та оптимізації програмного забезпечення.

Практична значущість одержаних результатів. Результати дослідження свідчать про високий потенціал розробленої системи для вирішення актуальних завдань у галузі комп'ютерного зору та систем безпеки. Система може бути використана для підвищення безпеки в різних об'єктах. Застосування в робототехніці, автомобільній промисловості та медицині.

Апробація результатів магістерської роботи. Ключові результати та висновки магістерської роботи були презентовані на двох наукових конференціях, організованих Державним університетом інформаційно-комунікаційних технологій.

РОЗДІЛ 1 ТЕНДЕНЦІЇ РОЗВИТКУ ТА ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ВІДЕОСПОСТЕРЕЖЕННІ

1.1 Особливості функціонування системи відеоспостереження

У сучасному суспільстві відеоспостереження повністю увійшло в наше повсякденне життя, поширившись з громадських місць на будинки. Широке використання систем відеоспостереження забезпечує людям більший захист, особливо з точки зору безпеки. Є старе китайське прислів'я, яке говорить: «Вуха порожні, очі справжні». Системи відеоспостереження ідеально підходять для задоволення потреб людей у візуальній інформації. Отже, як розвивалося відеоспостереження? Які етапи він пройшов, щоб виникнути? Система відеоспостереження першого покоління відноситься до традиційної аналогової системи відеоспостереження, в центрі якої лежить комутаційна матриця. Ця система складається з таких елементів, як аналогові камери, спеціальні відеокабелі, комутаційні матриці, монітори, аналогові відеореєстратори та магнітні стрічки. Ця система в основному використовувалася для вирішення вимог спостереження на малих і коротких відстанях. Камери безпеки OHWOAI — це 5-мегапіксельні бездротові системи запису NVR, які підтримують змішані входи 3-мегапіксельних і 5-мегапіксельних камер, а також можна додати сумісні старіші камери. Крім того, оскільки це бездротові камери безпеки, вони прості в установці та не вимагають проводки. Монітор може контролювати 10 каналів, тому можна додати до 10 камер загалом. особливості: Відеосигнал збирається, передається і зберігається в аналоговому форматі з високою якістю. Після десятиліть розвитку відповідні технології постійно вдосконалювалися та розвивалися. Однак системи відеоспостереження першого покоління мали очевидні недоліки: Вони мали обмежений радіус дії, їх можна було застосовувати лише на обмежених відстанях, ними не можна було керувати чи отримати доступ віддалено. Розширення та модернізація системи було складним і дорогим. Це вимагало широкого використання носіїв інформації, таких як магнітна стрічка, ручних запитів, а також

складного керування та обслуговування. Це не дозволяло ефективно інтегруватись з іншими системами безпеки.

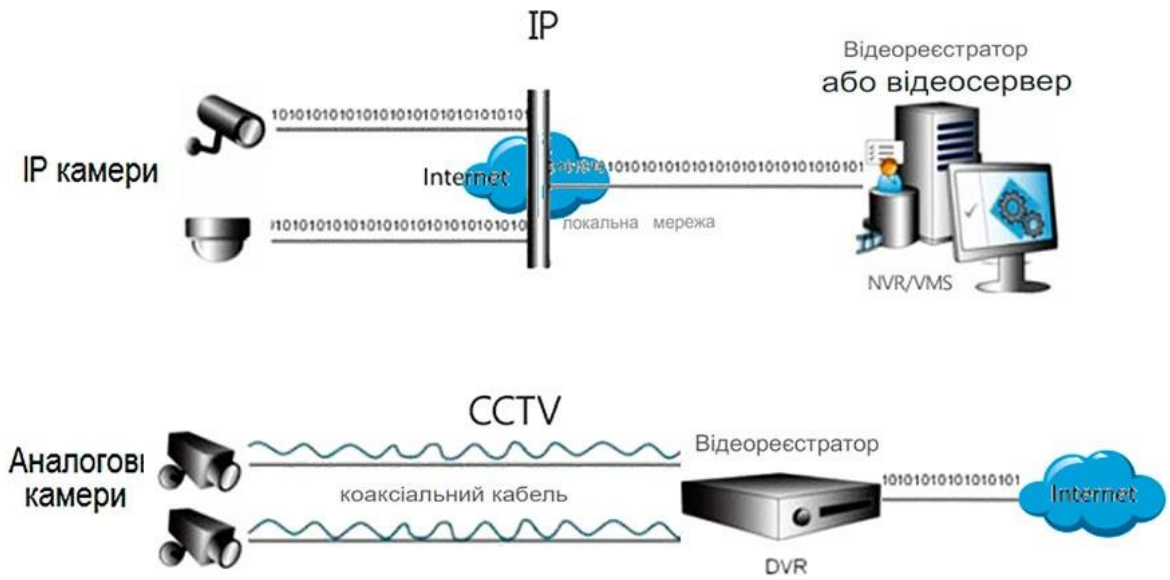


Рис.1.1 Порівняння принципу роботи аналогових та IP камер представлений

Друге покоління систем відеоспостереження відноситься до напівцифрових систем відеоспостереження, представлених реєстраторами на жорсткий диск. Це покоління систем повністю використовує комп'ютерні технології, надаючи користувачам більш зручні методи попереднього перегляду та керування. Цю систему можна вважати розширенням технології першого покоління. особливості: Відео та аудіосигнали збираються та зберігаються в цифровому форматі та мають високу якість. Цифрове збереження значно покращило можливості користувачів обробляти та шукати записану інформацію. Сумісний з аналоговими продуктами відеоспостереження першого покоління, що дозволяє здійснювати оновлення та модернізацію. Функція мережі та оптоволоконна передача системи запису на жорсткий диск вирішують проблему віддаленої передачі відеозображень, реалізуючи широкомасштабний моніторинг і спільне використання відеоресурсів у віддалених місцях. Вбудована система запису на жорсткий диск забезпечує високу надійність і легкість встановлення. У міру того, як технологія запису на жорсткий диск розвивається, відеоспостереження все більше застосовується в споживчому секторі.

Однак широкомасштабне впровадження систем відеоспостереження спричинило кілька проблем: На систему продовжував впливати аналоговий моніторинг першого покоління, де передача від точки моніторингу до центру була аналоговою, вимагаючи прокладання електричних проводів і оптоволоконних кабелів, і чим більша система, тим вищі витрати на будівництво та більше важко було підтримувати. Ємність кожного пристрою обмежена, і вони не підходять для централізованого запису у великих системах. Можливості мережі обмежені, а керування та обслуговування стає складним у великих програмах. Система відеоспостереження третього покоління відноситься до повністю цифрової системи відеоспостереження, яка використовує мережеве відео як ядро. Відео збирається як цифровий сигнал із переднього кінця та передається через мережу, і користувачі можуть переглядати, контролювати та зберігати всю систему через контрольний хост у мережі. особливості: Він має стекову структуру, є дуже гнучким і масштабованим і підтримує будь-яку топологію мережі. Економніша та ефективніша базова інфраструктура спрощує ієрархію керування та економить багато кабелів. Простий в установці та обслуговуванні, він пропонує широкий спектр рішень. Відзнятий матеріал передається, тиражується та зберігається без втрат, не має обмежень щодо відстані та доступний у будь-який час. Він використовує всі переваги зрілої технології мережі TCP/IP і пропонує різноманітні методи підключення. Його можна ефективно інтегрувати з існуючими та новими програмами та забезпечує уніфіковане керування кількома мережами. Зараз широко використовуються системи третього покоління, але в деяких областях все ще використовуються системи першого та другого покоління.

1.2 Огляд ринку систем відеоспостереження

Процес перегляду сценарію або подій і пошуку певної поведінки, яка є неправильною або може сигналізувати про початок або поширеність невідповідної поведінки, відомий як відеоспостереження. Моніторинг населення біля входу до спортивних заходів, громадського транспорту (вокзали, аеропорти тощо), а також

навколо кордону охоронюваних установ, особливо тих, які безпосередньо відмежовані громадськими зонами, є типовими застосуваннями відеоспостереження.

Аналіз відеозаписів безпеки під час моніторингу відеоспостереження в режимі реального часу дає користувачам більше знань про сценарій, дозволяючи їм краще судити про те, скільки та який тип активів відправляти. Такий одночасний моніторинг живої та архівної зйомки може підтвердити неправильне орієнтування чи будь-яку кримінальну дію до того, як клієнт, користувач послуг або підозрюваний зіткнеться з поліцейським підрозділом, залежно від кількості камер спостереження та їхнього розташування.

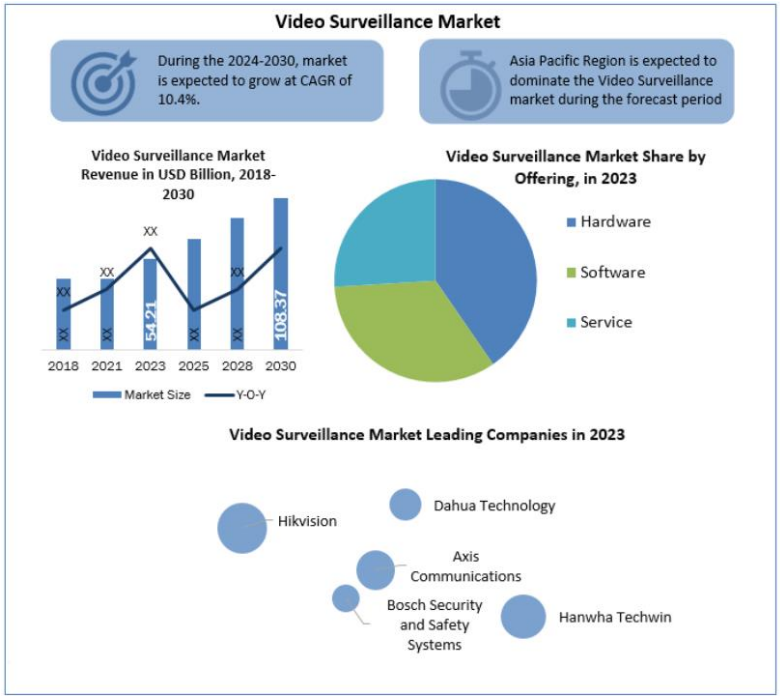


Рис.1.2 Ринок відеоспостереження

У цьому дослідженні аналізуються драйвери зростання ринку, а також сегменти ринку (пропозиція, система, вертикаль і регіон). Учасники ринку, регіони та особливі вимоги надали дані. Це дослідження ринку дає поглиблений погляд на всі важливі досягнення, які зараз відбуваються в усіх галузях промисловості. Статистика, інфографіка та презентації використовуються для аналізу ключових даних. Аналіз розглядає чинники ринку, обмеження, можливості та виклики для

ринку відеоспостереження. Дослідження допомагає оцінити рушії зростання ринку та визначити, як використовувати ці рушії як інструмент. Дослідження також допомагає у виправленні та вирішенні труднощів на світовому ринку відеоспостереження.

Динаміка ринку відеоспостереження. Постійно мінлива апаратна динаміка відеокамер є основною рушійною силою ринку відеоспостереження. Камери безпеки з роздільною здатністю 4K мають набагато чіткіший огляд, що означає більший рівень відновлення даних. Сучасні фотоапарати також мають об'єктиви з регульованим масштабуванням, що дозволяє звужувати поле зору. Зараз більшість камер мають вбудований мікрофон, який може передавати та записувати звук під час запису відео. Компанії також впровадили низку інновацій у апаратному забезпеченні, щоб зменшити ефективність відеокамер. Внутрішнє обігрів доступний на деяких камерах для використання в холодну погоду, тоді як на інших є детектори руху, щоб зафіксувати об'єкт у полі зору камери.

Занепокоєння конфіденційністю навколо відеоспостереження є основним стримуючим фактором на ринку відеоспостереження. Хоча системи відеоспостереження допомагають запобігти крадіжкам і забезпечити базову безпеку, вони часто розглядаються як порушення конфіденційності. Люди хочуть, щоб їхні приватні дані використовувалися виключно для певних і законних цілей, і є занепокоєння щодо того, як відеоматеріал може бути використаний або зловживань, особливо у сфері біометричної ідентифікації. Більше провінцій і територій запроваджують суворіші правила конфіденційності, які регулюють використання технологій відеоспостереження, що змушує фірми брати на себе більшу відповідальність за дані, які вони збирають. У разі віддаленого чи хмарного зберігання відеоматеріал може бути доступний або використаний неавторизованими людьми, зламаний або конфіскований, створюючи більший ризик.

Ключовою можливістю на ринку відеоспостереження є збільшення кількості тепловізійних камер після спалаху коронавірусу. Системи біометричних сканерів, тепловізори та алгоритми відстеження масок отримали безпрецедентне зростання

внаслідок пандемії COVID-19. Теплові камери, які широко використовуються в звичайних місцях, таких як аеропорти, можуть стати все більш поширеними та важливими в найближчі роки. Коли школи поступово відновлять роботу, заплановано встановлення тепловізійних камер. Через тривалу тривалість пандемії прогнозується значне збільшення використання тепловізійних камер у медичному секторі. Крім цих областей, високий попит на державні установи, муніципальні будівлі, заводи, розподільчі центри та приватні корпорації.

Аналіз сегменту ринку відеоспостереження. Згідно з Vertical, комерційний сегмент домінував на ринку відеоспостереження у 2023 році. Очікується, що середньорічний темп зростання буде 11,3% у вищезгаданий прогнозований період. Сегмент комерційного відеоспостереження обумовлений зростаючою актуальністю систем відеоспостереження внаслідок посилення проблем безпеки в різних секторах, наприклад у торгівлі, корпораціях, фінансових установах і банківських установах. Через порушення безпеки, такі як втрата активів, крадіжка зі зломом, незаконний доступ та інша незаконна поведінка, у бізнес-секторі спостерігався сплеск продуктів безпеки. Останнім часом потреба в системах відеоспостереження зросла через швидке зростання малих підприємств, а також закладів роздрібної торгівлі та супермаркетів.

1.3 Застосування штучного інтелекту в системах відеоспостереження

Інтеграція виявлення на основі морфології в камери відеоспостереження зі штучним інтелектом має вирішальне значення для мінімізації помилкових тривог, викликаних невідповідними об’єктами, такими як тварини, що тиняються на території, або природні елементи навколишнього середовища. Ця технологія гарантує, що сповіщення генеруються виключно для виявлення цікавих, наприклад людей і транспортних засобів.

Потенціал штучного інтелекту в камерах відеоспостереження:

- виявлення та розпізнавання обличчя: Розпізнавання обличчя можна використовувати в різних програмах. Це дозволяє запускати тривоги при виявленні

підозрілого обличчя, що зберігається в базі даних. Його також можна використовувати для виконання таких завдань, як контроль робочого часу та відвідуваності, що полегшує моніторинг робочого часу персоналу у визначеній робочій зоні.

- підрахунок людей : у багатьох магазинах важливо знати, скільки людей увійшло та вийшло з приміщення. Крім того, камери підрахунку людей пропонують статистичні звіти, які допоможуть бізнесу з маркетинговими та/або рекламними акціями.

- метадані відео. Будь-який судово-медичний пошук для ідентифікації особи чи транспортного засобу у разі правопорушення потребує часу та віддaності, щоб знайти підозрюваного. Метадані відео допомагають надати додаткову інформацію під час пошуку. Наприклад, його можна відфільтрувати, щоб ідентифікувати особу в синій кепці, рюкзаку та синьому одязі, яка нібито пограбувала магазин у торговому центрі.

- розпізнавання номерних знаків (LPR): камери з розпізнаванням номерних знаків можуть використовуватися для контролю на входах і виїздах, щоб, якщо потрібно, активувати шлагбаум доступу. Подібним чином камери LPR використовуються для контролю дорожнього руху в розумних містах, щоб співпрацювати з безпекою дорожнього руху або контролювати рух транспорту за допомогою періодичних звітів.

- виявлення людей і транспортних засобів: на сьогоднішній день оснастити установку штучним інтелектом нескладно. Виявлення людей і транспортних засобів додає додаткову безпеку установці, оскільки вони допомагають надіслати тривогу клієнту в разі вторгнення в обмежену або приватну зону.

Функціонування системи навчання камери ШІ. Штучний інтелект, який присутній у системі відеоспостереження, може бути розміщений у камері або в записуючому пристрої. Більшість систем, обладнаних штучним інтелектом, містять базу даних або навчені бібліотеки з багатьма зразками та часом. Вони були попередньо схвалені, щоб камера могла порівнювати та бути впевненою, що те, що вона бачить у цей момент, є ціллю відповідного розміру та морфології з бази даних.

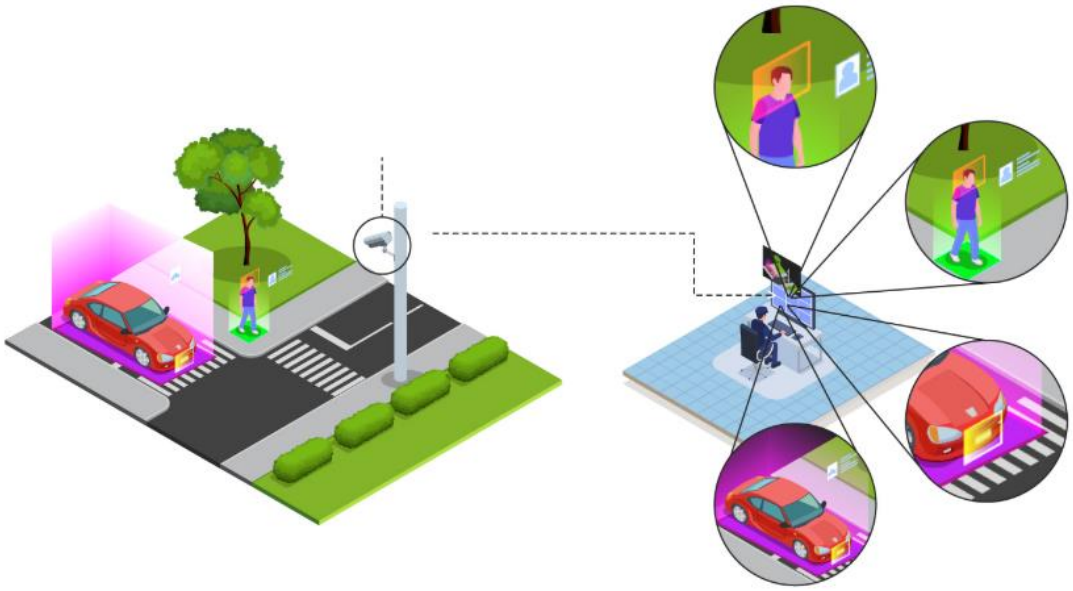


Рис.1.3 Приклад застосування ШІ у відеоспостереженні

Користь від застосування ШІ. Ця функція дійсна для різноманітних сценаріїв у приміщенні та на відкритому повітрі:

- склади/фабрики: за допомогою системи виявлення вторгнень ви можете контролювати простір, щоб генерувати сигнали тривоги з людьми, тож ви можете керувати контролем доступу людей за зонами.

- парки/пішохідні вулиці: на перетинах ліній ви можете налаштувати SIP для створення сповіщення, коли він виявляє транспортний засіб (автомобілі, мотоцикли...) у парку чи пішохідній зоні.

- вілли або окремі будинки: за допомогою перетину лінії та виявлення вторгнень ви можете стежити за зоною по периметру, щоб виявити людей, які стрибають, людей у саду або транспортні засоби, які заїжджають у гараж.

- громадські та приватні автостоянки : за допомогою системи виявлення вторгнень ви можете контролювати під’їзди транспортних засобів, генеруючи подію та тривогу, коли виявлено присутність пішоходів.

Рис.1.4 Технологія метаданих Safire Smart: штучний інтелект у PTZ-камерах

Виявлення на основі морфології, включене в PTZ-камери, має основне значення для зменшення помилкових тривог, створених об'єктами, які не становлять інтересу (тварини, що блукають навколо, елементи природного середовища тощо), надсилаючи сповіщення лише про виявлення людей і транспортних засобів:

1. Автоматичне супроводження: автоматичне супроводження цілі під час її проходження по приміщенню. На відміну від фіксованих камер, ці пристрої PTZ пропонують можливості руху: зокрема, інтелектуальне автоматичне відстеження є переконливою функцією, яка дозволяє не втратити положення виявленої цілі в обмежених зонах. Це автоматичне відстеження, яке починається, коли камери виявляють людину або транспортний засіб у цікавій зоні та відстежують його під час руху через інсталяцію, поки він не зникне з місця події.

2. Активне стримування: камери, що видають попередньо записане повідомлення, і спалах стримування. Нові камери Safire Smart PTZ оснащені функцією активного стримування, яка може транслювати попередньо записане повідомлення та спалахувати стримуючим спалахом, коли присутній зловмисник. Цю функцію можна запустити з власної смарт-події або з інтелектуальних подій з інших камер Safire Smart, які співіснують на тому самому запису (наприклад, якщо фіксована камера виявляє людину, вона може ініціювати «подію» для PTZ-камери, щоб поверніть і активуйте його динамік і відлякувальне світло).

3. Інфрачервоне підсвічування на великій відстані: дальність ІЧ-лазера до 500 м. PTZ-камера з лазерним ІЧ-випромінюванням до 500 м забезпечує покращене нічне бачення (лазерне ІЧ-випромінювання забезпечує інтенсивне освітлення в умовах слабкого освітлення або в повній темряві) і розширений радіус дії (ІЧ-лазери мають значно більший радіус дії порівняно зі звичайними інфрачервоними світлодіодами).

Функція SIP (Smart Intrusion Prevention), штучний інтелект Uniview.

Smart Intrusion Prevention (SIP) від Uniview — це технологія захисту периметра зі штучним інтелектом. Це більш професійне та вдосконалене рішення безпеки, ніж інші системи, такі як звичайне виявлення руху.

Uniview має функцію Smart Intrusion Prevention (SIP) на камері серії Prime. Ця технологія ШІ дозволяє налаштовувати 4 типи інтелектуальних функцій: виявлення перетину лінії; виявлення вторгнень; вхідна зона; зона виходу.

Технологія HIKVISION AcuSense. Інноваційна технологія AcuSense базується на розширеному алгоритмі глибокого навчання, розробленому командою досліджень і розробок Hikvision.

Технологія AcuSense вирізняється своєю здатністю ідентифікувати цілі, точно відрізняючи людей і транспортні засоби від інших рухомих елементів (домашніх тварин і погодних умов, таких як дощ).

AcuSense запускає тривогу лише тоді, коли відбувається заздалегідь визначений тип вторгнення (люди або транспортні засоби). Завдяки точній класифікації цілей система може відфільтрувати до 90% подій помилкової тривоги, досягаючи значного підвищення ефективності та результативності системи: фільтрація помилкової тривоги; пробне світло та настроювана звукова сигналізація; інтелектуальний пошук вмісту.

Теплові камери зі штучним інтелектом. Sunell AI застосовується до теплового зору.Лісові пожежі — це величезне стихійне лихо з руйнівними наслідками. Щоб забезпечити раннє попередження та запобігти пожежі, продукти Sunell активно використовують технологію штучного інтелекту.

Завдяки аналізу штучного інтелекту та алгоритмам, заснованим на глибокому навчанні, тепловізори Sunell можуть виявляти температурні аномалії в ключових областях і швидко активувати сповіщення для запобігання пожежам. У пожежонебезпечних регіонах тепловізори використовують інтелектуальні алгоритми виявлення пожежної точки для виявлення пожежі на ранніх стадіях.

1.4 Порівняння традиційного відеоспостереження з штучних інтелектом

Розвиток відеоспостереження кардинально змінив підхід до безпеки. Раніше камери просто фіксували події, а перегляд відзнятого матеріалу вимагав багато часу та людських ресурсів. З появою штучного інтелекту (ШІ) системи не лише записують, але й аналізують, прогнозують та реагують на події в реальному часі.

Традиційне відеоспостереження.

Витоки: перші системи з’явилися у 1940-х роках і поступово еволюціонували: від аналогових камер до цифрових систем із збереженням даних.

Переваги: допомагає запобігати злочинам, забезпечує відеодокази, спрощує моніторинг.

Недоліки: обмежена аналітика, залежність від людського фактору, низька якість зображень у старих системах, необхідність великих сховищ для записів.

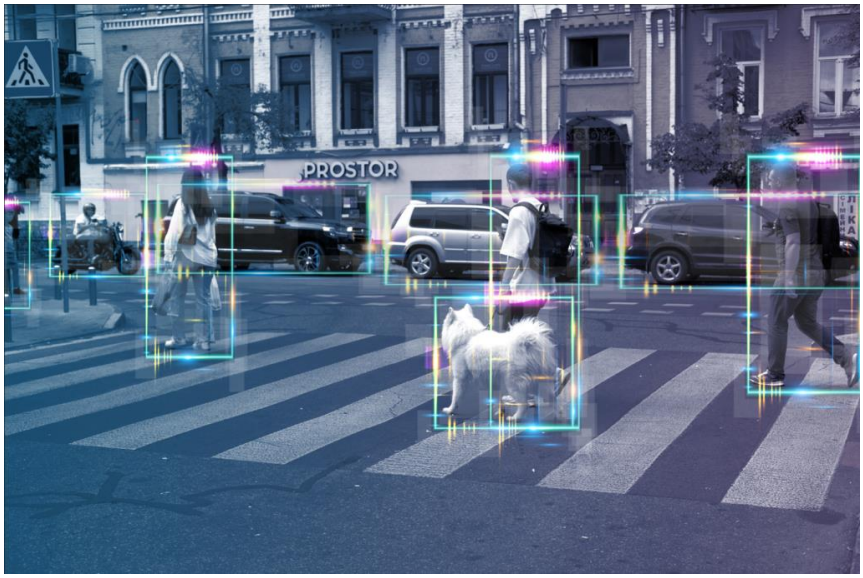


Рис.1.5 Застосування ШІ в розпізнанні об’єктів

Системи відеоспостереження з ШІ.

Можливості: автоматичне виявлення загроз; розпізнавання облич і об'єктів; аналіз поведінки та прогнозування подій.

Переваги: постійний моніторинг без втоми; швидка реакція на інциденти; інтеграція з іншими системами, такими як IoT; зберігання в хмарі та ефективни пошук даних; економія на персоналі та більша точність.

Рис.1.6 Контроль із використанням ШІ у відеоспостереженні

Недоліки: високі початкові витрати, питання конфіденційності.

Порівняння традиційних та ШІ-систем

Моніторинг: традиційний — залежний від людей, ШІ — працює 24/7.

Зберігання: традиційний — локальний, ШІ — хмарний із автоматизацією.

Точність: традиційний — людські помилки, ШІ — висока аналітика без емоцій.

Реакція: традиційний — затримки через людський фактор, ШІ — миттєві сповіщення.

Інтеграція: традиційний — обмежена, ШІ — гнучка та масштабована.

ШІ використовує глибоке навчання, нейронні мережі та алгоритми для аналізу великих обсягів даних. Камери стали більш точними, адаптивними та здатними працювати у складних умовах.

Впровадження штучного інтелекту у відеоспостереження зробило системи більш ефективними, розумними та надійними. Вони вже змінюють правила гри у сфері безпеки, підвищуючи стандарти захисту в домівках, бізнесі та громадських просторах.

1.5 Варіанти реалізації відеоспостереження з використанням штучного інтелекту

Відеоспостереження зі штучним інтелектом може надати численні переваги, зокрема спрощені операції, легші дослідження та швидке виявлення аномалій. Однак отримання цих переваг значною мірою залежить від типу системи, яку ви впроваджуєте. В епоху, коли безпека та ефективність бажані та очікувані, відеоспостереження AI стає важливим інструментом на сучасних підприємствах. За даними IBM Global Adoption Index, більше половини провідних організацій шукають шляхи впровадження ШІ. Це може варіюватися від інструментів управління проектами до платформ аналізу продажів і навіть безпеки відео. Однак більшість сучасних систем відеобезпеки повільно адаптуються або не можуть ефективно інтегрувати штучний інтелект у свої продукти. Штучний інтелект змінює наш спосіб роботи, а переваги, які він може надати в системі безпеки, ще майже не розкриті. Розуміння переходу до відеоспостереження зі штучним інтелектом Багато організацій стурбовані покращенням видимості, безпеки та проактивності свого відеозахисту. Технологія без цих якостей може викликати у вас відчуття дискомфорту, розчарування та стресу. Відеоспостереження зі штучним інтелектом має на меті вирішити цю проблему шляхом покращення всіх аспектів традиційної безпеки, включаючи моніторинг, управління та реагування. Ми бачили, як захист відео еволюціонував від DVR і NVR до хмари та гібридної хмари — тепер ми спостерігаємо перехід до справжнього захисту відео зі штучним інтелектом. AI дає змогу компаніям досягати більшого з меншими ресурсами, що спонукає багато організацій шукати способи інтегрувати AI у свій технологічний пакет. Штучний інтелект не тільки підвищує ефективність відеозахисту, але також може допомогти

інтерпретувати відео для вирішення інцидентів до їх ескалації. Коли трапляються інциденти, штучний інтелект може зрозуміти, що відбувається за допомогою сприйняття, подібного до людини, і викликати сповіщення в режимі реального часу для негайного реагування. Більшість постачальників відеозахисту пропонують певний рівень штучного інтелекту, але ці можливості дуже відрізняються в різних продуктах і часто обмежуються базовим розпізнаванням обличчя та номерних знаків або деякими варіантами пошуку зовнішнього вигляду. Існує безліч різних варіантів безпеки штучного інтелекту, але вибір потрібного значною мірою залежить від ваших потреб і вимог.

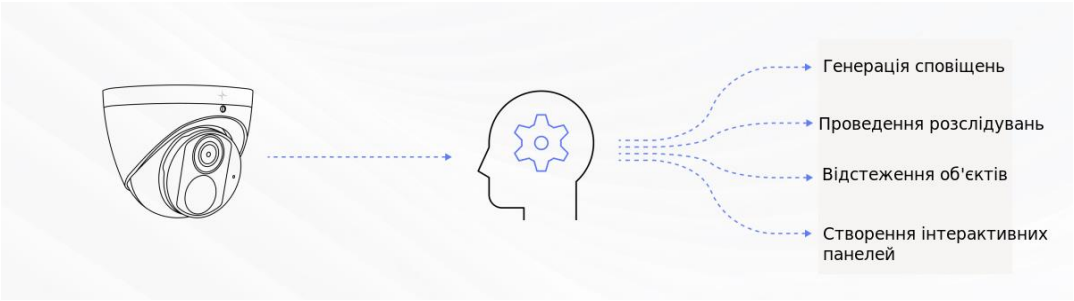


Рис.1.7 Функціонування системи спостереження

Як різні галузі можуть використовувати відеоспостереження AI Відеозахист історично використовувався виключно для запису відео. Навіть з останнім технологічним прогресом відеозахист все ще використовується таким чином — як інструмент відеозапису, який є реактивним і використовується лише після того, як подія вже сталася. Впроваджуючи штучний інтелект, організації можуть перетворити відеобезпеку на проактивне рішення для виявлення аномальних подій і відкриття нових можливостей. Відеоспостереження штучного інтелекту забезпечує індивідуальні переваги перед традиційними системами, щоб відобразити унікальні потреби різних галузей. Нижче наведено кілька практичних прикладів того, як організації можуть використовувати захист відео ШІ. Виробництво: відеоспостереження штучного інтелекту можна використовувати для моніторингу складальних ліній, забезпечення дотримання протоколів безпеки та виявлення неавторизованого персоналу або небезпечної поведінки для створення

безпечніших і продуктивніших робочих середовищ. Школи: відеоспостереження зі штучним інтелектом можна використовувати для захисту периметра кампусу та заборонених територій, а також для попередження в режимі реального часу шкільного персоналу безпеки, коли виявлено підозрілі обличчя, транспортні засоби чи поведінку. Комерційна нерухомість: відеоспостереження зі штучним інтелектом не тільки забезпечує покращену видимість, але також може використовуватися для надання інформації в режимі реального часу про пішохідний рух, зайнятість, вандалізм, бездіяльність, крадіжку та фальсифікацію будівельного обладнання. Як запровадити відеоспостереження AI Вибір правильного методу відеоспостереження зі штучним інтелектом має вирішальне значення для вашого успіху. Вибір неправильного методу може виснажити ресурси, призвести до низького рівня адаптації або забезпечити погану взаємодію з користувачем. Організації, які хочуть запровадити захист відео зі штучним інтелектом, зазвичай мають чотири шляхи: штучний інтелект у камері, аналітичні додатки штучного інтелекту, безпека відео в хмарі або уніфікована безпека відео зі штучним інтелектом. ШІ в камері Багато провідних брендів камер використовують технологію штучного інтелекту. Цей підхід часто є найбільш рентабельним способом почати використовувати ШІ. Однак вбудовані можливості ШІ зазвичай обмежені за обсягом і точністю. Це обмеження насамперед пов’язано з обмеженими ресурсами ШІ в цих пристроях і використанням загальних моделей, які рідко отримують оновлення. Аналітичні додатки ШІ Аналітичні додатки штучного інтелекту доповнюють наявну систему відеобезпеки. За допомогою цієї опції ви можете зберегти наявну систему камер і за потреби додати додаткові можливості AI. Аналітичні надбудови відрізняються за простотою впровадження та зазвичай вимагають додаткового апаратного чи програмного забезпечення, що призводить до вищих витрат. Крім того, цей метод часто забезпечує погану взаємодію з користувачем, оскільки він працює разом із вашою основною системою керування відео (VMS) і не забезпечує ефективного керування всіма наскрізними можливостями, які потрібні організаціям. Cloud Video Security Наступний варіант — нещодавній прогрес у сфері безпеки відео. Хмарні системи відеобезпеки мають

сучасніші функції та деякі можливості ШІ для спрощення керування безпекою порівняно зі застарілими локальними системами. Хоча системна архітектура є більш сучасною, технологія AI обмежена через вартість обробки та пропускну здатність, необхідну для надсилання відео в хмару. Крім того, через обмеження пропускної здатності хмарна аналітика AI забезпечує нижчу роздільну здатність і частоту кадрів за секунду (FPS), що призводить до нижчої точності виявлення. Хмарні системи відеозахисту зосереджені насамперед на забезпеченні кращого досвіду відеозахисту, не обов'язково розумнішого, який використовує вдосконалені моделі штучного інтелекту. Останній варіант, уніфікований захист відео зі штучним інтелектом, забезпечує комплексний досвід, який плавно інтегрує безпеку відео з розширеним штучним інтелектом. На відміну від трьох інших варіантів, в основі цього рішення лежить штучний інтелект. Потім повна система відеозахисту будується на основі штучного інтелекту, щоб забезпечити оптимальну роботу кінцевого користувача. Уніфікована відеозахист ШІ зазвичай не залежить від камери, тобто може працювати з будь-яким пристроєм. Він також має гібридну хмарну архітектуру, яка забезпечує сучасні хмарні функції зі значною обчислювальною потужністю для запуску складних моделей штучного інтелекту з високим FPS і роздільною здатністю. Уніфікована безпека відео зі штучним інтелектом — це найкращий спосіб об'єднати аналітичних постачальників штучного інтелекту та хмарні системи відеобезпеки без потенційних недоліків. Вибір зрештою залежить від потреб вашої організації, її готовності та довгострокового бачення безпеки. Якщо відеоспостереження штучного інтелекту є частиною цього бачення, то прийняття уніфікованого рішення забезпечить успіх як зараз, так і в майбутньому.

Основні міркування щодо ефективної стратегії відеоспостереження зі штучним інтелектом Розглянемо, як постачальник вирішує та обробляє наступні моменти під час впровадження рішення для відеоспостереження штучного інтелекту. Майбутня перевірка Штучний інтелект швидко розвивається, щодня з'являються нові та вражаючі програми. Вибір рішення штучного інтелекту, яке є динамічним і постійно розвивається, а не стоїть на місці, має вирішальне значення.

Виберіть рішення, яке постійно пропонує найсучасніші можливості та регулярні оновлення системи, гарантуючи, що воно завжди залишається сучасним і актуальним. Технічна сумісність Оцініть, наскільки ваша існуюча інфраструктура сумісна з вимогами ШІ. Не всі організації готові до повної перебудови системи; деякі, можливо, захочуть не поспішати на перехід. Перш ніж досліджувати життєздатні рішення, завжди краще запитати, чи ви вирішуєте для одного конкретного випадку використання, чи плануєте запровадити штучний інтелект у широкому масштабі, щоб змінити спосіб керування безпекою, безпекою та операціями. Обробка даних і конфіденційність Правильне рішення для вашої організації має завжди поводитися з вашими даними відповідально та з урахуванням конфіденційності. Шифрування корпоративного рівня, тести пера, сумісність із SOC 2, MFA та автоматичні оновлення мікропрограми мають бути стандартними для будь-якого рішення, яке ви шукаєте. Наслідки вартості Переконайтеся, що ви розумієте такі фінансові нюанси, як початкові витрати, загальна вартість володіння та потенційна економія завдяки підвищенню ефективності або запобіганню інцидентам. Багато організацій дивляться лише на початкові витрати, але при оцінці рішень штучного інтелекту важливо враховувати нематеріальні вигоди, які ваша організація отримає від впровадження штучного інтелекту. Ці переваги можуть включати запобігання інцидентам безпеки, економію грошей шляхом негайного реагування на подію або економію часу на загальне обслуговування системи. Врахування цих деталей забезпечить плавне й успішне впровадження передових технологій захисту відео зі штучним інтелектом. Як ШІ змінює виявлення аномалій і фізичну безпеку Зі стрімкою цифровізацією, зміною ризиків і зростаючою складністю операційного управління традиційних методів відеозахисту вже недостатньо. Використовуючи захист відео зі штучним інтелектом, організації можуть налаштовувати спеціальні тригери сповіщень, визначати, який тип події є аномальним або небажаним, і отримувати негайні сповіщення, коли інциденти потребують уваги. Ці тригери можна розширити для вирішення унікальних бізнес-проблем, надаючи нові та інноваційні можливості.

РОЗДІЛ 2 АНАЛІЗ АРХІТЕКТУРИ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ ТА ОСОБЛИВОСТІ ЇХ ЗАСТОСУВАННЯ

Підходи до глибокого навчання також виявилися придатними для аналізу великих даних із успішним застосуфванням до комп'ютерного зору, розпізнавання образів, розпізнавання мови, обробки природної мови та систем рекомендацій. У даному розділі досліджуються архітектури глибокого навчання та їх практичне застосування. Надається актуальний огляд чотирьох архітектур глибокого навчання, а саме автокодер, згорточна нейронна мережа, мережа глибоких переконань і обмежена машина Больцмана. Оглядаються різні типи глибоких нейронних мереж і підсумовуються останні досягнення. Виділено застосування методів глибокого навчання в деяких вибраних сферах (розпізнавання мови, розпізнавання образів і комп'ютерне бачення).

2.1 Архітектури глибокого навчання із застосуванням обмежених машин Больцмана

У цій частині наведено короткий огляд RBMs. RBM широко використовуються в мережах глибокого навчання через їхню історичну важливість і відносну простоту. RBM вперше був запропонований як концепція Смоленським і став відомим після того, як Хінтон опублікував свою роботу [15]. RBM використовувалися для створення стохастичних моделей ШНМ, які можуть вивчати розподіл ймовірностей щодо їхніх вхідних даних. RBM складаються з варіантів машин Больцмана (BM). BM можна інтерпретувати як НМ із стохастичними процесорами, підключеними двонаправлено. Оскільки важко вивчити аспекти невідомого розподілу ймовірностей, було запропоновано RBM для спрощення топології мережі та підвищення ефективності моделі. Добре відомо, що RBM є особливим типом марковських випадкових полів зі стохастичними видимими одиницями в одному шарі та стохастичними спостережуваними одиницями в іншому шарі.

Структура та алгоритм. Як показано на рис.2.1, нейрони обмежені для формування дводольного графа в RBM.

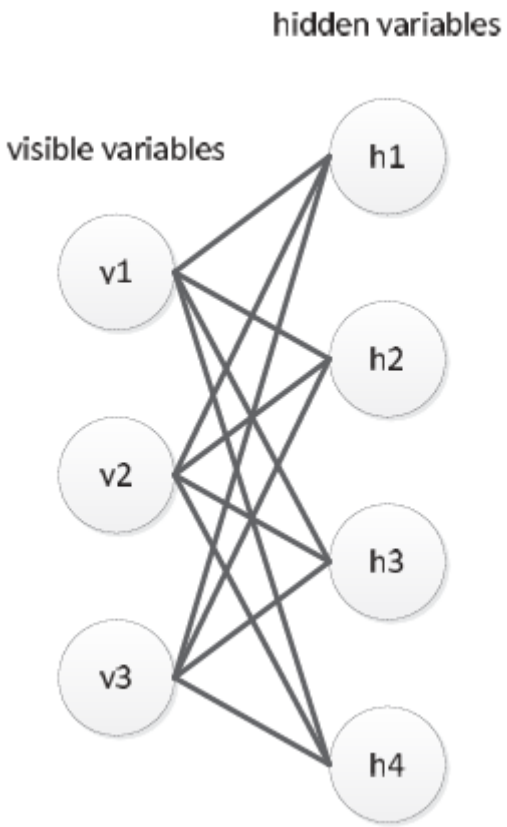


Рис. 2.1 Принципова схема RBMs

Можна побачити, що існує повний зв’язок між видимими та прихованими одиницями, тоді як між одиницями з одного шару зв’язку немає [165]. Для навчання RBM використовується семплер Гіббса. Починаючи з випадкового стану в одному шарі та виконуючи вибірку Гіббса, ми можемо генерувати дані з RBM. Після того, як стану одиниць на одному шарі задано, усі одиниці на інших рівнях будуть оновлені. Цей процес оновлення триватиме, доки не буде досягнуто рівноважний розподіл. Далі ваги в рамках RBM отримують шляхом максимізації ймовірності цього RBM. Зокрема, беручи градієнт логарифмічної ймовірності даних навчання, ваги можна оновити відповідно до:

$$\frac{\partial \log p(v^0)}{\partial \omega_{ij}} = \langle v_i^0 h_j^0 \rangle - \langle v_i^\infty h_j^\infty \rangle, \tag{2.1}$$

де ω_{ij} представляє вагу між видимою одиницею i та прихованою одиницею j. v0 ih0 j і v∞ (1) і h∞ j є кореляціями, коли видимі та приховані одиниці знаходяться на найнижчому та найвищому шарах відповідно. Детальний доказ можна знайти в [59]. Слід зазначити, що процес навчання буде більш ефективним при використанні алгоритму градієнтної контрастної дивергенції (CD). Алгоритм CD для навчання RBM був розроблений Хінтоном [56]. Процедура k-крокового алгоритму CD наведена в Алгоритмі 1, рис.2.2.

Input: RBM(V_1,...,V_m,H_1,...,H_n), training period T, learning rate ϵ
Output: The RBM weight matrix ω, gradient approximation $\Delta\omega_{ij}$, Δa_i and Δb_j for $i = 1, ..., m$, $j = 1, ..., n$
Initialize ω with random values distributed uniformly in $[0, 1]$, $\Delta\omega_{ij} = \Delta a_i = \Delta b_j = 0$ for $i = 1, ..., m$, $j = 1, ..., n$
For ∀ t = 1 : T do,
Sample $v_i^{(t)}$ $P(v_i|h^{(t)})$ when $i = 1, ..., m$
Sample $h_j^{(t+1)}$ $P(h_j|v^{(t)})$ when $j = 1, ..., n$
For $i = 1, ..., m$, $j = 1, ..., n$ do
$\Delta\omega_{ij} = \Delta\omega_{ij} + \epsilon \times [P(H_j = 1|v^{(0)}) \times v_i^{(0)} - P(H_j = 1|v^{(T)}) \times v_i^{(T)}]$
$\Delta a_i = \Delta a_i + \epsilon \times (v_i^{(0)} - v_i^{(T)})$
$\Delta b_j = \Delta b_j + \epsilon \times [P(H_j = 1|v^{(0)}) - P(H_j = 1|v^{(T)})]$
$\omega = \omega + \epsilon \times \Delta\omega_{ij}$
End For
End For

Рис.2.2 Алгоритм 1 k-крокової контрастної дивергенції для RBM

Припускаючи, що різниця між модельним і цільовим розподілом невелика, ми можемо використовувати вибірки, створені ланцюгом Гіббса, щоб наблизити негативний градієнт. В ідеалі зі збільшенням довжини ланцюга його внесок у ймовірність зменшується і прагне до нуля. Однак у [14] ми можемо виявити, що оцінка градієнта не може представляти сам градієнт. Крім того, більшість компонент CD і відповідний логарифмічний градієнт правдоподібності мають рівні знаки. Тому в [15] був запропонований більш практичний алгоритм під назвою стійка контрастна дивергенція. У цьому підході автори запропонували відстежувати стани постійних ланцюжків, а не шукати початкове значення ланцюга Маркова

Гіббса на заданому векторі даних. Стани прихованих і видимих одиниць у постійному ланцюжку оновлюються після оновлення кожної ваги. Таким чином, навіть невелика швидкість навчання не спричинить великої різниці між оновленнями та постійними станами ланцюга, забезпечуючи більш точні оцінки.

Варіації RBM. Сьогодні RBM відіграють важливу роль у різноманітних додатках, таких як тематичне моделювання, зменшення розмірності, спільна фільтрація, класифікація та вивчення функцій. Наприклад, RBM можна використовувати для кодування даних, а потім застосовувати до неконтрольованого навчання для регресії або класифікації. Крім того, RBM можна використовувати як генеративну модель. Ми можемо розрахувати спільний розподіл видимих і прихованих одиниць P(v,h) за законом Байєса. Умовну ймовірність окремої одиниці p(h|v) також можна обчислити за допомогою RBM. Тому RBM також можна використовувати як дискримінаційну модель. Як правило, RBM використовуються як екстрактори ознак у процесі попереднього навчання для завдань класифікації. Однак функції, отримані RBM під час неконтрольованого навчання, можуть бути некорисними в процесі навчання під контролем. Крім того, труднощі викличе підбір параметрів, критичних для продуктивності алгоритмів навчання. Щоб вирішити ці проблеми, Larochelle та Bengio запропонували дискримінаційні обмежені машини Больцмана (DRBM). Крім того, для онлайн-навчання з великими наборами даних модель гібридних DBRM (HDRBM) добре працює завдяки своїм сукупним перевагам як генеративності, так і дискримінації. навчання. Однак у задачах класифікації з кількома мітками продуктивність RBM незадовільна. Мніх та ін. [15] запропонували так звані умовно обмежені машини Больцмана (CRBM) для подальшого покращення продуктивності. Тим часом, у довговимірних часових рядах CRBM можна використовувати як нелінійні генеративні моделі. У [13] встановлено неорієнтовану модель з дійсними видимими змінними та двійковими прихованими. У цій моделі на видимі змінні на останніх кількох кроках часу можуть безпосередньо впливати латентні та видимі змінні на кожному кроці часу. З цією властивістю CRBM може ефективніше виконувати онлайновий висновок. Крім того, вивчаючи часові ряди, CRBM можуть отримувати багаті розподілені

представлення, щоб гарантувати ефективність точного висновку. Нещодавно Elfwing розробила самодостатню DRBM (називається FE-RBM) на основі нового алгоритму розрізнювального навчання [10]. У FE-RBM вихід для будь-яких вхідних і класових векторів обчислюється відповідно до негативної вільної енергії RBM. Мета навчання досягається шляхом мінімізації середньоквадратичної помилки навчання за допомогою методу стохастичного градієнтного спуску. Крім того, натхненний попередніми дослідженнями, вільна енергія масштабується константою на основі розміру мережі, щоб покращити надійність апроксимації функції в FE-RBM.

Коли RBM застосовуються до таких областей, як розпізнавання зображень і мовлення, їх продуктивність може бути серйозно погіршена через шуми в даних. У 2012 році Tang et al. [15] представили сучасну модель, надійну машину Больцмана (RoBM), яка може бути використана для роботи з шумами та оклюзіями у візуальному розпізнаванні. За допомогою RoBM можна досягти кращого узагальнення шляхом усунення впливу пошкоджених пікселів. Навчена на немаркованих даних із шумами за допомогою неконтрольованих алгоритмів нахилу, модель RoBM також може вивчати просторову структуру оклюдерів. Порівняно з традиційними алгоритмами, RoBM показали підвищену продуктивність у різних програмах, таких як малювання зображень і розпізнавання обличчя. Як ключовий фактор у розподілі Больцмана температура вперше враховується в графічній моделі DBN Лі та ін. [9]. Було запропоновано температурні обмежені машини Больцмана (TRBM), де температура виступає як незалежний параметр, який потрібно регулювати. Теоретичний аналіз показує, що температура є ключовим фактором, який контролює вибірковість активації нейронів у прихованих шарах. Доведено, що продуктивність запропонованих TRBM може бути покращена шляхом правильного встановлення параметра чіткості логістичної функції. Оскільки введено додатковий рівень гнучкості, TRBM можуть отримувати більш точні результати. Крім того, дослідження також дає деяке уявлення про RBM з фізичної точки зору, що вказує на те, що може існувати певний зв’язок між температурою та деякими реальними НМ.

2.2 Архітектури глибокого навчання із застосуванням мереж з глибокими переконаннями

Як згадувалося в попередньому розділі, приховані та видимі змінні не є взаємно незалежними [16]. Щоб дослідити залежності між цими змінними, Хінтон побудував DBN, склавши банк RBM. Зокрема, DBN складаються з кількох шарів стохастичних і латентних змінних і можуть розглядатися як особлива форма байєсівської ймовірнісної генеративної моделі. Порівняно з ANN, DBN є більш ефективними, особливо коли застосовуються до проблем із немаркованими даними.

Структура та алгоритм.

Принципова діаграма моделі показана нижче на рис.2.3.

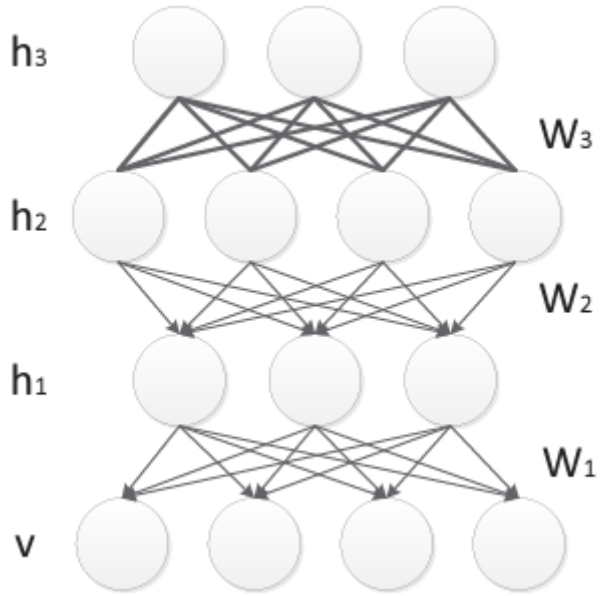


Рис. 2.3 Принципова схема DBN

З рис.2.3. видно, що в DBN кожні два сусідніх шари утворюють RBM. Видимий шар кожного RBM з’єднаний із прихованим шаром попереднього RBM, а два верхніх шари є ненаправленими. Спрямований зв'язок між вищезгаданим шаром і нижнім шаром здійснюється зверху вниз. Різні рівні RBM в DBN навчаються послідовно: спочатку навчаються нижчі RBM, потім вищі. Після того, як функції будуть виділені верхнім RBM, вони будуть поширені назад на нижні

рівні [30]. Порівняно з одним RBM, складена модель збільшить верхню межу логарифмічної правдоподібності, що передбачає більші здібності до навчання [5]. Процес навчання ДБН можна розділити на два етапи: етап попереднього навчання та етап тонкого налаштування. На етапі попереднього навчання проводиться неконтрольоване навчання на основі навчання в напрямку вниз-вгору для виділення ознак; в той час як на етапі тонкої настройки виконується керований алгоритм навчання на основі висхідного-вниз для подальшого коригування параметрів мережі. Ми зауважимо, що покращену продуктивність DBN можна значною мірою пояснити етапом попереднього навчання, на якому початкові ваги мережі вивчаються зі структури вхідних даних. Порівняно з випадково ініціалізованими, ці ваги ближчі до глобального оптимуму і тому можуть забезпечити кращу продуктивність. Алгоритм CD, представлений у попередньому підрозділі, можна використовувати для попереднього навчання DBN. Однак продуктивність зазвичай незадовільна, особливо коли вхідні дані обмежені. Щоб подолати цю проблему, було введено жадібний алгоритм пошарового навчання, який оптимізує ваги DBN за часової складності, лінійної до розміру та глибини мережі [9]. У жадібному пошаровому алгоритмі навчання RBM, які складають DBN, навчаються послідовно. Зокрема, видимий рівень найнижчого RBM навчається спочатку з h(0) як вхідним. Потім значення у видимому шарі імпортуються до прихованих шарів, де обчислюються ймовірності активації P(h|v) прихованих змінних. Представлення, отримане в попередньому RBM, буде використано як навчальні дані для наступного RBM, і цей процес навчання продовжується, доки не будуть пройдені всі шари. Оскільки в цьому алгоритмі апроксимація функції правдоподібності потрібна лише за один крок, час навчання було значно скорочено. Проблему недостатнього оснащення, яка зазвичай виникає в глибоких мережах, також можна подолати в процесі попереднього навчання. Цей алгоритм попереднього навчання також називають алгоритмом жадібного пошарового неконтрольованого навчання. Для наочності ми надали процедуру його реалізації в алгоритмі 2, рис.2.4.

Input: Input visible vector v_{in}, training period T, learning rate ϵ, number of layers J.
Output: Weight matrix ω^i of layer $i, i = 1, 2, ..., J-2$.
Initialize ω with random values from 0 to 1, $h^0 = v_{in}$; where h^0 denotes the value of units in the input layer. h^i represents the units' value of the ith layer. layer = 1;
for $\forall$ t = 1 : T do,
for $layer < L$, do gibbs sampling h^{layer} using $P(h^{layer}|h^{layer-1})$
Computing the CD in Algorithm 1, $\omega(layer)$ is achieved using $\omega(t+1) = \omega(t) + \epsilon \times \Delta\omega$
End for
End for

Рис.2.4 Алгоритм 2, пошаровий алгоритм для DBN

На етапі тонкого налаштування DBN навчаються з позначеними даними за допомогою алгоритму вгору-вниз, який є контрастною версією алгоритму пробудження-сну [17]. Щоб дізнатися межі категорій мережі, на верхньому рівні встановлюється набір міток для процесу вивчення ваг розпізнавання. Крім того, алгоритм зворотного розповсюдження використовується для точного налаштування вагових коефіцієнтів з позначеними даними. Порівняно з оригінальним алгоритмом пробудження та сну, алгоритм «вгору-вниз» не страждає від проблем усереднення режиму, що може призвести до поганої ваги розпізнавання. Підводячи підсумок, процес навчання DBN включає неконтрольовану пошарову процедуру попереднього навчання, що виконується за принципом «знизу вгору», і контрольований процес тонкого налаштування «вгору-вниз». Процес попереднього навчання можна розглядати як вивчення функцій, за допомогою якого можна отримати кращі початкові значення вагових коефіцієнтів, а потім алгоритм «вгору-вниз» використовується для налаштування всієї мережі. Варто зазначити, що за допомогою DBN немарковані дані обробляються ефективно. Крім того, можна уникнути проблем з надмірним і недостатнім оснащенням.

C. Варіації DBN У 2009 році Nair і Hinton [21] представили модель верхнього рівня для DBN і оцінили її в задачі розпізнавання 3D-об'єктів. Машина Больцмана третього порядку використовується як модель верхнього рівня та навчається за допомогою гібридного алгоритму, який поєднує як генеративні, так і дискримінаційні градієнти. Базуючись на роботі Індівері та Лю [15], стверджується, що архітектури процесорів, навіяні мозком, є моделями підтримки DNN і

кортикальних мереж. Крім того, було доведено, що комплементарні апріори можуть бути використані для подолання труднощів висновку в щільно пов’язаних мережах переконань. У 2008 році Салахутдінов і Хінтон [13] представили метод вивчення хорошого ядра коваріації для процесу Гаусса з даними без міток і DBN. У порівнянні зі звичайним ядром, заснованим на необроблених вхідних даних, ядро Гауса працює краще, якщо набори даних мають великі розміри та добре структуровані. Завдяки успішному застосуванню DBN до еталонного тесту TIMIT Acoustic-Phonetic Continuous Speech Corpus, дослідники мотивовані вирішувати набагато складніше завдання – тему великого словникового запасу. Було доведено, що для такого завдання навчання DBN обчислювально складніше. Хоча зворотне розповсюдження стохастичного градієнтного спуску показало свою силу на етапі тонкого налаштування, важко змінити процес навчання, особливо для великомасштабного набору даних. На основі надзвичайно потужного графічного процесора можна навчити глибоку архітектуру для десятків розпізнавачів мовлення, використовуючи велику кількість даних навчання мовлення з чудовими результатами. Однак він не зможе отримати прийнятні результати лише з одним GPU, оскільки поточні архітектури не можуть гарантувати ефективність навчання. Таким чином, Ден і Ю [13] запропонували нову глибоку архітектуру, яку називають глибоко опуклими мережами (DCN), щоб подолати недоліки в масштабованості навчання. DCN складаються з різноманітних багатошарових модулів. Один модуль формується з одного прихованого шару, а також двох наборів ваг у спеціальній нейронній мережі. Точніше, найнижчий модуль складається з двох лінійних шарів і нелінійного шару. Один лінійний шар містить вхідні змінні, а інший містить вихідні змінні. Крім того, нелінійний шар містить нелінійні вхідні змінні. Метод навчання в DCN базується на пакетному режимі, що призводить до паралельного навчання. Крім того, продуктивність DCN можна покращити за допомогою процесу тонкого налаштування структури. Порівняно зі стандартними алгоритмами класифікації, такими як SVM і KNN, DBN також можна використовувати для класифікації зображень завдяки їхній видатній продуктивності в навчанні функцій. Ґрунтуючись на жадібному пошаровому неконтрольованому алгоритмі навчання, Abdel et al. [11]

запропонував автоматичну систему діагностики, яка включає DBN для попереднього навчання та NN зворотного поширення для точного налаштування. Порівняно зі стандартною NN лише з однією контрольованою фазою, система діагностики може досягти вищої точності класифікації. Нещодавно Liao et al. [10] запропонував новий метод пошуку зображень, який базується на DBN і класифікаторі Softmax. Для отримання схожих зображень із бази даних використовується стандартний алгоритм отримання зображень на основі вмісту (CBIR), який використовує автоматизовані методи вилучення ознак. Однак представлення функцій зображення не таке добре, як очікувалося. Показано, що модель DBN-Softmax забезпечує вищу точність і краще запам'ятовування, ніж попередні, такі як алгоритм на основі форми та алгоритм перцептивного хешування. Загалом, базуючись на моделюванні архітектури візуальної системи людини, DBN-Softmax може забезпечити дійсне представлення та вимірювання вилучення більш ефективно, ніж стандартні алгоритми, у яких порогове значення потрібно встановлювати вручну на основі обчислення відстані Хеммінга.

Щоб підвищити гнучкість DBN, була введена нова модель згорткових мереж глибокого переконання (CDBN). Оскільки вхідні дані мають бути векторизовані як матриця зображення, інформацію про двовимірну (2-D) структуру, таку як вхідне зображення, не можна імпортувати як вхідні дані безпосередньо в DBN. Однак у CDBN можна витягнути особливості багатовимірних зображень. Хоча жадібний пошаровий алгоритм відіграє важливу роль у навчанні DBN, багато інших методів глибокого навчання також були досліджені. У роботі Бенгіо [10] заявив, що ми можемо розглядати кожну пару шарів DBN як шумопоглинаючий автокодер (DAE).

2.3 Аналіз архітектури глибокого навчання із застосуванням автокодеру

Автокодер (AE), який є іншим типом ШНМ, також називається автоасоціатором. Це неконтрольований алгоритм навчання, який використовується для ефективного кодування набору даних з метою зменшення розмірності.

Протягом останніх кількох десятиліть АЕ були на передньому краї серед досліджень ШНМ. У своїй роботі Бурлард і Камп [15] виявили, що багатошаровий перцептрон (MLP) у режимі автоасоціації може досягти стиснення даних і зменшення розмірності в таких областях, як обробка інформації. Нещодавно АЕ були використані для вивчення генеративних моделей даних. Вхідні дані спочатку перетворюються на абстрактне представлення, яке потім перетворюється назад у вихідний формат за допомогою функції кодувальника. Точніше, він навчений кодувати вхідні дані в деяке представлення, щоб вхідні дані можна було реконструювати з цього представлення. По суті, АЕ намагається наблизити функцію ідентичності в цьому процесі. Однією з ключових переваг АЕ є те, що ця модель може безперервно отримувати корисні функції під час розповсюдження та фільтрувати непотрібну інформацію. Крім того, оскільки в процесі кодування вхідний вектор перетворюється на представлення нижчої розмірності, ефективність процесу навчання може бути підвищена.

Структура та алгоритм. АЕ — це одношарова нейронна мережа прямого зв’язку, подібна до MLP [13]. Різниця між MLP і АЕ полягає в тому, що метою АЕ є реконструкція вхідних даних, тоді як метою MLP є прогнозування цільових значень за допомогою певних вхідних даних. Кількість вузлів у вхідному та вихідному шарах однакова. У процесі кодування АЕ спочатку перетворює вхідний вектор x у приховане представлення h за допомогою вагової матриці ω; тоді в процесі декодування АЕ повертає h до вихідного формату, щоб отримати ˜ x з іншою ваговою матрицею ω′. Теоретично ω′ має бути транспонуванням ω. Оптимізація параметрів прийнята для того, щоб мінімізувати середню помилку реконструкції між x і ˜ x. Середні квадратичні помилки (MSE) використовуються для вимірювання точності реконструкції відповідно до припущеного розподілу вхідних характеристик. Схематична діаграма моделі показана нижче на рис.2.5.

Введення коду реконструкції Подібно до DBN, процес навчання для АЕ також можна розділити на два етапи: перший етап полягає у вивченні функцій за допомогою неконтрольованого навчання, а другий – для точного налаштування мережі за допомогою контрольованого навчання.

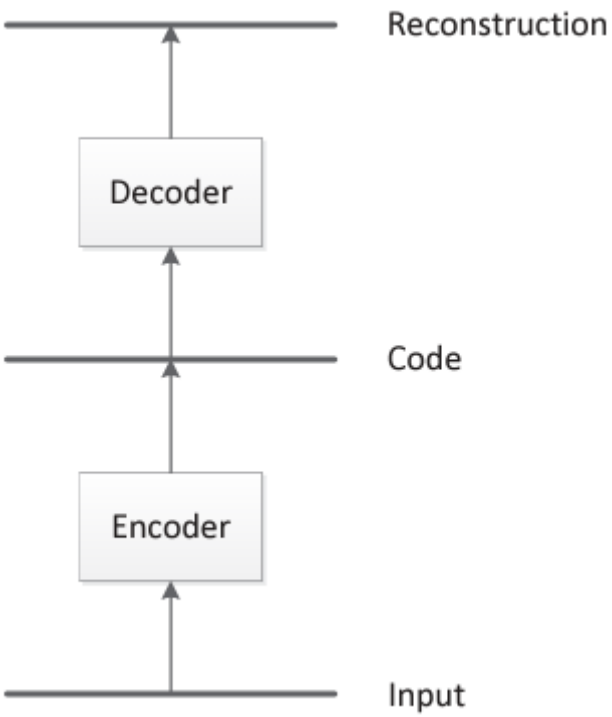


Рис.2.5 Принципова схема AE

Щоб бути конкретнішим, на першому етапі спочатку виконується пряме розповсюдження для кожного входу, щоб отримати вихідне значення ˜ x. Потім квадрат помилки використовується для вимірювання відхилення ˜ x від вхідного значення. Нарешті, помилка буде передано через мережу для оновлення вагових коефіцієнтів. На етапі тонкого налаштування, коли мережа має відповідні функції на кожному рівні, ми можемо прийняти стандартний метод навчання під наглядом і алгоритм градієнтного спуску для налаштування параметрів на кожному рівні.

Варіації AE. У роботі Vincent et al. [14] запропонували DAE для знешумлення традиційних AE. DAE навмисно додає шуми до навчальних даних і навчає AE з цими пошкодженими даними. За допомогою процесу навчання DAE може відновити безшумну версію навчальних даних, що передбачає підвищену надійність. У порівнянні з RBM, деякі стандартні методи оптимізації можуть бути використані в DAE. Слід зазначити, що використовуючи статистичні залежності, притаманні вхідним даних, DAE скасовує несприятливі наслідки зашумлених вхідних даних, спотворених стохастичним чином. Цільова функція для оптимізації в DAE показана у (2.2)

$$J_{DAE} = \sum_{t} IE_{q(\tilde{x}|x^{(t)})}[L(x^{(t)}, g_\theta(f_\theta(\tilde{x})))], \tag{2.2}$$

де IEq(˜ x|x(t))[L(x(t), gθ(fθ(˜ x)))] представляє середнє значення за пошкодженими даними ˜ x, отримане з процедури пошкодження q(˜ x|x(t)). На практиці стохастичний градієнтний спуск використовується для оптимізації цільової функції. Нова архітектура була розроблена в [156] на основі складених шарів DAE. За допомогою стекової моделі реалізація DAE стає легшою, оскільки нам потрібно лише визначити тип і рівень спотворюючого шуму. Нещодавно було помічено, що продуктивність класифікаційних завдань буде покращена, якщо заохочувати розрідженість для вивчення уявлень. Розріджені представлення використовуються для простої інтерпретації вхідних даних шляхом вилучення прихованої структури даних. Алгоритм навчання для розрідженого представлення вперше був запропонований Ранцато в [128]. Щоб налаштувати вектор коду на квазідвійковий розріджений, між лінійним кодером і лінійним декодером додається нелінійна розрідженість. Ми зауважимо, що для двійкових вхідних даних потрібні великі ваги, щоб мінімізувати помилку реконструкції. Загальна функція витрат у розрідженому AE показана у (3.3):

$$J_{sparse}(\omega, b) = J(\omega, b) + \beta \sum_{j=1}^{N} KL(\rho \| \rho'_j) \tag{2.3}$$

де ρ — параметр розрідженості, як правило, невелика величина, близька до нуля, N — кількість нейронів у прихованому шарі, ρ′ j — середня активація прихованої одиниці j, а Jsparse(ω,b) — попередня функція вартості . β контролює вагу терміну штрафу за розрідженість. Крім того, Махзані та Фрей [10] запропонували k-розріджений AE у 2013 році. K-розріджений AE складається з базової архітектури стандартного AE, зберігаючи лише найвищі k активації в прихованих шарах. Отримані результати показують, що k-розріджені AE працюють

краще, ніж DAE та RBM. Вони стверджували, що k-розріджені AE можна легко навчити, а просунутий процес кодування сприятиме досягненню задовільної продуктивності для великомасштабних проблем. У роботі Rifai et al. [133] запропонували скорочувальні автокодери (CAE), де добре вибраний термін штрафу додається до стандартної функції вартості на етапі реконструкції. Цей штрафний термін використовується для покарання за чутливість функцій щодо вхідних даних.

Таким чином, відображення вхідного вектора в представлення буде сходитися з більшою ймовірністю. Результати, отримані за допомогою CAE, ідентичні або навіть кращі за результати, отримані за допомогою інших упорядкованих AE, таких як DAE. Навчальна мета CAE показана (2.4) :

$$J_{CAE} = \sum_{t} L(x^{(t)}, g_\theta(f_\theta(x^{(t)}))) + \lambda \| J(x^{(t)}) \|_F^2, \qquad (2.4)$$

де L(·) є функцією вартості, λ є параметром для контролю сили регуляризації, а J(x) є функцією, яка представляє матрицю Якобі кодера. Rifai та ін. виявив, що термін штрафу створить надійні функції на рівні активації. Крім того, покарання можна використовувати для вирішення компромісу між надійністю та точністю реконструкції. Також показано, що DAE з невеликими шумами, що пошкоджують, можна розглядати як CAE, в якому вся функція реконструкції штрафується [11]. Крім того, у 2016 році Sun та ін. [146] запропонував роздільний глибокий автокодер (SDAE), який використовується для оцінки невидимого шуму. Загальна помилка реконструкції зашумленого мовного спектру може бути мінімізована шляхом коригування невідомих параметрів DAE та оцінки чистого мовного спектру [19].

2.4 Огляд архітектури глибокого навчання на базі згорткової нейронної мережі

CNN є підтипом дискримінаційної глибокої архітектури [3] і показали задовільну продуктивність в обробці двовимірних даних із сітчастою топологією, таких як зображення та відео. Архітектура CNN натхненна організацією зорової кори тварин. У 1960-х роках Hubel і Wisel [73] запропонували концепцію під назвою рецептивні поля. Вони виявили, що складні механізми клітин містяться в зоровій корі тварин, відповідальні за виявлення світла в перекриваються та малих підобластях поля зору. Крім того, обчислювальна модель Neocognitron була представлена в [46] з ієрархічно організованими перетвореннями зображень. Однак Neocognitron відрізняється від CNN тим, що не потребує спільної ваги. Концепція CNN натхненна нейронними мережами із затримкою часу (TDNN). У TDNN ваги розподіляються в часовому вимірі, що призводить до скорочення обчислень. У CNN згортка замінила загальне матричне множення в стандартних NN. Таким чином, кількість ваг зменшується, тим самим зменшуючи складність мережі. Крім того, зображення, як необроблені вхідні дані, можна безпосередньо імпортувати в мережу, таким чином уникаючи процедури вилучення ознак у стандартних алгоритмах навчання. Слід зазначити, що CNN є першою справді успішною архітектурою глибокого навчання завдяки успішному навчанню ієрархічних рівнів. Топологія CNN використовує просторові зв’язки, щоб зменшити кількість параметрів у мережі, а отже, продуктивність покращується за допомогою стандартних алгоритмів зворотного поширення. Ще одна перевага моделі CNN полягає в тому, що вона вимагає мінімальної попередньої обробки. Зі швидким розвитком обчислювальних методів обчислювальні методи, прискорені GPU, використовувалися для більш ефективного навчання CNN. Сьогодні CNN вже успішно застосовуються для розпізнавання рукописного тексту, розпізнавання обличчя, розпізнавання поведінки, розпізнавання мови, систем рекомендацій, класифікації зображень і NLP.

Структура та алгоритм. Три фактори відіграють ключову роль у процесі навчання CNN: розріджена взаємодія, спільне використання параметрів та еквіваріантне представлення [74]. На відміну від традиційних мережевих мереж, де зв’язок між вхідними та вихідними одиницями виводиться шляхом множення

матриці, CNN зменшують обчислювальний тягар завдяки розрідженій взаємодії, де ядра робляться меншими за вхідні дані та використовуються для всього зображення. Основна ідея спільного використання параметрів полягає в тому, що замість того, щоб вивчати окремий набір параметрів у кожному місці, нам потрібно вивчати лише один набір із них, що означає кращу продуктивність CNN. Спільне використання параметрів також наділило CNN привабливою властивістю, яка називається еквіваріантністю, тобто щоразу, коли змінюється вхід, вихід змінюється таким же чином. Отже, для CNN потрібно менше параметрів порівняно з іншими традиційними алгоритмами NN, що призводить до зменшення пам’яті та підвищення ефективності. Компоненти стандартного рівня CNN показано на рис.2.6., а концептуальну схематичну діаграму стандартного CNN показано на рис.2.7.

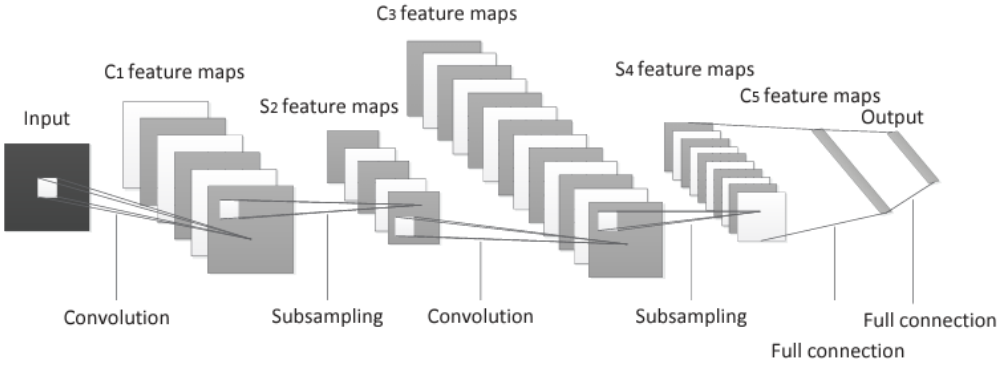


Рис.2.6 Схематична структура CHC

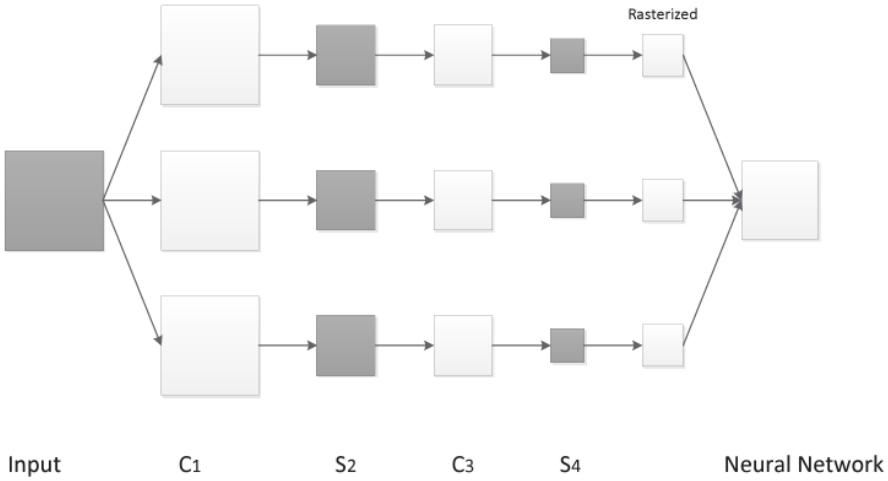


Рис.2.7 Концептуальна структура CNN

Як показано на рис.2.7, CNN — це багаторівнева нейронна мережа, яка складається з двох різних типів шари, тобто шари згортки (c-шари) і шари підвибірки (s-шари). C-рівні та s-рівні з'єднані по черзі і утворюють середню частину мережі. Як показано на рис.2.6, вхідне зображення згортається за допомогою фільтрів, які можна навчити, на всіх можливих зсувах, щоб створити карти функцій у першому c-шарі. Кожен фільтр містить шар з'єднувальних вантажів. Зазвичай чотири пікселі на карті об'єктів утворюють групу. Проходячи через сигмоїдну функцію, ці пікселі створюють додаткові карти функцій у першому s-шарі. Ця процедура продовжується, і таким чином ми можемо отримати карти функцій у наступних c-шарах і s-шарах. Нарешті, значення цих пікселів растеризуються і відображаються в одному векторі як вхідні дані мережі [13].

Як правило, c-рівні використовуються для виділення ознак, коли вхід кожного нейрона пов'язаний із локальним рецептивним полем попереднього шару. Після того, як усі локальні об'єкти вилучені, можна з'ясувати співвідношення між ними. S-шар, по суті, є шаром для відображення функцій. Ці шари відображення функцій мають спільні ваги та утворюють площину. Крім того, для досягнення незмінності масштабу сигмоїдна функція вибирається як функція активації через її незначний вплив на ядро функції. Слід також зазначити, що фільтри в цій моделі використовуються для з'єднання серії сприйнятливих полів, що перекриваються, і перетворення пакетного вхідного двовимірного зображення в єдиний блок на виході. Однак, коли розмірність вхідних даних дорівнює розмірності вихідних даних фільтра, буде важко підтримувати інваріантність перекладу за допомогою додаткових фільтрів. Через високу розмірність застосування класифікатора може спричинити переобладнання. Щоб вирішити цю проблему, запроваджено процес об'єднання, який також називають субдискретизацією або пониженням дискретизації, щоб зменшити загальний розмір сигналу. Насправді субдискретизація вже була успішно застосована для зменшення розміру даних під час стиснення аудіо. У 2-D фільтрі для підвищення незмінності позиції також використовується субдискретизація. Процедура навчання для CNN подібна до процедури для стандартної NN із використанням зворотного поширення. Зокрема,

Lecun et al. [10] представив градієнт помилок для навчання CNN. На першому етапі інформація поширюється в прямому напрямку через різні шари. Основні характеристики досягаються завдяки застосуванню цифрових фільтрів на кожному шарі. Потім обчислюються вихідні значення. На другому етапі обчислюється похибка між очікуваним і фактичним значеннями виходу. Завдяки зворотному поширенню та мінімізації цієї помилки вагова матриця додатково коригується, і мережа, таким чином, точно налаштовується. На відміну від інших стандартних алгоритмів у класифікації зображень, попередня обробка не часто виконується в CNN. Замість встановлення параметрів, як у випадку з традиційними NN, нам просто потрібно навчити фільтри в CNN. Крім того, при виділенні ознак CNN не залежать від попередніх знань і людського втручання. Метод максимального об’єднання був запропонований у LeNets для підвибірки [12]. Узагальнюючи статистику найближчих виходів, функція об’єднання використовується для заміни виходу мережі в певній позиції. Використовуючи метод max-pooling, ми можемо отримати максимальний вихід у прямокутному околі. Процедура об'єднання також може зробити представлення інваріантним до перекладів вхідних даних. Тепер, додавши максимальний шар об’єднання між згортковими шарами, просторова абстрактність збільшується разом із збільшенням абстрактності об’єктів. Як зазначено в [17], об'єднання використовується для отримання інваріантності в перетвореннях зображень. Цей процес призведе до кращої стійкості до шуму. Зазначається, що продуктивність різних методів об’єднання залежить від кількох факторів, таких як роздільна здатність, з якою виділяються низькорівневі ознаки, і зв’язки між потужностями вибірки. У 2011 році Бурео [16] виявив, що навіть якщо об’єкти дуже несхожі, їх можна об’єднати разом, якщо вони розташовані близько. Крім того, виявлено, що кращу продуктивність можна досягти, виконавши кластеризацію перед стадією об’єднання. У [18] показано, що кращої продуктивності об’єднання можна досягти шляхом більш адаптивного навчання рецептивних полів. Зокрема, використовуючи концепцію надмірної повноти, пропонується ефективний алгоритм навчання для прискорення процесу навчання на основі поступового вибору ознак.

Варіації CNN. За останні кілька років CNN став популярною темою дослідження. У роботі Eigen et al. [20] представив нову модель під назвою рекурсивні згорточні мережі (RCN). Архітектуру RCN можна розглядати як CNN з однаковою кількістю карт функцій у всіх шарах і пов’язаними вагами фільтрів між шарами. Показано, що більша кількість шарів передбачає збільшене обчислювальне навантаження, тому немає сенсу точно вказувати розміри карт функцій. CNN також використовувався для виділення ознак у таких сферах, як розпізнавання об’єктів. У роботі Jarrett et al. [21] запропонував нову модель, яка поєднує згортку з AE. Базуючись на архітектурі AE, використовується неконтрольоване навчання ознак із прогнозною розрідженою декомпозицією з обмеженнями розрідженості вектора ознак. Етап виділення ознак включає банк фільтрів, нелінійне перетворення та рівень об’єднання функцій. У процесі навчання кожною згортковою AE використовується звичайний алгоритм градієнтного спуску без додавання додаткових членів регуляризації. Доведено, що складена згортка AE може досягти задовільних ініціалізацій CNN, уникаючи локальних мінімумів сильно невипуклих цільових функцій. Великий успіх був досягнутий, коли CNN застосовувалися для дослідження комп'ютерного зору. У CRBM згортка обчислюється зі звичайним RBM як ядром. Хоча кількість параметрів у RBM залежить від розмірності вхідного зображення, складність CRBM залежить лише від кількості ознак, які потрібно виділити, та розміру сприйнятливого поля. Алгоритм CD також можна застосовувати для навчання CRBM. Видимий шар ініціалізується введенням зображення. Прохід угору виконується для обчислення стану пікселів у прихованому шарі. Порівняно зі стандартними RBM у програмах бачення, CRBM можуть досягти вищої швидкості збіжності з меншим значенням функції негативної правдоподібності. Крім того, також були розроблені згорточні мережі глибоких переконань (CDBN) і застосовані до масштабованого неконтрольованого навчання для ієрархічних представлень і неконтрольованого навчання ознак для аудіокласифікації. Нещодавно швидке перетворення Фур'є (ШПФ) було використано в оригінальних CNN.

2.5 Приклад застосування нейронного навчання

У цьому розділі ми розглянемо деякі практичні застосування архітектур глибокого навчання. Насправді, завдяки своїй здатності обробляти великі обсяги немаркованих даних, методи глибокого навчання надали потужні інструменти для роботи з аналізом великих дани. В останні роки було зібрано величезну кількість даних у різних сферах, включаючи кібербезпеку, медичну інформатику та соціальні медіа. Алгоритми глибокого навчання використовуються для вилучення функцій високого рівня з цих даних, щоб отримати ієрархічні представлення. Останнім часом глибоке навчання привернуло увагу багатьох високотехнологічних підприємств, таких як Google, Facebook і Microsoft. Архітектура глибоких мереж широко застосовувалася в розпізнаванні мови та акустичному моделюванні для класифікації звуку. Крім того, підходи до глибокого навчання також відіграють важливу роль у сфері обробки зображень, наприклад рукописна класифікація [84], класифікація сцен за допомогою дистанційного зондування з високою роздільною здатністю, надроздільна здатність одного зображення (SR) , багатокатегорія швидке послідовне візуальне представлення Brain Computer Interfaces (BCI), адаптація домену redand для великомасштабної класифікації настроїв. Крім того, глибокі архітектури також використовувалися в багатозадачному навчанні для НЛП з підвищеною надійністю логічного висновку. Далі ми зробимо загальний огляд кількох вибраних застосувань глибоких мереж: розпізнавання мови, комп’ютерного зору та розпізнавання образів.

Розпізнавання мовлення. Протягом останніх кількох десятиліть алгоритми машинного навчання широко використовувалися в таких сферах, як автоматичне розпізнавання мови (ASR) і акустичне моделювання. ASR можна розглядати як стандартну задачу класифікації, яка ідентифікує послідовності слів із послідовностей ознак або сигналів мовлення. У деяких чітко визначених програмах, таких як транскрипція та диктування, широко використовуються комерційні засоби розпізнавання мовлення. Щоб забезпечити задовільну продуктивність ASR, необхідно розглянути багато проблем, наприклад, шумне

середовище, розпізнавання кількох моделей і багатомовне розпізнавання. Як правило, перед застосуванням алгоритмів розпізнавання мови дані мають бути попередньо оброблені за допомогою методів видалення шуму. Сінгх та ін. [14] розглянули деякі загальні підходи до видалення шуму та покращення мови, такі як спектральне віднімання, фільтрація Вінера, вікна та оцінка спектральної амплітуди. Традиційні алгоритми машинного навчання, такі як SVM і NN, забезпечили багатообіцяючі результати в розпізнаванні мови. Наприклад, моделі суміші Гауса (GMM) використовувалися для розробки систем розпізнавання мовлення шляхом представлення зв'язку між акустичним входом і прихованими станами прихованої моделі Маркова (HMM) . 1) Стандартна архітектура та алгоритми розпізнавання мовлення: Стандартна архітектура системи ASR наведена на рис.2.8.

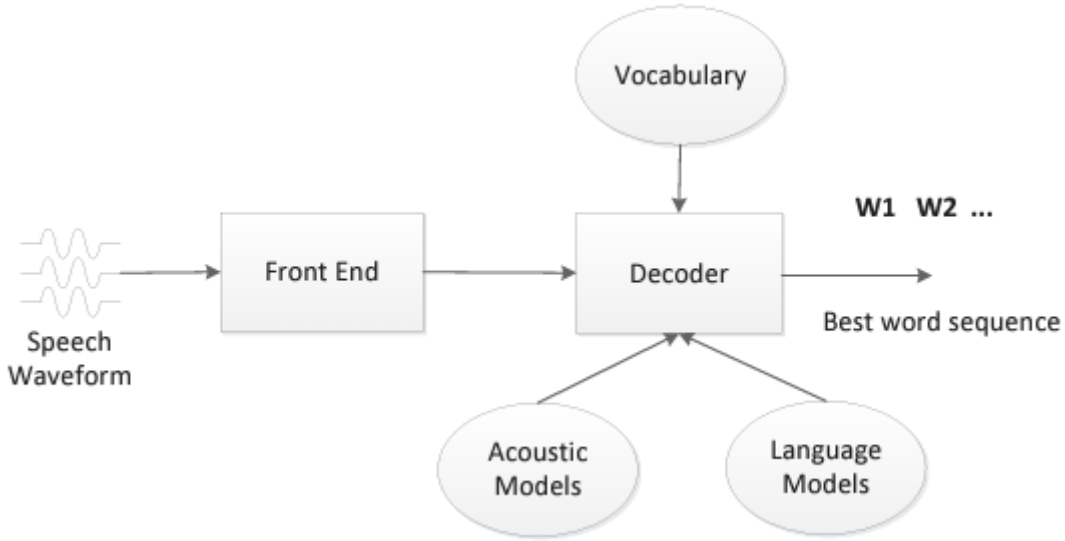


Рис.2.8 Архітектура системи розпізнавання мовлення

По-перше, мовний хвиля проходить через слуховий передній кінець, де сигнал попередньо обробляється та виробляються спектральні характеристики. Потім функції будуть передані до оцінювача ймовірності телефону, щоб оцінити ймовірність кожного телефону. Після цього декодер розкодує мову з імовірностями телефону за допомогою мовної моделі n-gram (LM) і HMM. Нарешті, вихідні дані будуть надіслані до синтаксичного аналізатора, перетворені на найкращу послідовність слів і перетворені у читабельний формат. Як згадувалося раніше,

традиційні схеми машинного навчання досягли задовільних результатів для ASR. Серед них HMM та GMM широко використовуються для акустичного моделювання для генерації низькорівневого акустичного вмісту з високорівневих акустичних вхідних даних. Тут ми коротко розповімо про ці дві моделі. Оскільки мовний сигнал можна розглядати як короткочасний або кусково-стаціонарний сигнал, то протягом короткого періоду часу можна вважати процес мовлення стаціонарним. Таким чином, модель Маркова може описувати стохастичний процес мовлення. Крім того, процес навчання HMM автоматичний і простий у реалізації. Ми можемо використовувати послідовність прихованих станів HMM для представлення акустичних особливостей з нестаціонарним розподілом. HMM генерує послідовність векторів, що представляють імовірність кожного стану. Слід зазначити, що на продуктивність HMM може сильно вплинути невідповідність між умовами навчання та тестування. У такому випадку потрібні великі обсяги даних. У [26] GMM були використані для оцінки вихідної щільності станів HMM. Крім того, GMM відіграють важливу роль у завданнях генерації мовлення та часто використовуються в покадровому відображенні, особливо для покращення мовлення, відображення артикуляції в акустику та перетворення голосу. Системи GMM-HMM значно підвищили точність класифікації і також можуть бути застосовані для видалення шуму в галасливих мовних висловлюваннях. Слід визнати, що GMM-HMM все ще має деякі обмеження. Для GMM-HMM важко представити нелінійні або більш складні зв’язки між акустичними характеристиками та мовними вхідними сигналами. Ефективність моделювання зазвичай дуже низька для даних поблизу нелінійного різноманіття. Крім того, припущення про умовну незалежність є ще одним відомим недоліком GMM. Крім того, втрата необробленої інформації також може погіршити продуктивність систем GMM-HMM.

Загальновідомо, що НМ, розташовані на нелінійному колекторі або поблизу нього, можуть забезпечити кращу продуктивність, ніж системи GMM-HMM. Елегантні результати були досягнуті два десятиліття тому, коли дослідники прийняли ШНМ з одним шаром нелінійних прихованих одиниць для прогнозування

станів HMM за вікнами акустичних коефіцієнтів. Однак через обмежені обчислювальні ресурси в той час було важко реалізувати стандартні NN з багатьма прихованими рівнями. Протягом останніх кількох років технології обчислень швидко розвивалися, що призвело до більш ефективних способів навчання DNN. З 2006 року глибоке навчання стало новою областю досліджень машинного навчання. Як ми щойно згадували, алгоритми глибокого навчання можуть принести задовільні результати у вилученні та перетворенні ознак, і їх успішно застосували для розпізнавання образів. Композиційні моделі створюються за допомогою моделей DNN, де характеристики отримуються з нижчих рівнів. За допомогою деяких добре відомих наборів даних, включаючи великі набори словникових даних, дослідники показали, що DNN можуть досягти кращої продуктивності, ніж GMM, щодо акустичного моделювання для розпізнавання мовлення. Завдяки їхній видатній продуктивності в моделюванні кореляції даних, архітектури глибокого навчання тепер замінюють GMM у розпізнаванні мови.

Найпростішим підходом до визначення співвідношення між акустичним входом і артикуляційним виходом є метод лінійного відображення, такий як GLM. Однак, як згадувалося раніше, відображення звуку та артикуляції є нелінійним, що вказує на те, що GLM не може досягти ідеальної продуктивності. У процесі навчання метод HMM вимагає фонетичної інформації як обмежень для вирішення проблеми відображення. Зв'язок між акустичними та артикуляційними характеристиками розглядається як лінійне відображення в кожному стані HMM. У цьому випадку GMM використовується для моделювання спільного розподілу артикуляційних і акустичних характеристик для вирішення проблеми необмеженого відображення. Як показано в [26], ШНМ можна використовувати для побудови системи розмовних аватарів у режимі реального часу, керованої мовленням, через їх відносно короткий час обчислення. Порівняно з іншими моделями, ШНМ мають чудову продуктивність. Однак навчання ШНМ із кількома прихованими рівнями зазвичай займає багато часу, і процес навчання зазвичай потрапляє в пастку поганого локального оптимуму. Крім того, на продуктивність може значно вплинути спосіб ініціалізації ШНМ. Мотивовані цими фактами, DNN

прийняті для ефективної обробки великої кількості немаркованих даних. За допомогою DNN виконується більш розумна ініціалізація та виконується більш ефективне попереднє навчання, що також певною мірою вирішило проблему переобладнання.

Комп'ютерне бачення та розпізнавання образів. Комп'ютерний зір спрямований на те, щоб змусити комп'ютери точно розуміти та ефективно обробляти візуальні дані, такі як відео та зображення [14]. Машина повинна сприймати реальні високовимірні дані та створювати символічну або числову інформацію відповідно. Кінцева мета комп'ютерного зору полягає в тому, щоб наділити комп'ютери здатністю сприйняття людини. Концептуально комп'ютерний зір відноситься до наукової дисципліни, яка досліджує, як отримати інформацію із зображень у штучних системах. Наступні області включені як суб-області комп'ютерного зору: виявлення подій, реконструкція сцени, виявлення та розпізнавання об'єктів, оцінка положення об'єктів, відновлення зображення, статистичне навчання, редагування зображень та покращення відео. Розпізнавання образів – це наукова дисципліна, метою якої є ідентифікація шаблону заданого вхідного значення [14]. Це досить загальна концепція, яка охоплює кілька піддоменів, таких як класифікація, регресія, маркування послідовності та тегування мови. У зв'язку зі швидким промисловим розвитком постійно зростають вимоги до можливостей пошуку та обробки інформації, що породило нові виклики для розпізнавання образів. Останнім часом розвиток архітектур глибокого навчання забезпечив нові підходи до проблеми розпізнавання образів, які будуть обговорюватися далі.

Визнання: За останні кілька років методи глибокого навчання досягли величезного прогресу в областях комп'ютерного зору та розпізнавання образів, особливо в таких сферах, як розпізнавання об'єктів. Ми обговоримо деякі класичні проблеми комп'ютерного зору щодо завдань розпізнавання. У програмах класифікації вибір ознак є важливим питанням. Зазвичай функції вказуються вручну в традиційних алгоритмах класифікації, які мають обмежену загальність. Деякі типові архітектури глибокого навчання, такі як CNN, можуть автоматично

вибирати функції та досягати видатної продуктивності на основі обчислювальних ресурсів із прискоренням GPU. Зверніть увагу, що системи зору людини відрізняються від систем комп'ютерного зору, і було показано, що DNN можна легко обдурити нерозпізнаними зображеннями [24]. Однак це не означає, що методи глибокого навчання не підходять для класифікаційних завдань. Нещодавні дослідження показали, що в задачах класифікації методи глибокого навчання можуть отримати багатообіцяючі результати. У розпізнаванні об'єктів, яке також називають класифікацією об'єктів, методи глибокого навчання досягли кращої продуктивності порівняно зі звичайними алгоритмами класифікації. Тут ми розглядаємо деякі нещодавні досягнення в класифікаційних завданнях. Для розпізнавання дорожніх знаків Німеччини було запропоновано багатоколонковий DNN. Для вивчення нейропсихіатричних станів на основі моделей функціонального зв'язку (FC) широко використовувалися стандартні класифікатори, такі як SVM. Нещодавно DNN були використані для класифікації патернів FC у стані спокою всього мозку при шизофренії (SZ). Щоб покращити продуктивність класифікації, була запропонована нова мультимодальна глибока нейронна мережа з максимальним запасом (3mDNN), яка використовує переваги кількох локальних дескрипторів зображення. У порівнянні зі стандартними алгоритмами, цей метод, враховуючи інформацію кількох дескрипторів, може досягти дискримінаційної здатності. DNN також можна використовувати для класифікації моделей швидкості вітру та контрольованої багатоспектральної класифікації землекористування. У комп'ютерному зорі та розпізнаванні образів іноді потрібно будувати та обробляти 3D-моделі. Розуміння меша є одним із ключових факторів у цій галузі. Зокрема, маркування сітки можна використовувати, щоб дізнатися властиві характеристики сітки. Слід зазначити, що методи глибокого навчання також можна застосовувати для розпізнавання пози рук (HPR). Оскільки функції, створені традиційними алгоритмами, обмежені, і важко виявити та відстежити руки за допомогою звичайних камер, DNN використовуються для створення розширених функцій.

Виявлення: Виявлення є одним із найвідоміших піддоменів комп'ютерного зору. Він намагається точно визначити місцезнаходження та класифікувати цільові об'єкти на зображенні. У завданнях виявлення зображення сканується, щоб виявити певні особливі проблеми. Наприклад, ми можемо використовувати виявлення зображень, щоб виявити можливі аномальні тканини або клітини на медичних зображеннях. Модель на основі деформованих частин (DPM), запропонована Felzenszwalb, є одним із найпопулярніших методів. Як показано в [18], завдяки їхнім сильним можливостям фіксувати геометричну інформацію, таку як розташування об'єктів, DNN широко використовувалися для виявлення та показали видатну продуктивність. Як зазначено в [20], DBN використовуються в системах автоматизованої діагностики (CAD) для раннього виявлення раку молочної залози. У цьому випадку точність класифікатора є найважливішим фактором для САПР. Порівняно зі стандартними алгоритмами класифікації, такими як метод дерева рішень C4.5, техніка керованої нечіткої кластеризації (SFC), підхід Fuzzy-GA, метод нейронної мережі радіально-базисної функції (RBFNN) і оптимізована вейвлет-нейронна мережа з роєм частинок (PSOWNN), DBN можуть досягти кращої продуктивності системи САПР. DBN також можна використовувати для зменшення нелінійної розмірності вхідних характеристик. Дослідження з виявлення пухлин головного мозку та задачі сегментації приділяють все більшу увагу протягом останніх кількох років. Завдяки своїй обчислювальній ефективності метод магнітно-резонансної томографії (МРТ) для клінічного виявлення пухлин головного мозку було впроваджено за допомогою методів глибокого навчання. Однак метод виявлення пухлин головного мозку на основі МРТ страждає від невідповідності між передбачуваним розміром і формою пухлин головного мозку. CNN використовуються для вирішення цієї проблеми через його потужну здатність до навчання. Слід зазначити, що багатообіцяючі результати були досягнуті CNN у таких сферах, як виявлення людини та класифікація активності на основі доплерівського радара. Подібним чином, методи глибокого навчання також можна застосовувати для анотування генетичних варіантів для ідентифікації патогенних варіантів. Зазвичай комбінований алгоритм виснаження, що залежить від анотації

(CADD) найбільш широко використовується для анотації варіантів кодування та некодування. У методі CADD лінійне ядро SVM навчається як класифікатор. Однак через обмеження SVM метод CADD не може охопити нелінійні зв’язки між функціями. Тому DANN використовуються замість класифікатора SVM. DANN підходять для великої кількості зразків і функцій. Останніми роками моделі виявлення помітності привернули все більшу увагу дослідників у прогнозуванні місць спостереження людського ока в полі зору. Традиційні підходи базуються на механізмах контрастного висновку та ручних функціях. Для методів глибокого навчання потрібні лише необроблені дані зображення. Крім того, методи глибокого навчання також застосовувалися для виявлення глаукоми і систем взаємодії людини з роботом із багатообіцяючими результатами. Як ще одне важливе застосування комп’ютерного зору, виявлення зміни зображення відіграє важливу роль не лише у цивільній, а й у військовій сферах. Метою виявлення зміни зображення є сортування відмінностей між двома зображеннями, зробленими в різний час для однієї сцени. Виявлення зображень широко використовується в дистанційному зондуванні, медичній діагностиці, оцінці катастроф і відеоспостереження. Зокрема, обробка зображень радара із синтетичною апертурою (SAR) є широко використовуваним додатком для виявлення змін. За допомогою найсучасніших методів різницеве зображення (DI) створюється між мультичасовими зображеннями SAR для виявлення змін. Однак DI може негативно вплинути на продуктивність виявлення змін. Щоб уникнути цього, методи глибокого навчання ігнорували процес генерації DI. Для управління рухом і моніторингу безпеки на морі широко використовується виявлення кораблів на космічних знімках. Завдяки візуалізованому вмісту та властивостям високої роздільної здатності космічні зображення перевершують інші зображення дистанційного зондування у виявленні об’єктів. Однак, порівняно з інфрачервоними та радіолокаційними зображеннями з синтетичною апертурою, космічні зображення легко піддаються впливу погодних умов. Крім того, складність обробки зображень зростає, оскільки більша база даних обробляється для більш високої роздільної здатності.

Інші програми: вирівнювання обличчя відіграє важливу роль у різних візуальних програмах, таких як розпізнавання обличчя. Однак у екстремальних ситуаціях, коли знімаються зображення обличчя, вирівнювання обличчя може призвести до труднощів під час процесу аналізу. Тому для вирішення цієї проблеми розглядалися різні моделі для варіації форми та зовнішнього вигляду. Залежно від використовуваної моделі стандартні підходи можна умовно розділити на три групи: модель активного вигляду, локальна модель з обмеженнями та регресія. У порівнянні зі звичайними методами на основі регресії, адаптивні каскадні глибокі згорткові нейронні мережі (ACDCNN), запропоновані Донгом для виявлення точок обличчя, значно зменшили складність системи [39]. Dong та ін. покращив базові DCNN, використовуючи адаптивний спосіб навчання з різними мережевими архітектурами. Результати експерименту показали, що їхні мережі можуть досягти кращої продуктивності, ніж DCNN або інші новітні методи. Слід зазначити, що анотація зображення з кількома мітками є актуальною темою в області комп'ютерного зору. Крім того, методи глибокого навчання нещодавно були застосовані до програм пошуку зображень на основі вмісту. Більш конкретно, для вирішення завдань крос-модального пошуку було введено AE відповідності (Corr-AE) шляхом кореляції прихованих представлень двох унімодальних AE .

У комп'ютерному зорі видалення шумів є важливою проблемою, оскільки цифрові зображення спотворюються шумом під час отримання та передачі. Хоча існує багато алгоритмів усунення шуму, більшість із них розроблено для особливих випадків і їм бракує загальності. Фільтр Вінера добре працює для видалення гаусових шумів [18]. Однак це вимагає знання про автокореляційні функції вхідних даних. Що стосується придушення зашумлених зображень із краями, медіанна фільтрація ефективно бореться з шумами солі та перцю. Таким чином, були запропоновані багатообіцяючі автокодери з розрідженим шумозаглушенням (SSDAE) з багатообіцяючою ефективністю видалення шуму. Крім того, адаптивні багатоколонкові SSDAE (AMC-SSDAE) були представлені для підвищення надійності фільтра.

РОЗДІЛ 3 АНАЛІЗ РЕЗУЛЬТАТІВ ВПРОВАДЖЕННЯ МОДЕЛІ СИСТЕМИ ВІДЕОСПОСТЕРЕЖЕННЯ ЗІ ШТУЧНИМ ІНТЕЛЕКТОМ

У сучасному світі системи відеоспостереження відіграють ключову роль у забезпеченні безпеки та моніторингу як у публічних, так і у приватних просторах. Традиційні системи, які базуються на ручному аналізі відеопотоків, мають значні обмеження, пов’язані зі швидкістю обробки даних та необхідністю постійного контролю з боку оператора. Тому зростає потреба у впровадженні розумних систем відеоспостереження, що базуються на алгоритмах штучного інтелекту (ШІ) та глибокого навчання.

Запропонована модель системи відеоспостереження, розглянута у цьому дослідженні, використовує архітектуру MobileNet-SSD для виявлення та підрахунку об’єктів у відеопотоці, а також банк фільтрів Калмана для надійного відстеження їх руху. Завдяки цьому система здатна працювати в реальному часі та забезпечувати високу точність обробки даних навіть за умов обмежених обчислювальних ресурсів. Основою реалізації такої моделі є вбудовані пристрої з низьким енергоспоживанням, зокрема граничні вузли на базі пристрою UpSquared2 із блоком процесора зору (VPU), що дозволяє ефективно прискорювати виконання завдань на згорткових нейронних мережах (CNN).

У цьому розділі розглянуто особливості проєктування, впровадження та аналізу результатів роботи запропонованої моделі відеоспостереження зі штучним інтелектом. Важлива увага приділена таким аспектам, як:

- розробка архітектури системи та алгоритмів обробки зображень;

- ефективність роботи моделі у реальному часі при використанні різних апаратних ресурсів;

- точність виявлення та відстеження об’єктів у складних сценаріях (перекриття, зміна освітлення);

- енергоспоживання системи при обробці відеопотоків.

Завдяки отриманим результатам підтверджено доцільність впровадження розумної системи відеоспостереження у розподілені інфраструктури безпеки, що

забезпечують оптимальне поєднання продуктивності, точності та енергоефективності.

3.1 Розробка та дизайн моделі відеоспостереження зі штучним інтелектом

Розробка сучасних систем відеоспостереження зі штучним інтелектом передбачає інтеграцію передових методів комп'ютерного зору та оптимізованих апаратних засобів, що забезпечують обробку відеопотоків у реальному часі. Запропонована модель має на меті досягти високої точності виявлення та відстеження об'єктів при мінімальних ресурсних витратах, що особливо важливо для вбудованих систем із низьким енергоспоживанням.

Архітектура системи відеоспостереження.

Основу системи складає розподілена архітектура, що використовує граничні обчислення (Edge Computing). Головними елементами системи є:

- граничний вузол (Edge Node): Виконує основні завдання, включаючи обробку відеопотоків, виявлення об'єктів та їх подальше відстеження. Для реалізації використано платформу UpSquared2, яка включає процесор Intel Atom x7-E3950 і блок процесора зору Intel Myriad-X VPU.

- камери відеоспостереження: Забезпечують подачу відеопотоків у реальному часі, що обробляються на рівні граничного вузла.

- алгоритми обробки зображень: Інтегрована нейронна мережа MobileNet-SSD відповідає за виявлення людей у відеопотоці, тоді як фільтри Калмана забезпечують надійне відстеження об'єктів у динамічних умовах.

Програмне забезпечення та алгоритми.

Для оптимізації виконання завдань на обмежених апаратних ресурсах було обрано фреймворк OpenVino. Його функціональні можливості дозволяють:

- оптимізувати модель нейронної мережі за допомогою Model Optimizer, що перетворює архітектуру у проміжне представлення (IR).

- виконувати інференс на різних апаратних модулях, зокрема VPU та CPU, завдяки Inference Engine, що балансує навантаження системи.

Основний алгоритм обробки відеопотоків включає наступні етапи:

1. Попередня обробка зображень: Зменшення шуму, нормалізація піксельних значень, визначення регіону інтересу (ROI) та зміна розміру зображення до 300 × 300 пікселів для адаптації до входу нейронної мережі MobileNet-SSD.

2. Виявлення об’єктів: Нейронна мережа MobileNet-SSD виконує виявлення людей, визначаючи координати обмежувальних рамок (bounding boxes).

3. Відстеження об’єктів: За допомогою банку фільтрів Калмана відбувається прогнозування руху об’єктів та корекція позицій у випадку часткових перекриттів або тимчасової відсутності об’єкта у кадрі.

4. Паралельна обробка потоків: Завдяки можливостям Myriad-X VPU, система підтримує паралельну обробку до 12 відеопотоків, що значно збільшує продуктивність у реальному часі.

Переваги розробленої моделі:

1. Висока точність виявлення: Архітектура MobileNet-SSD, навчена на датасетах COCO та Pascal VOC, демонструє високу точність при низьких обчислювальних витратах.

- реальний час: Завдяки паралельній обробці відеопотоків та використанню VPU, система досягає обробки зі швидкістю понад 30 кадрів на секунду (fps).

2. Енергоефективність: Система споживає лише 12-15 Вт, що дозволяє її роботу від портативних акумуляторів або альтернативних джерел енергії.

Таким чином, розробка та дизайн запропонованої системи відеоспостереження зі штучним інтелектом забезпечують баланс між високою продуктивністю, точністю та енергоефективністю.

3.2 Аналіз алгоритмів виявлення та відстеження об'єктів у системах відеоспостереження

Оскільки однією з вимог запропонованих систем є їх портативність і гнучкість у впровадженні, вибір вбудованої апаратної платформи був однією з основних задач, розглянутих у цьому дослідженні. Нижче наведено короткий опис

характеристик апаратних компонентів і програмного забезпечення, яке використовувалося.

Вбудованою платформою, обраною для інтелектуального вузла, є система UpSquared2 [12], спеціалізоване апаратне забезпечення, що включає:

- мікропроцесор Intel Atom x7-E3950,
- 8 Гігабайт (ГБ) оперативної пам’яті (RAM),
- 64 ГБ вбудованої MultiMediaCard (eMMC) пам’яті.

Крім того, платформа включає модуль глибокого навчання Intel Movidius Myriad X VPU, систему на кристалі (SoC), яка використовується для прискорення виконання висновків штучного інтелекту (ШІ) при низькому енергоспоживанні.

Мета пристрою Myriad-X полягає в обробці висновків глибоких нейронних мереж (DNN) на високій швидкості та з мінімальним енергоспоживанням. Згідно з описом виробника, архітектура пристрою дозволяє виконувати понад 4 трильйони операцій за секунду (TOPS). Такий рівень обчислювальної потужності досягається завдяки комбінації нейронного обчислювального двигуна та 16 128-бітних VLIW SHAVE (streaming hybrid architecture vector engine) процесорів, які входять до складу пристрою. Крім того, система складається з двох LEON4 Core CPUs (архітектура RISC; SPARC v8). Як зазначалося раніше, цей пристрій дозволяє забезпечити високу швидкість обробки висновків із низьким енергоспоживанням, що робить його ідеальним для реалізації висновків на граничних вузлах і у вбудованих системах, що працюють від батарей.

Для виконання алгоритмів ШІ у VPU використовується фреймворк OpenVino, який полегшує оптимізацію та розгортання згорткових нейронних мереж (CNN). Фреймворк OpenVino включає два основні інструменти, як показано на рис.3.1:

- оптимізатор моделей (Model Optimizer).
- двигун висновків (Inference Engine).

Ці інструменти дозволяють оптимізувати модель та виконувати її на різних апаратних платформах, таких як Intel CPUs, VPUs або GPUs, скорочуючи час виконання. Однією з переваг цього фреймворку є те, що він може бути встановлений на будь-якому пристрої, який відповідає мінімальним вимогам. Крім

того, ці два модулі можна встановлювати окремо, дозволяючи використовувати оптимізатор моделей на комп'ютері, що використовується для навчання мережі, і двигун висновків — у вбудованій системі.

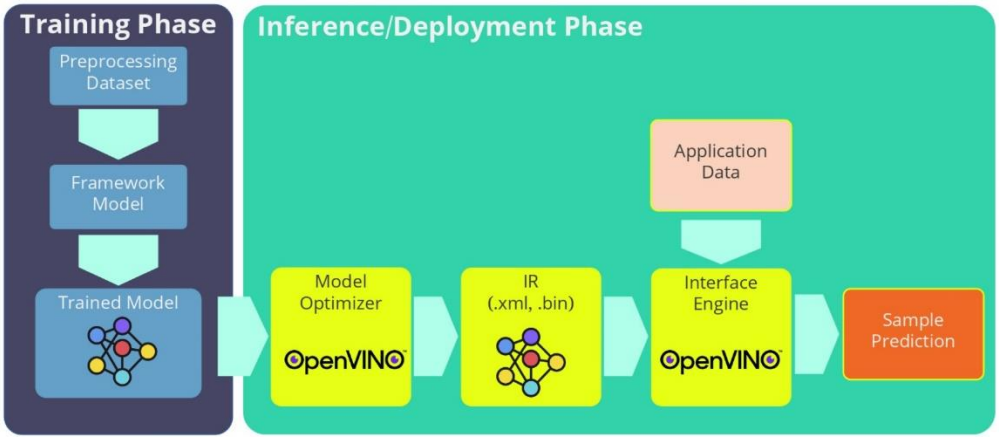


Рис.3.1 Структура OpenVino

Оптимізатор моделей — це застосунок, що працює через командний рядок і дозволяє налаштовувати та оптимізувати нейронні моделі для прискорення виконання висновків у системі. Оптимізатор моделей OpenVino сумісний із різними бібліотеками, такими як Tensorflow, Pytorch або Caffe. У процесі роботи оптимізатор створює файли проміжного представлення (Intermediate Representation, IR). До них входять:

- файл із розширенням .xml, що визначає шари, розміри та з'єднання архітектури;

- файл із розширенням .bin, що задає ваги кожного параметра архітектури.

Щодо двигуна висновків (Inference Engine), як зазначено вище, цей модуль OpenVino можна встановити на будь-який пристрій незалежно від оптимізатора моделей. Двигун завантажує файли IR та виконує висновки на обраному апаратному забезпеченні (CPU, VPU чи GPU). Крім того, двигун висновків відповідає за балансування навантаження під час виконання висновків, щоб уникнути перевантаження будь-якого пристрою. Таким чином, навантаження розподіляється між ядрами процесора (CPU) або між групою VPUs.

3.3 Впровадження моделі системи відеоспостереження на основі штучного інтелекту

Впровадження інтелектуальної системи відеоспостереження зі штучним інтелектом (ШІ) є ключовим етапом у реалізації сучасних рішень для виявлення, відстеження та підрахунку об'єктів у реальному часі. Основна задача полягає у розробці архітектури, яка ефективно використовує граничні обчислення для обробки відеопотоків із мінімальним енергоспоживанням та високою продуктивністю. Система базується на поєднанні алгоритмів MobileNet-SSD для виявлення об'єктів та фільтрів Калмана для їх відстеження. При цьому використовується спеціалізована вбудована платформа UpSquared2 з процесором Intel Movidius Myriad-X VPU, що дозволяє досягти високої ефективності та гнучкості у роботі системи.

Сучасна архітектура систем відеоспостереження на основі штучного інтелекту потребує ретельного підходу до проєктування як апаратної, так і програмної частини, оскільки такі системи мають виконувати завдання у реальному часі, забезпечуючи високу продуктивність, надійність та енергоефективність. Запропонована архітектура орієнтована на граничні обчислення, що дозволяє обробляти дані локально, без необхідності передачі інформації у хмарні сервери, тим самим зменшуючи затримки та навантаження на мережу.

Основу архітектури складають три ключові компоненти:

1. Передобробка даних,
2. Виконання інференції на нейронній мережі,
3. Постобробка результатів за допомогою фільтра Калмана.

Загальний потік даних та обробки у системі наведено на рис.3.2, де показано розподіл завдань між центральним процесором (CPU) та блоком процесора зору (VPU).

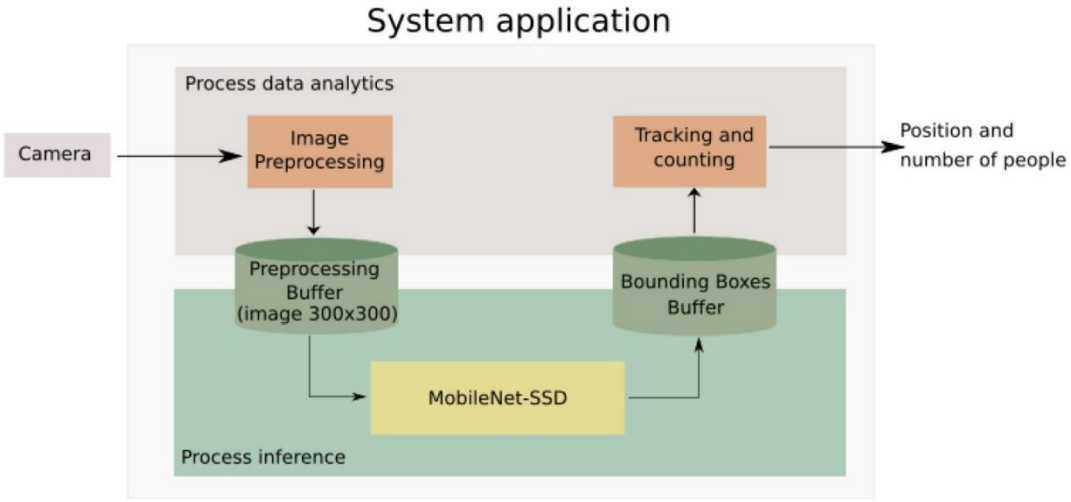


Рис.3.2 Потік даних і обробка за допомогою багатопоточності

Передобробка даних є першим етапом обробки у системі, де здійснюється підготовка відеопотоку для подальшого аналізу нейронною мережею. Вихідний сигнал від камер відеоспостереження надходить до системи у форматі RGB-відеопотоку з високою роздільною здатністю. Однак, для ефективної роботи MobileNet-SSD**,** розмір вхідного зображення має бути 300 × 300 пікселів**.**

Основні кроки передобробки включають:

1. Вибір області інтересу (ROI) – виділення частини зображення, що містить ключові об'єкти для аналізу. Це дозволяє скоротити обчислювальні ресурси, зменшуючи обсяг оброблюваних даних.

2. Зменшення розміру зображення – зміна масштабу відеокадру до 300 × 300 пікселів для відповідності вимогам нейронної мережі MobileNet-SSD. Цей етап мінімізує обчислювальну складність, зберігаючи при цьому важливі ознаки об'єктів.

3. Нормалізація значень пікселів – переведення значень пікселів у діапазон від -1 до +1. Це дозволяє покращити роботу нейронної мережі, оскільки вона краще обробляє відмасштабовані дані**.**

4. Мінімізація шуму (опційно) – у випадках зображень низької якості або із сильним шумом можуть застосовуватися алгоритми фільтрації, наприклад, медіанна фільтрація або розмиття по Гаусу**.**

Передобробка виконується на CPU, оскільки ця задача вимагає менше обчислювальних ресурсів порівняно з нейронною інференцією.

Інференція є найважливішим етапом у роботі системи, оскільки вона відповідає за виявлення об'єктів (людей) на відеопотоці. Для цієї задачі обрано архітектуру MobileNet-SSD, що поєднує у собі ефективність та швидкість роботи, оптимізовану для граничних обчислень.

Основні особливості MobileNet-SSD:

- глибокі (depthwise) згортки – зменшують обчислювальну складність, обробляючи кожен канал зображення окремо.

- точкові (pointwise) згортки – використовуються для об'єднання інформації з усіх каналів, створюючи вихідні ознаки для подальшої класифікації.

- оптимізація для вбудованих систем – завдяки малій кількості параметрів MobileNet-SSD забезпечує високу продуктивність навіть на пристроях із обмеженими ресурсами.

Реалізація інференції здійснюється на VPU Myriad-X, що інтегрований у платформу UpSquared2. Myriad-X має здатність обробляти до 4 відеопотоків одночасно завдяки своїй архітектурі, що підтримує паралельне виконання. Використання OpenVino для оптимізації та розгортання нейронної мережі дозволяє досягти мінімального часу обробки кожного кадру (~8.71 мс для MobileNet-SSD).

Постобробка виконується після завершення інференції на VPU. Результати, отримані від MobileNet-SSD у вигляді bounding boxes, передаються для подальшого аналізу. На цьому етапі використовується фільтр Калмана, який забезпечує відстеження об'єктів у відеопотоці.

Основні функції постобробки:

1. Прогнозування руху – фільтр Калмана дозволяє передбачити майбутнє положення об'єкта, навіть якщо він тимчасово втрачається детектором через оклюзії або інші перешкоди.

2. Ідентифікація об'єктів – система асоціює кожне нове виявлення з існуючим трекером, забезпечуючи безперервність відстеження.

3. Видалення трекерів – якщо об'єкт відсутній у кадрах протягом певного періоду, відповідний трекер завершує свою роботу, звільняючи ресурси системи.

Постобробка виконується на CPU, що дозволяє уникнути перевантаження VPU і збалансувати роботу системи.

Запропонована архітектура підтримує масштабовану обробку відеопотоків, що є ключовою перевагою для сучасних систем відеоспостереження. У випадку використання лише CPU, система може обробляти до 2 відеопотоків зі швидкістю 30 кадрів/с. При використанні VPU Myriad-X кількість оброблюваних потоків зростає до 12, забезпечуючи роботу у реальному часі.

Завдяки підтримці паралельної обробки, система може легко адаптуватися до різних сценаріїв, включаючи:

- моніторинг малих приміщень (до 4 камер),

- розподілені системи відеоспостереження для великих площ, де потрібна синхронна обробка багатьох потоків.

Таким чином, архітектура системи відеоспостереження поєднує у собі гнучкість, надійність та енергоефективність, роблячи її ідеальною для застосування у реальних умовах, де необхідна робота у реальному часі та з низьким енергоспоживанням.

Використання MobileNet-SSD для виявлення об’єктів

У системах відеоспостереження на основі штучного інтелекту одним із найважливіших завдань є виявлення об'єктів у реальному часі. Це завдання є основою для подальших процесів, таких як відстеження, підрахунок об'єктів та аналіз поведінки. У нашій системі для реалізації функції виявлення об'єктів було обрано MobileNet-SSD – ефективну нейронну мережу, оптимізовану для роботи на пристроях із обмеженими ресурсами.

MobileNet-SSD поєднує в собі швидкість обробки та високу точність виявлення при мінімальних обчислювальних витратах, що робить її ідеальною для граничних обчислень (Edge Computing). У цьому розділі детально розглянемо роботу MobileNet-SSD, її архітектуру, принципи функціонування, а також особливості її впровадження у нашій системі відеоспостереження.

MobileNet – це легка згорткова нейронна мережа, створена спеціально для мобільних і вбудованих пристроїв. Вона використовує унікальну структуру глибинно-сепарабельних згорток (Depthwise Separable Convolutions), яка дозволяє значно зменшити обчислювальну складність у порівнянні зі стандартними згортками.

З іншого боку, SSD (Single Shot MultiBox Detector) є методом детекції об'єктів, який дозволяє одночасно виконувати локалізацію об'єктів та класифікацію за один прохід мережі. Ця архітектура є ефективнішою у порівнянні з традиційними методами, такими як Faster R-CNN, оскільки вона не потребує додаткового етапу регіонального пропонування (Region Proposal).

Архітектура MobileNet-SSD складається з двох основних частин:

1. Feature Extractor (екстрактор ознак)

А. Перша частина мережі відповідає за вилучення ознак з вхідного зображення. У MobileNet цей процес виконується за допомогою глибинно-сепарабельних згорток.

Б. Глибинно-сепарабельна згортка складається з двох етапів:

- Depthwise Convolution – застосування окремого ядра до кожного каналу вхідного зображення.

- Pointwise Convolution – застосування 1×1 згортки для об'єднання отриманих ознак.

Обчислювальна вартість глибинно-сепарабельної згортки визначається як:

$$C_{\{DW+PW\}} = D_K^2 \cdot M \cdot D_F^2 + M \cdot N \cdot D_F^2 \quad (3.1)$$

D_K – розмір ядра,

M – кількість вхідних каналів,

N – кількість вихідних каналів,

D_F – розмір вхідного зображення.

Ця формула показує значне скорочення обчислень у порівнянні зі стандартною згорткою, де вартість дорівнює

$$C_{\{DW+PW\}} = D_K^2 \cdot M \cdot D_F^2 + M \cdot N \cdot D_F^2 \quad (3.2)$$

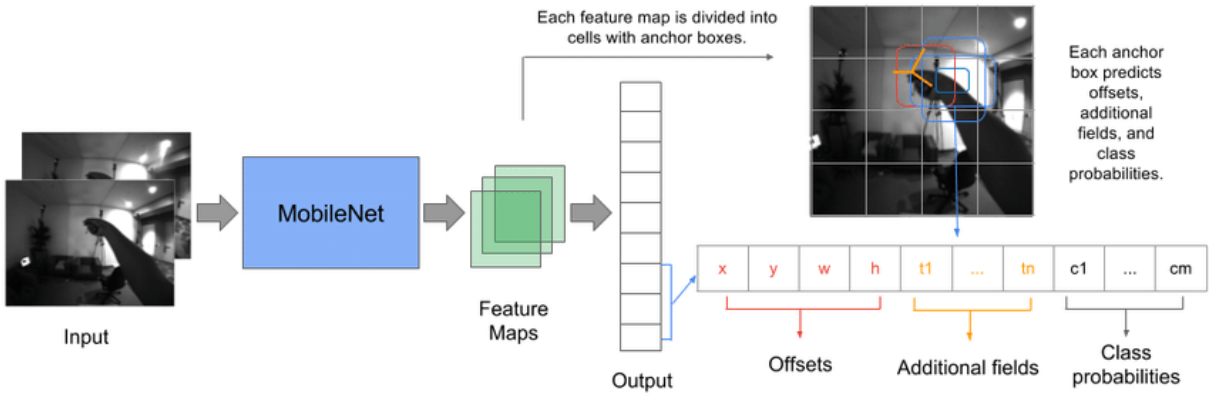


Рис.3.3 Архітектура MobileNet-SSD із глибинно-сепарабельними згортками.

2. Prediction Layers (шари передбачення)

а). Після вилучення ознак мережа передає дані до шарів передбачення, які виконують локалізацію та класифікацію об'єктів.

б). SSD використовує кілька шарів ознак з різними масштабами для виявлення об'єктів різних розмірів:

- Ранні шари використовуються для детекції дрібних об'єктів,
- Пізні шари відповідають за детекцію великих об'єктів.

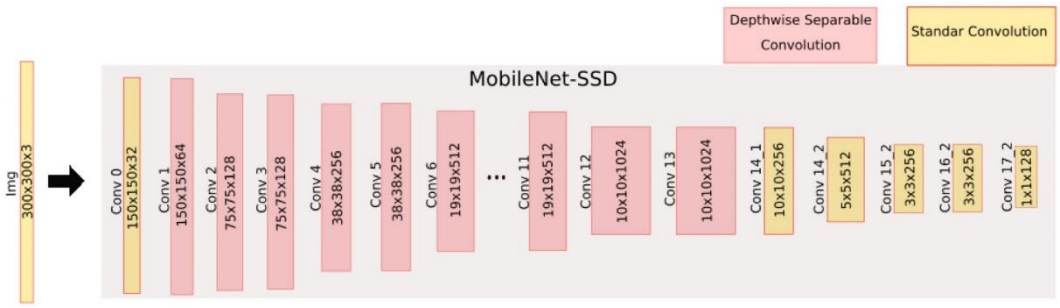


Рис.3.4 Потік даних у шарах передбачення SSD

Для кожного вікна (bounding box) мережа генерує наступну інформацію:

- координати меж прямокутника (локалізація об'єкта);
- ймовірність належності об'єкта до певного класу (класифікація).

Переваги MobileNet-SSD у системі відеоспостереження

1. Швидкістьобробки: Завдяки глибинно-сепарабельним згорткам MobileNet-SSD забезпечує високу швидкість роботи, що дозволяє обробляти відеопотік у реальному часі. На нашій платформі UpSquared2 з процесором Myriad-X VPU середній час інференції становить 8.71 мс на кадр.

2. Енергоефективність: Використання MobileNet-SSD у поєднанні з VPU дозволяє мінімізувати енергоспоживання системи, що робить її ідеальною для портативних застосувань, де живлення здійснюється від батарей.

3. Висока точність: Завдяки оптимізації та використанню COCO та Pascal VOC як навчальних наборів даних, мережа MobileNet-SSD демонструє середню точність (mAP) на рівні 72.7%. Це забезпечує надійне виявлення об'єктів навіть у складних умовах, таких як часткові оклюзії, зміни освітлення або велика кількість об'єктів на сцені.

4. Масштабованість: MobileNet-SSD може обробляти кілька відеопотоків одночасно завдяки паралельному виконанню на Myriad-X VPU. Як показано на рис.3.5, система може працювати з 12 відеопотоками зі швидкістю 30 кадрів/с.

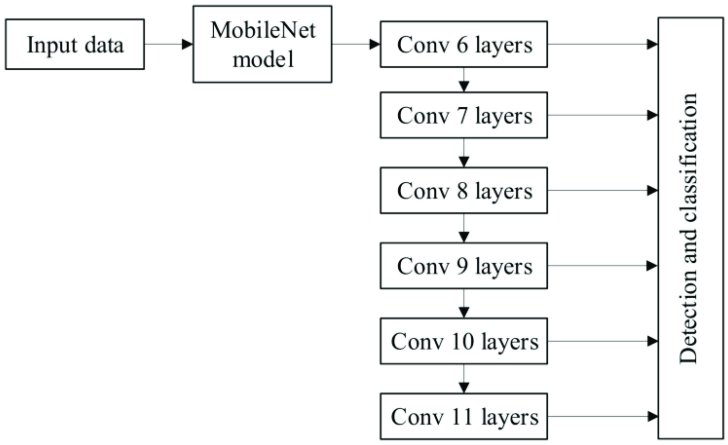


Рис.3.5 Масштабована обробка кількох відеопотоків у MobileNet-SSD на VPU

Процес навчання MobileNet-SSD. Для підготовки мережі MobileNet-SSD до виконання завдань відеоспостереження було використано наступні етапи:

1. Базове навчання: Мережа навчалася з нуля на наборі даних MS-COCO, що містить понад 80 000 зображень із різноманітними об'єктами, включаючи людей.

2. Тонке налаштування: Після базового навчання мережа була тонко налаштована на наборах Pascal VOC 2007 та VOC 2012 для підвищення точності виявлення людей.

3. Оптимізація для VPU: Навчені ваги мережі були конвертовані у формат IR (Intermediate Representation) за допомогою OpenVino Model Optimizer для подальшого розгортання на Myriad-X.

4. Валідація результатів: Точність роботи мережі була перевірена на наборі EPFL із двома складними сценаріями: лабораторія та коридор. Як показано в Таблиці 1, точність та відгук для мережі MobileNet-SSD склали понад 80% у більшості випадків.

Реалізація відстеження за допомогою фільтрів Калмана

Відстеження об’єктів у системах відеоспостереження є наступним етапом після виявлення об’єктів. Воно дозволяє не тільки локалізувати положення об’єкта у певний момент часу, а й прогнозувати його майбутнє положення, забезпечуючи плавне та безперервне відстеження навіть у випадках часткової або короткочасної втрати видимості. Для реалізації функції відстеження у нашій системі було використано фільтри Калмана – потужний і ефективний метод для оцінювання стану об’єктів у динамічних системах.

Фільтр Калмана – це рекурсивний алгоритм, який на основі попередніх спостережень і модельованого руху об’єкта оцінює його поточний стан та передбачає майбутнє положення. Це робиться шляхом двох основних етапів:

1. Прогнозування (Prediction): На основі моделі руху об’єкта (наприклад, лінійний рух із постійною швидкістю) фільтр Калмана обчислює очікуваний стан об'єкта у наступний момент часу.

2. Оновлення (Correction). Фільтр отримує реальні спостереження з вхідних даних (координати прямокутника виявлення) і оновлює прогноз, зменшуючи похибку за допомогою оцінки невизначеності вимірювань.

Фільтр Калмана описується системою рівнянь, що включають:

- стан об'єкта $x_k = [x_k; v_k]$, що містить координати положення та швидкість.

Матрицю переходу станів F яка моделює рух об'єкта.

Вимірювання z_k, які надаються мережею MobileNet-SSD.

Матрицю вимірювань H , що перетворює стан об'єкта у вимірювані величини.

Коваріацію похибок P_k і матрицю шуму Q.

Рівняння прогнозування (Prediction):

$$x_{\{k|k-1\}} = F\, x_{\{k-1|k-1\}} + w_k \tag{3.3}$$

$$P_{\{k|k-1\}} = F\, P^T_{\{k-1|k-1\}F} + Q \tag{3.4}$$

Рівняння оновлення (Correction):

$$K_k = P^T_{\{k|k-1\}H} \left(H\, P^T_{\{k|k-1\}H} + R\right)^{\{-1\}} \tag{3.5}$$

$$x_{\{k|k\}} = x_{\{k|k-1\}} + K_k \left(z_k - H\, x_{\{k|k-1\}}\right) \tag{3.6}$$

Де:

K_k – Калманівський коефіцієнт посилення,

R– коваріація шуму вимірювань,

w_k – шум процесу.

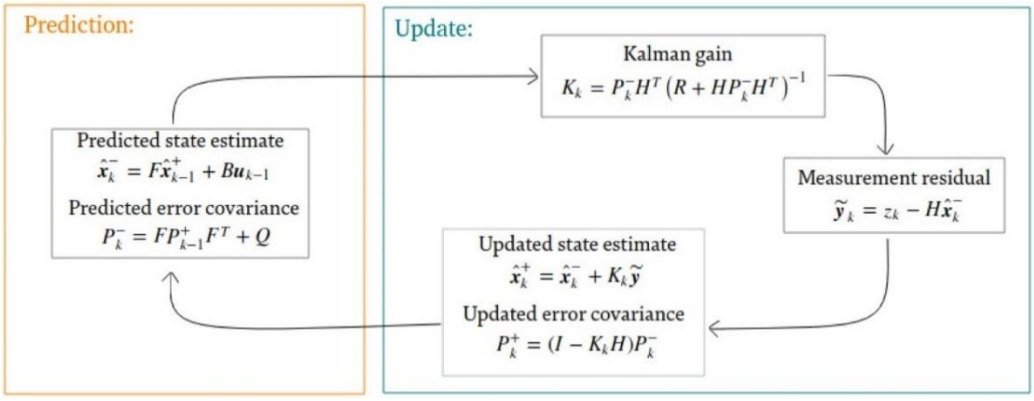


Рис.3.6 Схематична структура фільтра Калмана, що включає етапи прогнозування та оновлення

У нашій системі кожен об'єкт, який виявляється мережею MobileNet-SSD, отримує власний фільтр Калмана. Процес реалізації включає наступні кроки:

1. Ініціалізація трекера: Після того, як об'єкт виявлено протягом NNN послідовних кадрів, система створює новий трекер з використанням фільтра Калмана. Ініціалізація включає задавання початкових координат об'єкта та коваріації похибок.

2. Асоціація об'єкта з фільтром: Для кожного нового кадру система повинна зіставити виявлені об'єкти з існуючими трекерами. Це завдання вирішується за допомогою алгоритму Hungarian, який мінімізує різницю між прогнозами трекерів та новими вимірюваннями.

3. Прогноз стану: На основі попереднього стану об'єкта фільтр Калмана прогнозує його нові координати. Це дозволяє системі підтримувати трек навіть тоді, коли об'єкт тимчасово зникає з поля зору камери через оклюзію.

4. Оновлення стану: Якщо об'єкт знову з'являється у кадрі, фільтр Калмана оновлює свій стан з урахуванням нових вимірювань. Цей процес мінімізує похибку прогнозу та забезпечує точне відстеження.

5. Завершення трекера: Якщо об'єкт не виявляється протягом тривалого часу, система видаляє трекер для оптимізації використання ресурсів.

Рис.3.7 Алгоритм роботи фільтрів Калмана у системі відеоспостереження.

Переваги використання фільтрів Калмана:

1. Стійкість до перешкод і шумів: Фільтр Калмана є ефективним інструментом для відстеження у реальному світі, де присутні шуми вимірювань та оклюзії. Навіть якщо виявлення об'єкта на деякий час припиняється, фільтр продовжує прогнозувати його майбутнє положення.

2. Мінімальні обчислювальні витрати: На відміну від більш складних методів, фільтр Калмана має низькі обчислювальні вимоги, що дозволяє йому працювати у реальному часі навіть на вбудованих пристроях із обмеженими ресурсами.

3. Гнучкість: Фільтри Калмана можна легко адаптувати до різних сценаріїв, наприклад, для відстеження об'єктів, що рухаються з змінною швидкістю або у випадках, коли об’єкт періодично зникає зі сцени.

4. Інтеграція з MobileNet-SSD: Використання результатів виявлення з MobileNet-SSD як вхідних даних для фільтрів Калмана забезпечує синергію між виявленням та відстеженням, підвищуючи надійність системи.

У нашій системі кожен об'єкт отримує індивідуальний фільтр Калмана. Для забезпечення ефективного масштабування система використовує багатопоточну обробку:

- потік 1: Відповідає за обробку виявлень від MobileNet-SSD.

- потік 2: Виконує прогноз та оновлення фільтрів Калмана для кожного об'єкта.

- потік 3: Здійснює перевірку на завершення треків.

Ця архітектура дозволяє системі одночасно відстежувати велику кількість об'єктів у режимі реального часу.

Паралельна обробка відеопотоків. Паралельна обробка відеопотоків є ключовим етапом у розробці сучасних систем відеоспостереження на основі штучного інтелекту (ШІ). У реальних умовах, коли система відеоспостереження повинна обробляти декілька відеопотоків одночасно, важливо забезпечити високу продуктивність, мінімальні затримки та оптимальне використання апаратних ресурсів. У цьому розділі буде детально описано принципи реалізації паралельної обробки на граничних вузлах, а також архітектурні рішення, що дозволяють досягти ефективної роботи системи.

У системі відеоспостереження, що базується на граничних обчисленнях, апаратна платформа UpSquared2 забезпечує можливість паралельної обробки завдяки використанню багатоядерного CPU та спеціалізованого Vision Processing Unit (VPU).

Кожен відеопотік у системі проходить через послідовність етапів обробки, які можна виконувати незалежно для кожного потоку. Це дозволяє реалізувати конвеєрну архітектуру із багатопотоковою обробкою, оптимізованою для граничних вузлів.

Основні етапи обробки відеопотоків включають:

- пвідеоданих (шумозаглушення, нормалізація, зміна розмірів).

- виявлення об'єктів за допомогою нейронної мережі MobileNet-SSD.

- відстеження об'єктів за допомогою фільтрів Калмана.

- зберігання та передачу результатів на центральний сервер або користувацький інтерфейс.

Використання багатоядерного процесора (CPU).

Процесор Intel Atom x7-E3950, інтегрований у платформу UpSquared2, має 4 ядра, що дозволяє виконувати декілька потоків одночасно. У нашій системі кожне

ядро процесора відповідає за обробку окремого відеопотоку або окремого етапу обробки в межах одного потоку. Це дозволяє досягти високої пропускної здатності системи та ефективно обробляти відеодані у реальному часі.

Наприклад:

- ядро 1: передобробка відеопотоку 1 (зміна розміру зображення, нормалізація).

- ядро 2: передобробка відеопотоку 2.

- ядро 3: обробка результатів інференції для відеопотоку 1 (фільтри Калмана).

- ядро 4: обробка результатів для відеопотоку 2.

Ця схема дозволяє повністю завантажити ресурси CPU та забезпечити конвеєрну обробку декількох потоків.

Використання Vision Processing Unit (VPU). Vision Processing Unit (VPU) Myriad-X забезпечує високу ефективність обробки нейронних мереж завдяки паралельному виконанню обчислень на спеціалізованих ядрах SHAVE. Однією з головних переваг VPU є можливість одночасної обробки кількох відеопотоків у реальному часі.

Основні характеристики Myriad-X:

- використання 16 ядер SHAVE для паралельної обробки.

- підтримка FP16 обчислень, що забезпечує баланс між точністю та продуктивністю.

- ефективна інтеграція з OpenVINO, що дозволяє оптимізувати моделі нейронних мереж та запускати їх на VPU з мінімальним споживанням енергії.

У нашій системі VPU відповідає за виконання інференції нейронної мережі MobileNet-SSD для до чотирьох відеопотоків одночасно. Це дозволяє значно розвантажити CPU і підвищити загальну продуктивність системи.

Для максимально ефективного використання апаратних ресурсів система реалізує конвеєрну архітектуру, у якій різні етапи обробки відеопотоків виконуються паралельно на різних ядрах процесора та модулях VPU.

Послідовність обробки відеопотоків у конвеєрній архітектурі:

1. Захоплення відеопотоку: Відеокамери передають відеопотік на граничний вузол через мережеве з'єднання.

2. Передобробка кадрів: Кожен відеопотік проходить передобробку, включаючи:

- зміна розмірів зображення до 300 × 300 пікселів.

- нормалізація піксельних значень у діапазоні$[-1.0, +1.0]$.

- вибір області інтересу (ROI) для зменшення обсягу оброблюваних даних.

3. Інференція нейронної мережі (VPU): Оброблені кадри передаються на VPU, де виконується інференція MobileNet-SSD для виявлення об'єктів.

Завдяки паралельній обробці, VPU може одночасно обробляти до чотирьох кадрів.

4. Відстеження об'єктів (CPU): Результати виявлення об'єктів передаються на CPU, де за допомогою фільтрів Калмана виконується відстеження руху об'єктів.

5. Агрегація результатів та передача даних. Оброблені дані (координати об'єктів, траєкторії руху) передаються на центральний сервер або відображаються у користувацькому інтерфейсі.

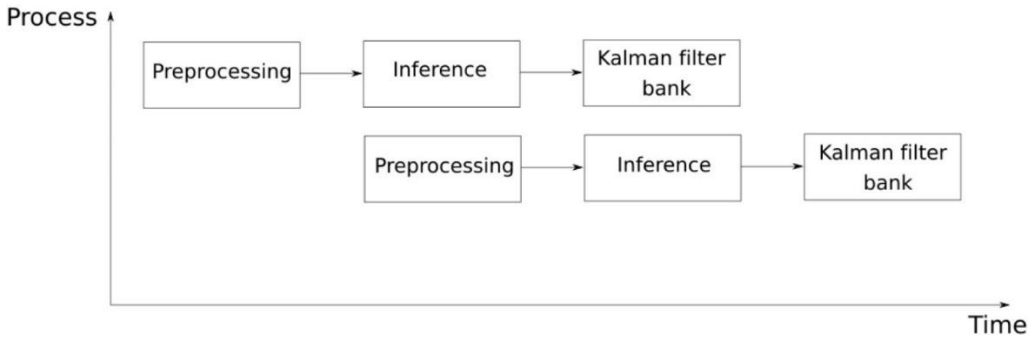


Рис.3.8 Етапи обробки відеопотоків у конвеєрній архітектурі

Система була розроблена з урахуванням масштабованості для обробки великої кількості відеопотоків. Завдяки використанню багатоядерного CPU та VPU система здатна обробляти до 12 відеопотоків одночасно при частоті 30 кадрів за

секунду (fps). Це робить її ідеальною для розподілених систем відеоспостереження, які потребують обробки даних у реальному часі.

Теоретична модель масштабованості передбачає, що за наявності більшої кількості граничних вузлів або додаткових VPUs продуктивність системи можна збільшити пропорційно кількості доступних ресурсів.

Переваги паралельної обробки відеопотоків:

1. Висока продуктивність: Завдяки конвеєрній архітектурі система досягає мінімальних затримок при обробці відеопотоків.

2. Оптимальне використання ресурсів: Розподіл завдань між CPU і VPU забезпечує ефективне навантаження апаратних компонентів.

3. Енергоефективність. Використання VPU Myriad-X дозволяє знизити енергоспоживання при одночасному підвищенні продуктивності.

4. Масштабованість: Система може бути легко масштабована для обробки більшої кількості відеопотоків за допомогою додаткових граничних вузлів.

3.4 Результати тестування системи відеоспостереження у реальному часі

Основною метою дослідження, представленого в даній роботі, є розробка портативної та високопродуктивної системи відеоспостереження, що дозволяє виявляти, відстежувати та підраховувати людей у реальному часі. Ця система базується на вбудованому обладнанні з використанням периферійних обчислень (Edge computing) та штучного інтелекту (ШІ).

Для детальної перевірки працездатності, ефективності та надійності запропонованої системи було розгорнуто повноцінну тестову платформу, яка дозволяє оцінити можливості системи в реальному часі за різних умов. Ключовою вимогою цього етапу було наближення умов тестування до реальних сценаріїв, що включають зміну освітлення, обмежені простори, перекриття об’єктів та обробку одночасно кількох відеопотоків. У процесі тестування проводилися експерименти з використанням реальних відеоданих, а також були реалізовані сценарії із різними конфігураціями апаратного забезпечення для оптимізації параметрів.

Тестова система була побудована на основі високопродуктивної та енергоефективної платформи UpSquared2, яка поєднує у собі апаратні модулі для обчислень, необхідних для роботи ШІ-моделей на периферійних вузлах.

Основні характеристики UpSquared2:

- процесор: Intel Atom x7-E3950, що має 4 ядра та забезпечує оптимальний баланс між продуктивністю та енергоспоживанням. Цей процесор підтримує багатопотокову обробку, що є критичним для обробки кількох відеопотоків одночасно.

- пам'ять: 8 ГБ оперативної пам'яті (RAM) забезпечує швидке завантаження нейронних мереж і обробку відеоданих без затримок. Для зберігання проміжних результатів обчислень використовується 64 ГБ eMMC ROM, що гарантує стабільну роботу системи.

- ШІ-прискорювач: Intel Movidius Myriad-X VPU (Vision Processing Unit) — ключовий компонент для паралельної обробки завдань машинного навчання. Myriad-X здатен обробляти до 4 трильйонів операцій на секунду (TOPS), що робить його ідеальним для інференції глибоких нейронних мереж із низьким енергоспоживанням.

- фреймворк OpenVino: для оптимізації моделей ШІ та прискорення обчислень використовується фреймворк OpenVino. Цей інструмент дозволяє адаптувати нейронні мережі, такі як MobileNet-SSD, для виконання на різних типах апаратного забезпечення (CPU, GPU, VPU).

Особливості апаратного налаштування:

1. Вбудована підтримка VPU дозволяє виконувати нейронні обчислення на SHAVE-ядрах Myriad-X, зменшуючи навантаження на центральний процесор.

2. Розподіл обчислювальних ресурсів між CPU і VPU забезпечує паралельну обробку відеопотоків.

3. Модульна структура системи дозволяє інтегрувати додаткові камери або датчики для розширення функціональності.

Переваги платформи UpSquared2:

- компактність і портативність: система має невеликий розмір, що дозволяє легко розміщувати її у будь-якому середовищі.

- енергоефективність: підтримка роботи від портативних батарей.

- продуктивність: одночасна обробка кількох потоків завдяки VPU.

Для оцінки можливостей системи та її адаптації до реальних сценаріїв використовувався набір даних EPFL, який є широко визнаним стандартом у галузі відеоаналітики для завдань виявлення та відстеження об'єктів. Цей набір даних був обраний завдяки його складності та різноманітності умов, що дозволяє оцінити стійкість системи до змін навколишнього середовища.

Характеристики тестового середовища:

1. EPFL-LAB (Лабораторія):

- Просторі умови: мінімальна кількість об'єктів у кадрі, велика площа та відсутність серйозних перешкод.

- Рух об'єктів: чотири людини, які змінюють свої позиції у кадрі, рухаючись назад і вперед до камери.

- Мінімальний рівень перекриття: об'єкти рідко перекривають один одного, що дозволяє оцінити точність системи у базових умовах.

- Рівномірне освітлення: стабільні умови освітлення без значних змін протягом зйомки.

Рис.3.10 b демонструє приклад лабораторного середовища, де умови сприятливі для виявлення та відстеження об'єктів.

2. EPFL-CORRIDOR (Коридор):

- обмежений простір: вузький коридор з великою кількістю об'єктів у кадрі.

- рух об'єктів: до восьми людей у різних частинах кадру, що створює високий рівень перекриття.

- зміни освітлення: коливання рівнів освітлення між різними відеопотоками, що створює додаткові виклики для алгоритмів.

- складність сценарію: висока кількість перекриттів і невеликі відстані між об'єктами дозволяють перевірити ефективність відстеження у складних умовах.

Рис.3.10a показує коридорне середовище з високим рівнем завантаженості сцени.

Переваги використання набору EPFL:

- забезпечує різноманітність сценаріїв для тестування.

- включає реалістичні умови, що імітують відеоспостереження у громадських та приватних місцях.

- дозволяє порівнювати результати з іншими методами завдяки широкому використанню цього набору даних у наукових дослідженнях.

Методи тестування та оцінки. Для тестування запропонованої системи використовувалися наступні метрики продуктивності:

1. Точність (Precision) — показник правильності виявлення об'єктів.

2. Відгук (Recall) — показник повноти виявлення об'єктів.

3. PR-криві (Precision-Recall Curves) — графічна візуалізація балансу між точністю та відгуком.

4. Час обробки — вимірювався для різних апаратних конфігурацій (CPU, VPU) при обробці одного або кількох відеопотоків.

5. Енергоспоживання — оцінювалося у різних режимах роботи системи (простой, CPU, CPU+VPU).

Для кожного тесту було виконано кілька ітерацій для отримання усереднених результатів. Це дозволяє знизити похибки вимірювань та забезпечити репрезентативність отриманих даних.

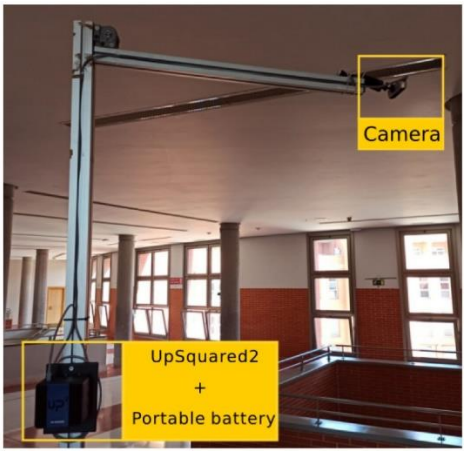


Рис.3.9 Встановлена тестова система

Підготовка системи до тестування.

1. Завантаження моделей: модель MobileNet-SSD була навчена на наборах даних COCO та уточнена на Pascal VOC.

2. Оптимізація моделей: фреймворк OpenVino використовувався для конвертації моделей у формат IR (Intermediate Representation), що дозволяє виконувати інференцію на VPU з максимальною швидкістю.

3. Налаштування камер: для тестування було використано кілька відеокамер, які передавали відеопотоки у реальному часі з роздільною здатністю 720p.

Під час тестування проводилася моніторингова перевірка системи для фіксації результатів обробки, зокрема:

- швидкість обробки кадрів (FPS);

- точність локалізації об’єктів;

- час виконання інференції на кожному обчислювальному модулі.

Таким чином, експериментальне налаштування забезпечило реалістичні умови тестування, що дозволяють оцінити ефективність системи відеоспостереження у реальному часі та порівняти результати із сучасними методами.

Для оцінки точності та продуктивності запропонованої системи було проведено аналіз PR-кривих (precision-recall), які дозволяють об’єктивно оцінити якість виявлення людей у різних середовищах. Тестування виконували на основі EPFL-набору даних, що складається з двох сценаріїв: лабораторного середовища (EPFL-LAB) і університетського коридору (EPFL-CORRIDOR). В обох випадках наявні умови, які характеризуються різним рівнем освітлення, перекриттям об'єктів і кількістю осіб, що ускладнюють задачу виявлення.

Для обґрунтованої оцінки запропонованої системи на базі MobileNet-SSD з фільтром Калмана, було обрано порівняння з іншими сучасними методами виявлення людей, що включають як класичні підходи, так і моделі глибокого навчання. До методів порівняння увійшли наступні алгоритми:

1. ACF (*Aggregate Channel Features*) – класичний метод, що поєднує AdaBoost і HOG-ознаки для RGB-зображень. Основною перевагою цього підходу є його

швидкість та низькі обчислювальні вимоги, що дозволяє використовувати його на обмежених апаратних платформах. Однак його основним недоліком є низька стійкість до складних умов, таких як перекриття об'єктів і зміни освітлення.

2. PCL-MUNARO – метод, що використовує RGB-D детектор на основі модифікованих HOG-ознак і сегментації по глибині. Цей підхід показує гарні результати у середовищах, де доступна глибина завдяки камерам типу RGB-D. Недоліком є висока залежність від наявності додаткових даних глибини, що обмежує універсальність системи.

3. DPOM (*Depth-based Presence On Ground Plane*) – алгоритм, що базується на використанні глибинної інформації та байєсівського виведення для визначення присутності людей на площині землі. DPOM показує високу точність у середовищах з великим ступенем перекриття об’єктів, проте є обчислювально складним і вимагає високої точності у вимірюванні глибини.

4. YOLO-v3 – сучасна згорткова нейронна мережа (CNN), що працює з RGB-зображеннями та є оптимізованою для обробки об’єктів у реальному часі. YOLO-v3 показує високу продуктивність на кольорових даних, але його недоліком є вразливість до складних умов освітлення та перекриття.

5. YOLO-depth – модифікована версія YOLO-v3, що перенавчена для обробки глибинних даних. Основною перевагою є можливість забезпечення конфіденційності, оскільки глибинна інформація не розкриває особистість людей. Однак точність моделі знижується у відсутності якісних глибинних зображень.

Для усіх моделей використовувалися однакові вхідні дані та поріг накладення (IoU) 40%, що дозволяє об'єктивно порівнювати результати між різними методами.

Результати тестування представлені у вигляді PR-кривих, рис.3.11, та зведених значень точності (precision) і відгуку (recall) для обох середовищ (EPFL-LAB та EPFL-CORRIDOR) у табл.3.3.

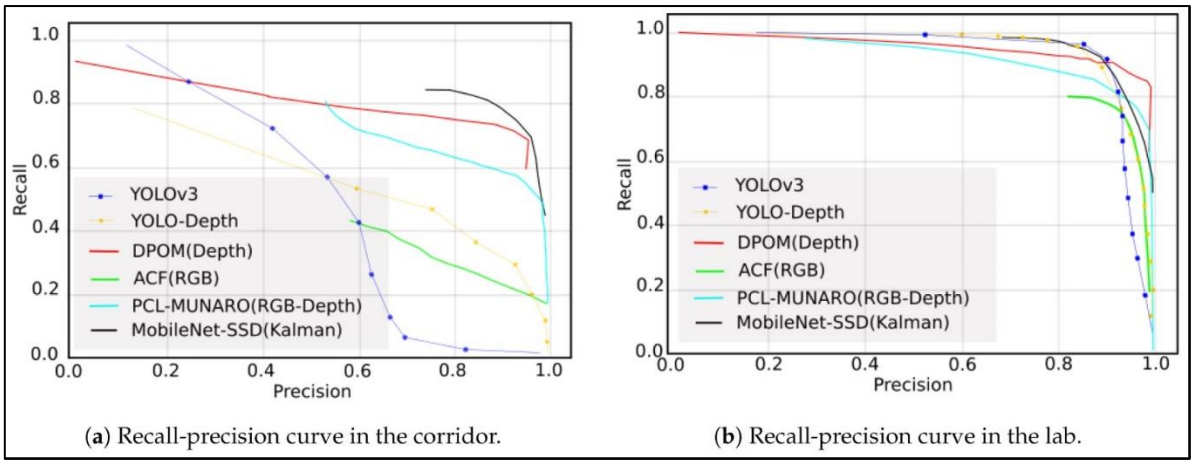


Рис.3.10 Криві точності та відгуку для різних середовищ у наборі EPFL

У лабораторному середовищі (EPFL-LAB) система MobileNet-SSD + Kalman досягла 87.82% точності та 88.14% відгуку. Це демонструє, що запропонована система є стабільною для середовищ із мінімальним рівнем перекриття об'єктів і стабільним освітленням. Універсальність і продуктивність системи дозволяє успішно виконувати завдання виявлення за умов обмежених обчислювальних ресурсів.

Таблиця 3.1

Результати точності та відгуку для набору EPFL

Метод	EPFL-LAB Точність	EPFL-LAB Відгук	EPFL-CORRIDOR Точність	EPFL-CORRIDOR Відгук
MobileNet-SSD+ Kalman	87.82	88.14	81.3	80.6
ACF	83.8	86.4	66.3	40.3
DPOM	98.5	85.4	96.3	70.9
PCL-MUNARO	88.61	82.36	92	56
YOLO-v3	89.7	90.3	58.6	59.8
YOLO-depth	89.8	88.6	78.4	47.9

У коридорному середовищі (EPFL-CORRIDOR) система показала 81.3% точності та 80.6% відгуку, що є найкращим результатом серед методів на основі RGB, поступаючись лише DPOM, який досяг 96.3% точності. Проте відгук DPOM

нижчий (70.9%), що свідчить про пропуск частини об'єктів у складних сценаріях з освітленням і високим ступенем перекриття.

Методи ACF та PCL-MUNARO, які є класичними підходами, демонструють значно гірші результати у порівнянні із нейронними мережами. Наприклад, у коридорному середовищі ACF показав лише 66.3% точності та 40.3% відгуку, що підтверджує їх низьку ефективність у сучасних динамічних умовах.

MobileNet-SSD з фільтрами Калмана показала оптимальний баланс між точністю та відгуком у двох тестових середовищах. Це робить запропоновану модель надійною та універсальною для реалізації у вбудованих системах відеоспостереження з обмеженими ресурсами. Важливо підкреслити, що завдяки використанню фільтрів Калмана, система ефективно відстежує об’єкти навіть при частковому перекритті, а MobileNet-SSD забезпечує високу швидкість обробки даних при низьких обчислювальних витратах.

Таким чином, результати тестування підтверджують, що розроблена система є конкурентоспроможною серед сучасних методів виявлення людей, забезпечуючи високу продуктивність та ефективність у різних середовищах.

Ефективність обробки відеопотоків у реальному часі є ключовим критерієм для сучасних систем відеоспостереження на основі штучного інтелекту. У цьому підрозділі проводиться детальний аналіз продуктивності системи при обробці одного та множинних відеопотоків, використовуючи апаратну платформу UpSquared2 із центральним процесором (CPU) та модулем VPU (Vision Processing Unit). Метою дослідження є визначення меж можливостей системи в умовах реального часу та її потенціалу для масштабування на більшу кількість камер.

Для оцінки продуктивності було проведено серію експериментів із використанням MobileNet-SSD і YOLO-v3 на різних апаратних ресурсах. Основні параметри, що вимірювалися, включають:

- час інференції (Inference Time) – час, необхідний для обробки одного кадру відео.

- кількість оброблюваних відеопотоків (Streams) – максимальна кількість відеопотоків, яку система здатна обробити при частоті 30 кадрів/с (fps).

- завантаження апаратних ресурсів (CPU/VPU Usage) – ступінь використання ядер процесора та VPU під час виконання завдань.

-затримка обробки (Latency) – час між надходженням кадру і отриманням результату інференції.

Для вимірювання були використані:

- CPU: Intel Atom x7-E3950.

- VPU: Intel Movidius Myriad-X.

- фреймворк: OpenVino, який оптимізує згорткові нейронні мережі (CNN) для ефективної роботи на CPU/VPU.

- набір даних: EPFL-LAB та EPFL-CORRIDOR, що містять реальні сценарії з різною кількістю людей і складністю умов.

Результати продуктивності інференції.

Порівняльний аналіз часу інференції для MobileNet-SSD та YOLO-v3 наведено у табл.3.2

Таблиця 3.2

Обчислювальні витрати для MobileNet-SSD та YOLO-v3

Мережа	tCPU (мс)	tVPU (мс)
MobileNet-SSD	13.93	8.71
YOLO-v3	47.54	30.07

Аналіз результатів.

1. Час інференції MobileNet-SSD:

- використання VPU дозволяє досягти часу обробки 8.71 мс для одного кадру, що є майже у 5 разів швидшим, ніж інференція YOLO-v3 на CPU (47.54 мс).

- на CPU обробка MobileNet-SSD займає 13.93 мс, що вдвічі перевищує швидкодію VPU, але все ще значно ефективніше у порівнянні з YOLO-v3.

2. Оптимізація за рахунок VPU:

- при використанні VPU спостерігається 64% скорочення часу інференції у порівнянні з виконанням на CPU.

- це досягається завдяки апаратній архітектурі Myriad-X, яка підтримує паралельну обробку за допомогою SHAVE-процесорів.

Обробка відеопотоків:

- система, що працює виключно на CPU, обмежена обробкою двох відеопотоків у реальному часі при 30 fps, оскільки кожне ядро процесора завантажене на 58.83%, рис.3.11.

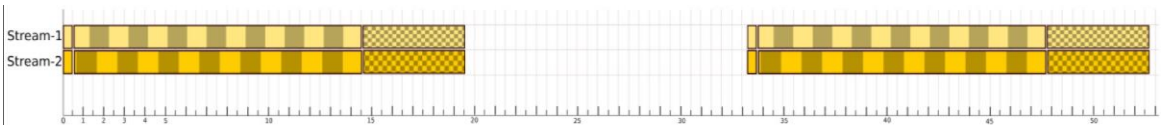


Рис.3.11 Обробка двох відеопотоків за допомогою CPU

- використання VPU дозволяє масштабувати систему для обробки до 12 відеопотоків одночасно, з частотою 30 fps. Завантаження VPU у такому режимі досягає 73%, що залишає резерв для подальшого підвищення продуктивності.

4. Графік обробки 12 відеопотоків:

- на рис.3.12 продемонстровано, як система розподіляє обробку між CPU та VPU, забезпечуючи одночасну обробку 12 відеопотоків.

- попередня обробка кадрів і фільтрація Калмана виконується на CPU, тоді як інференція MobileNet-SSD повністю відбувається на VPU.

- паралельна архітектура забезпечує мінімальну затримку між надходженням відео та отриманням результатів інференції.

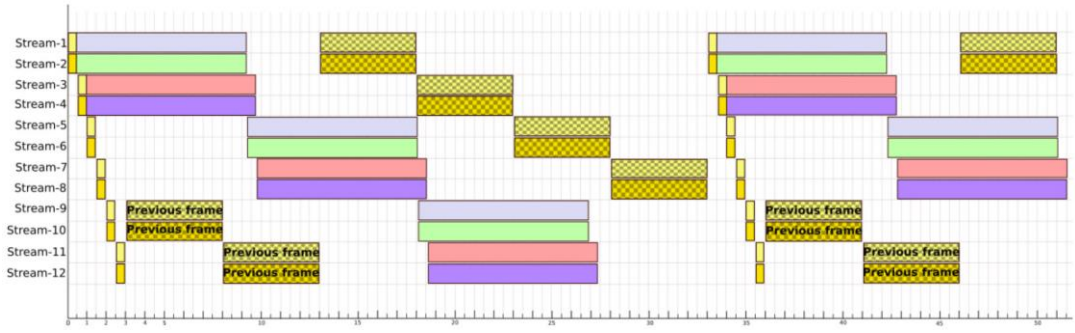


Рис.3.12 Обробка 12 відеопотоків із використанням CPU та VPU.

Система використовує паралельну архітектуру обробки, що дозволяє оптимізувати використання апаратних ресурсів. Основні переваги полягають у наступному:

1.Масштабованість:

- завдяки підтримці VPU, продуктивність системи може бути збільшена шляхом додавання нових відеокамер без суттєвого зниження швидкодії.

- використання глобальних буферів дозволяє синхронізувати обробку відеопотоків та балансувати навантаження між CPU та VPU.

2. Енергоефективність:

- паралельна обробка на VPU споживає менше енергії у порівнянні з використанням лише CPU.

- для обробки 12 відеопотоків середнє енергоспоживання системи залишається у межах 14.41 Вт.

3. Низька затримка:

- завдяки оптимізації алгоритмів та використанню OpenVino, система забезпечує мінімальну затримку обробки кадрів – до 19 мс на один кадр при використанні CPU та 8.71 мс на VPU.

4. Реальний час:

- система ефективно обробляє відеопотоки з частотою 30 fps, що відповідає вимогам сучасних додатків відеоспостереження.

Результати тестування показали, що система на основі MobileNet-SSD та VPU Myriad-X демонструє значну перевагу над традиційними методами обробки відео. Використання паралельної архітектури дозволяє одночасно обробляти до 12 відеопотоків у реальному часі з мінімальним енергоспоживанням. У порівнянні з YOLO-v3, час інференції скорочується у 4 рази, а використання VPU забезпечує додаткові 64% оптимізації у швидкодії.

Таким чином, система є ефективним рішенням для розподілених інтелектуальних відеосистем з можливістю роботи на обмежених апаратних ресурсах.

Енергоспоживання системи.

Ефективне використання енергії є одним із ключових факторів для сучасних вбудованих систем відеоспостереження, особливо коли мова йде про портативні пристрої, що працюють на батарейному живленні. Ретельний аналіз енергоспоживання системи було проведено для трьох основних режимів роботи: режим простою, виконання інференції виключно на CPU та комбіноване використання CPU та VPU для інференції. Результати вимірювань подані у таюл.3.5, а детальний аналіз надається у наступних підрозділах.

Система UpSquared2, яка є основою для нашого відеоспостереження, отримувала живлення від зовнішнього джерела із постійною напругою 5 В. Для вимірювання енергоспоживання було використано:

- ультиметр високої точності для моніторингу середнього струму (А).

- осцилограф для аналізу динамічних коливань споживаної потужності в процесі роботи.

- програмний моніторинг для оцінки навантаження на процесорні ядра (CPU) та ядра SHAVE модуля VPU під час різних режимів.

Енергоспоживання оцінювалося на основі таких параметрів:

- середня потужність (Вт) — загальна кількість енергії, що споживається системою у певному режимі роботи.

- енергія на завдання (W·ms) — інтегральне значення, яке показує витрати енергії для обробки одного кадру у кожному режимі.

Результати вимірювань у трьох режимах роботи.

Таблиця 3.3

Споживання енергії системи у різних режимах

Режим роботи	Середня потужність (Вт)	Енергія на завдання (W·ms)
Режим простою	5.41	-
AI + TRACK на CPU	12.05	234.25
AI на VPU + TRACK на CPU	14.41	204.91

Аналіз результатів та порівняння режимів.

Режим простою.У режимі простою система споживає 5.41 Вт. Це мінімальне значення енергоспоживання для платформи UpSquared2, коли жодні ресурси CPU чи VPU не використовуються для обчислень. Низький рівень потужності забезпечується завдяки:

- енергоефективному мікропроцесору Intel Atom x7-E3950, що автоматично переходить у режим зниженого споживання при відсутності навантаження.

- апаратній оптимізації материнської плати UpSquared2, яка мінімізує споживання енергії під час бездіяльності.

Виконання інференції на CPU.

У режимі виконання інференції виключно на CPU, система споживає 12.05 Вт, а енергія на завдання становить 234.25 W·ms. Високе споживання енергії пояснюється наступними факторами:

- відсутність апаратного прискорення: Центральний процесор виконує усі обчислювальні завдання, включаючи передобробку, інференцію та постобробку, що призводить до високого навантаження на ядра CPU.

- лінійне використання ресурсів: Обмеження у паралельній обробці призводить до значного збільшення загального часу обробки кадрів, що в свою чергу збільшує витрати енергії.

Виконання інференції на VPU+CPU.

Найкращі результати енергоспоживання було отримано при комбінованому використанні CPU та VPU. У цьому режимі середня споживана потужність становить 14.41 Вт, але енергія на завдання зменшується до 204.91 W·ms, що на 25% менше, ніж у режимі виключно на CPU. Це пояснюється наступними причинами:

1. Апаратне прискорення VPU: Модуль Myriad-X обробляє інференцію значно швидше завдяки 16 ядрам SHAVE та нейронному процесору для оптимізованої обробки згорткових нейронних мереж (CNN).

2. Паралельна обробка: Передобробка кадрів і постобробка (фільтри Калмана) виконуються на CPU, тоді як інференція повністю перенесена на VPU, що дозволяє рівномірно розподілити навантаження.

3. Швидше виконання завдань: Завдяки VPU загальний час обробки одного кадру скорочується до 8.71 мс, що знижує тривалість активного споживання енергії.

Порівняння з іншими системами

Для оцінки енергоефективності запропонованої системи було проведено порівняння з існуючими аналогами на ринку. Наприклад, система на базі NVIDIA Jetson Nano із мережею YOLOv3-tiny демонструє наступні характеристики:

- середня потужність: 9 Вт.

- продуктивність: 9 кадрів на секунду (fps).

Наша система на UpSquared2 із MobileNet-SSD забезпечує:

- середню потужність: 14.41 Вт.

- продуктивність: 30 кадрів на секунду (fps) для 12 відеопотоків.

Таким чином, співвідношення потужності до продуктивності є значно ефективнішим у нашій системі. Енергія на кадр, виміряна для Jetson Nano, становить 1 Вт·с на 9 кадрів або 111.1 W·ms на кадр, тоді як для нашої системи це значення лише 17.08 W·ms на кадр (при 12 потоках).

На основі проведеного аналізу можна виділити наступні переваги системи:

1. Оптимізація енергоспоживання:

- використання VPU дозволяє досягти 25% економії енергії порівняно з використанням лише CPU.

- енергоспоживання на один відеопотік становить 17.08 W·ms, що є оптимальним для портативних систем відеоспостереження.

2. Висока продуктивність:

- система здатна обробляти до 12 відеопотоків у реальному часі при 30 fps.

-в исока швидкодія інференції забезпечується завдяки паралельній обробці на CPU та VPU.

3. Ефективне використання ресурсів:

- платформа UpSquared2 дозволяє розподілити обчислювальні завдання між CPU та VPU, мінімізуючи час обробки та енергетичні витрати.

4. Порівняння з альтернативами:

Запропонована система перевершує аналогічні рішення на ринку за показниками ефективності енерговикористання та швидкодії.

Таким чином, система є не лише продуктивною, але й енергоефективною, що робить її ідеальним рішенням для розподілених периферійних систем відеоспостереження.

ВИСНОВКИ

Виконуючи перше поставлене завдання, у магістерській роботі, мною досліджено ринок розвитку систем відеоспостереження, особливості функціонування та перспективи розвитку.

Проаналізовано особливості впровадження ШІ в ситстеми відеоспостереження, виконано порівняння з традиційними ситстемами спостереження.

У другому розділі мною було розглянуто останні розробки глибоких нейронних мереж. Досліджено деякі широко використовувані архітектури глибокого навчання та виділено окремі програми для комп'ютерного зору, розпізнавання образів і розпізнавання мови. Більш конкретно, детально обговорюються чотири класи архітектур глибокого навчання, а саме обмежена машина Больцмана, мережі глибоких переконань, автокодер і згорточна нейронна мережа. Оскільки в додатках, що включають аналіз великих даних, рідко можливо отримати дані з мітками, алгоритми керованого навчання навряд чи можуть забезпечити задовільну продуктивність у таких випадках.

На основі цих підходів до глибокого навчання можливо використовувати алгоритми неконтрольованого навчання для обробки немаркованих даних. Крім того, компроміс між точністю та обчислювальною складністю можна гнучко регулювати в більшості алгоритмів глибокого навчання. Зі стрімким розвитком апаратних ресурсів і обчислювальних технологій ми впевнені, що глибокі нейронні мережі отримають все більшу увагу та знайдуть ширші застосування в майбутньому.

У третьому розділі досліджено портативна система відеоспостереження з обробкою AI CNN на краю, яка може виявляти та відстежувати людей надійним і надійним способом. Нова комп'ютерна технологія, що використовується на краю, — це апаратні модулі VPU, які дозволяють виконувати висновок CNN швидше й ефективніше, ніж центральний процесор у пристроях нижчого класу. Досліджувана система дозволяє реалізувати програму комп'ютерного зору з низьким енергоспоживанням і високою обчислювальною продуктивністю. Для досягнення

цього застосовується вбудований пристрій UpSquared2, який інтегрує Myriad-X VPU, де виконується обраний CNN, MobileNet-SSD. Однією з вимог було виконання системи в режимі реального часу. Для досягнення реального часу, використано фреймворк OpenVino 2020.2, який дозволяє оптимізувати CNN, спрощуючи його висновки та полегшуючи виконання на VPU. Крім того, ця структура дозволяє паралельно виводити CNN на VPU SHAVES. Висновок було виконано як на VPU, так і на CPU. Поліпшення швидкості обробки на 64,59% було отримано, коли обробку виконували за допомогою VPU замість CPU.

У роботі також представлено програмну систему на основі глибокого навчання (MobileNet-SSD) і банки фільтрів Калмана. Систему порівнювали з іншими найсучаснішими методами машинного навчання (ACF, PCL-MUNARO) і глибокого навчання (DPOM, YOLOv3, YOLO-depth) для виявлення людей. Досягнуті результати відповідають поточному рівню техніки, маючи точність 81,43% і відкликання 80,6% у наборі даних коридору EPFL, тоді як у лабораторії EPFL як точність, так і відкликання були вище 87%.

Іншим важливим моментом системи є обробка кількох відеопотоків у реальному часі (+30 кадрів в секунду), підтримка до 12 потоків за допомогою апаратного забезпечення UpSquared2 за допомогою вбудованого VPU. На додаток до цього, система була розроблена таким чином, щоб бути портативною та легкою в маніпулюванні, як у налаштуванні внутрішніх параметрів програмного забезпечення, таких як налаштування області інтересу та порогу виявлення, так і в споживанні електроенергії системою. Маючи середнє споживання електроенергії 12 Вт, з піками до 15 Вт, під час запуску AI inference на VPU. Можливість живити систему за допомогою портативних батарей або відновлюваних систем забезпечує велику гнучкість розподілених систем відеоспостереження.

ПЕРЕЛІК ПОСИЛАНЬ

1. Heaton, J.B.; Polson, N.G.; Witte, J.H. Deep learning for finance: Deep portfolios. Appl. Stoch. Model. Bus. Ind. 2017,pp. 3–12.

2. Ching, T.; Himmelstein, D.S.; Beaulieu-Jones, B.K.; Kalinin, A.A.; Do, B.T.; Way, G.P.; Ferrero, E.; Agapow, P.M.; Zietz, M.; Hoffman, M.M.; et al. Opportunities and obstacles for deep learning in biology and medicine. J. R. Soc. Interface 2018,pp.11-15

3. Zheng, Q.; Zhao, P.; Li, Y.; Wang, H.; Yang, Y. Spectrum interference-based two-level data augmentation method in deep learning for automatic modulation classification. Neural Comput. Appl. 2020,pp. 1–23.

4. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A survey of deep neural network architectures and their applications. Neurocomputing 2017, pp. 11–26.

5. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision. In Science and Information Conference; Springer: Berlin/Heidelberg, Germany, 2019, pp. 128–144.

6. Räty, T.D. Survey on contemporary remote surveillance systems for public safety. IEEE Trans. Syst. Man, Cybern. Part C (Appl. Rev.) 2010, pp. 493–515.

7. Al-Nawashi, M.; Al-Hazaimeh, O.M.; Saraee, M. A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments. Neural Comput. Appl. 2017, pp. 565–572.

8. Ibrahim, S.W. A comprehensive review on intelligent surveillance systems. Commun. Sci. Technol. 2016, pp.33-40.

9. Gautam, K.; Thangavel, S.K. Video analytics-based intelligent surveillance system for smart buildings. Soft Comput. 2019, pp. 281–283

10. Yu, W.; Liang, F.; He, X.; Hatcher, W.G.; Lu, C.; Lin, J.; Yang, X. A survey on the edge computing for the Internet of Things. IEEE Access 2017,pp. 6, 69-74.

11. Santamaria, A.F.; Raimondo, P.; Tropea, M.; De Rango, F.; Aiello, C. An IoT Surveillance System Based on a Decentralised Architecture. Sensors 2019, pp.19-23.

12. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv 2017, pp.143-146.

13. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge computing: Vision and challenges. IEEE Internet Things J. 2016, pp 63–64.

14. Kristiani, E.; Yang, C.T.; Huang, C.Y. iSEC: An Optimized Deep Learning Model for Image Classification on Edge Computing. IEEE Access 2020, pp.23-26.

15. Agarwal, S.; Hervas-Martin, E.; Byrne, J.; Dunne, A.; Luis Espinosa-Aranda, J.; Rijlaarsdam, D. An Evaluation of Low-Cost Vision Processors for Efficient Star Identification. Sensors 2020, pp.76-79.

16. Brunetti, A.; Buongiorno, D.; Trotta, G.F.; Bevilacqua, V. Computer vision and deep learning techniques for pedestrian detection and tracking: A survey. Neurocomputing 2018, pp. 17–33.

17. Victoria Priscilla, C.; Agnes Sheila, S.P. Pedestrian Detection—A Survey. In Intelligent Computing Paradigm and Cutting-Edge Technologies; Jain, L.C., Peng, S.L., Alhadidi, B., Pal, S., Eds.; Springer: Cham, Seitzerland, 2020, pp. 349–358.

18. Bedo, J.; Sanderson, C.; Kowalczyk, A. An efficient alternative to svm based recursive feature elimination with applications in natural language processing and bioinformatics. In Australasian Joint Conference on Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2006; pp. 170–180.

19. Smith, N.; Gales, M. Speech recognition using SVMs. In Proceedings of the Advances in Neural Information Processing Systems, Orlando, FL, USA, 13–17 May 2022, pp. 11–12.

20. Zeng, C.; Ma, H. Robust head-shoulder detection by pca-based multilevel hog-lbp detector for people counting. In Proceedings of the 2020 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2020, pp. 206–207.

21. Li, C.; Guo, L.; Hu, Y. A new method combining HOG and Kalman filter for video-based human detection and tracking. In Proceedings of the 2020 3rd International Congress on Image and Signal Processing, Yantai, China, 16–18 October 2020, pp. 290–293.

22. Thombre, D.; Nirmal, J.; Lekha, D. Human detection and tracking using image segmentation and Kalman filter. In Proceedings of the 2019 International Conference on Intelligent Agent & Multi-Agent Systems, Chennai, India, 22–24 July 2019, pp. 1–5.

23. Sell, J.; O'Connor, P. The xbox one system on a chip and kinect sensor. IEEE Micro 2024, pp. 44–53.

24. Choi, B.; Meriçli, C.; Biswas, J.; Veloso, M. Fast human detection for indoor mobile robots using depth images. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2023; pp. 108–113.

25. Zhang, H.; Reardon, C.; Parker, L.E. Real-time multiple human perception with color-depth cameras on a mobile robot. IEEE Trans. Cybern. 2023, pp. 142–144.

26. Luna, C.A.; Losada-Gutiérrez, C.; Fuentes-Jiménez, D.; Mazo, M. Fast heuristic method to detect people in frontal depth images. Expert Syst. Appl. 2021, pp.16-24.

27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Commun. ACM 2017, pp. 84–90.

28. Tian, Y.; Luo, P.; Wang, X.; Tang, X. Deep learning strong parts for pedestrian detection. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–14 December 2015, pp. 190–191.

29. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–14 December 2015; pp. 140–148.

30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans. Pattern Anal. Mach. Intell. 2017, pp.113–114.

ДЕМОНСТРАЦІЙНІ МАТЕРІАЛИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕЛЕКОМУНІКАЦІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ
Кафедра Інформаційних систем та технологій

ПРЕЗЕНТАЦІЯ ДО МАГІСТЕРСЬКОЇ РОБОТИ

на тему:
«МОДЕЛЬ СИСТЕМИ ВІДЕОСПОСТЕРЕЖЕННЯ ЗІ ШТУЧНИМ ІНТЕЛЕКТОМ»

Студент: Дегтяр О.М.
Керівник: Полоневич О.В.

Київ 2025р.

МЕТА ТА ЗАВДАННЯ

Слайд 2

Мета роботи - реалізація інтелектуальної системи виявлення, відстеження та підрахунку об'єктів із застосуванням ШІ.

Об'єкт дослідження – процес виявлення об'єктів.

Предмет дослідження – інтелектуальні системи виявлення об'єктів.

Для виконання поставленої мети, у магістерській роботі розроблено та виконано наступні завдання:

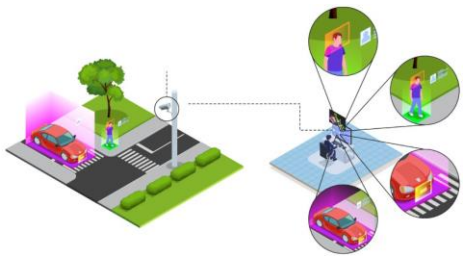
- огляд тенденцій розвитку та особливостей застосування штучного інтелекту у відеоспостереженні;

- аналіз архітектури глибоких нейронних мереж та особливості їх застосування;

- впровадження моделі системи відеоспостереження зі штучним інтелектом.

Застосування штучного інтелекту в системах відеоспостереження

Слайд 3



Інтеграція виявлення на основі морфології в камери відеоспостереження зі штучним інтелектом має вирішальне значення для мінімізації помилкових тривог, викликаних невідповідними об'єктами, такими як тварини, що тиняються на території, або природні елементи навколишнього середовища.

Ця технологія гарантує, що сповіщення генеруються виключно для виявлення цікавих, наприклад людей і транспортних засобів.

Рисунок 1 – Приклад застосування ШІ у відеоспостереженні

Потенціал штучного інтелекту в камерах відеоспостереження;

- виявлення та розпізнавання обличчя;
- підрахунок людей;
- метадані відео;
- розпізнавання номерних знаків (LPR);
- виявлення людей і транспортних засобів.

Порівняння традиційного відеоспостереження з штучних інтелектом

Слайд 4

Рисунок 2 – Контроль із використанням ШІ у відеоспостереженні

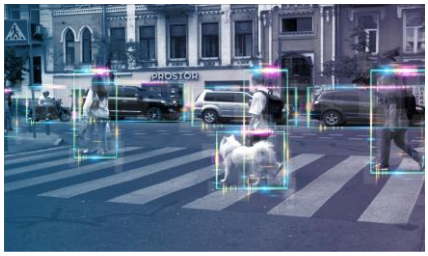


Рисунок 3 – Застосування ШІ в розпізнанні об'єктів

Характеристика	Традиційний підхід	ШІ-системи
Моніторинг	Залежний від людей	Працює 24/7
Зберігання	Локальне	Хмарне із автоматизацією
Точність	Можливі людські помилки	Висока аналітика без емоцій
Реакція	Затримки через людський фактор	Миттєві сповіщення
Інтеграція	Обмежена	Гнучка та масштабована
Недоліки	—	Високі початкові витрати, питання конфіденційності

Варіанти реалізації відеоспостереження з використанням штучного інтелекту

Слайд 5

Варіант реалізації	Особливості
ШІ в камері	Рентабельний спосіб почати використання ШІ. Обмежені можливості через апаратні ресурси та загальні моделі.
Аналітичні додатки ШІ	Доповнюють наявну систему відеобезпеки. Потребують додаткового обладнання та збільшують витрати.
Cloud Video Security (Хмарна безпека відео)	Сучасні функції для управління безпекою. Обмеження через вартість обробки, пропускну здатність, низька точність.
Уніфікована безпека відео зі ШІ	Інтегрує всі функції ШІ. Не залежить від конкретної камери. Забезпечує високу обчислювальну потужність та плавну інтеграцію.

Таблиця 1 – Варіанти реалізації відеоспостереження з використанням штучного інтелекту та їх особливості

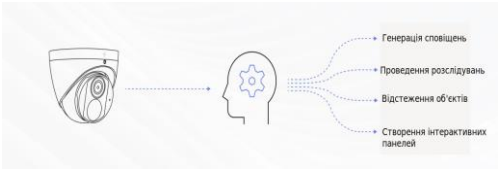


Рисунок 4 – Функціонування системи спостереження

Архітектури глибокого навчання із застосуванням обмежених машин Больцмана

Слайд 6

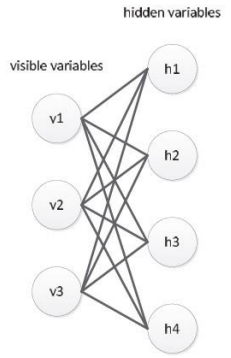


Рисунок 5 – Принципова схема RBMs

Input: RBM($V_1,...,V_m,H_1,...,H_n$), training period T, learning rate ϵ
Output: The RBM weight matrix ω, gradient approximation $\Delta\omega_{ij}$, Δa_i and Δb_j for $i = 1,...,m$, $j = 1,...,n$
Initialize ω with random values distributed uniformly in $[0,1]$, $\Delta\omega_{ij} = \Delta a_i = \Delta b_j = 0$ for $i = 1,...,m$, $j = 1,...,n$
For $\forall$ t = 1 : T do,
Sample $v_i^{(t)} \quad P(v_i|h^{(t)})$ when $i = 1,...,m$
Sample $h_j^{(t+1)} \quad P(h_j|v^{(t)})$ when $j = 1,...,n$
For $i = 1,...,m$, $j = 1,...,n$ do
$\Delta\omega_{ij} = \Delta\omega_{ij} + \epsilon \times [P(H_j = 1|v^{(0)}) \times v_i^{(0)} - P(H_j = 1|v^{(T)}) \times v_i^{(T)}]$
$\Delta a_i = \Delta a_i + \epsilon \times (v_i^{(0)} - v_i^{(T)})$
$\Delta b_j = \Delta b_j + \epsilon \times [P(H_j = 1|v^{(0)}) - P(H_j = 1|v^{(T)})]$
$\omega = \omega + \epsilon \times \Delta\omega_{ij}$
End For
End For

Рисунок 6 – Алгоритм 1 k-крокової контрастної дивергенції для RBM

- Обмежені машини Больцмана (RBM) — це стохастичні нейронні мережі, що використовуються для вивчення розподілу ймовірностей, класифікації, тематичного моделювання та зменшення розмірності.
- Навчання RBM здійснюється за допомогою алгоритму контрастної дивергенції (CD), що забезпечує ефективне оновлення ваг і наближення розподілу.
- Варіації RBM: CRBM для часових рядів, FE -RBM для класифікації, RoBM для роботи з шумами, TRBM з параметром температури для точнішої активації.
- RBM застосовуються в розпізнаванні зображень і мовлення, генеративних моделях, спільній фільтрації та задачах класифікації.
- Сучасні моделі (TRBM, RoBM) підвищують продуктивність RBM завдяки інноваційним алгоритмам та регульованим параметрам.

Архітектури глибокого навчання із застосуванням мереж з глибокими переконаннями

Слайд 7

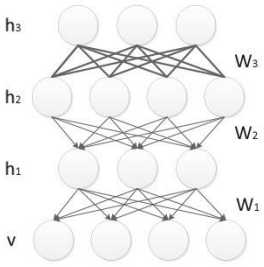


Рисунок 7- Принципова схема DBN

Input: Input visible vector v_{in}, training period T, learning rate ϵ, number of layers J.
Output: Weight matrix ω^i of layer $i, i = 1, 2, ..., J-2$.
Initialize ω with random values from 0 to 1, $h^0 = v_{in}$; where h^0 denotes the value of units in the input layer.
h^i represents the units' value of the ith layer. layer = 1;
for $\forall$ t = 1 : T do,
for $layer < L$, do gibbs sampling h^{layer} using $P(h^{layer}|h^{layer-1})$
Computing the CD in Algorithm 1, $\omega(layer)$ is achieved using $\omega(t+1) = \omega(t) + \epsilon \times \Delta\omega$
End for
End for

Рисунок 8 – Алгоритм 2, пошаровий алгоритм для DBN

- Глибокі вірувальні мережі (DBN) складаються з кількох шарів обмежених машин Больцмана (RBM) і використовуються для обробки немаркованих даних та класифікації.
- Алгоритм навчання DBN включає два етапи: неконтрольоване пошарове попереднє навчання (знизу вгору) та контрольоване тонке налаштування (зверху вниз).
- Жадібний пошаровий алгоритм забезпечує ефективне навчання DBN із високою продуктивністю завдяки покращенню ваг на основі вхідних даних.
- Варіації DBN: включають моделі, такі як DCN (глибокі опуклі мережі) для масштабованого навчання, та CDBN (згорткові мережі глибокого переконання) для аналізу багатовимірних зображень.
- DBN у практиці: використовуються в задачах розпізнавання зображень, обробки мовлення, діагностичних системах, пошуку зображень та складних завданнях класифікації.
- Переваги DBN: забезпечують високу точність навчання, ефективну обробку великих наборів даних та адаптацію до різних додатків завдяки інноваційним архітектурам.

Аналіз архітектури глибокого навчання із застосуванням автокодеру та на базі згорткової нейронної мережі

Слайд 8

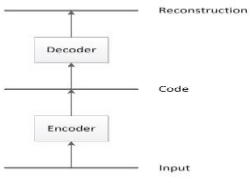


Рисунок 9 – Принципова схема AE

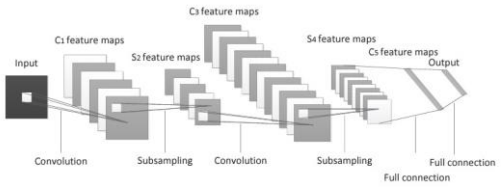


Рисунок 10 – Схематична структура CHC

- Автокодери (AE) використовуються для зменшення розмірності даних, фільтрації шуму (DAE), розріджених представлень та ефективного кодування з високою точністю реконструкції.
- Згорткові нейронні мережі (CNN) — це потужна архітектура для роботи з двовимірними даними (зображення, відео), що забезпечує ефективне навчання завдяки згортці, спільному використанню параметрів і пониженню дискретизації.
- Варіації CNN, такі як рекурсивні згорткові мережі (RCN) і CRBM, адаптовані для виділення ознак у складних задачах комп'ютерного зору, розпізнавання об'єктів і класифікації аудіо.
- Сучасні застосування CNN включають розпізнавання тексту, облич, поведінки, мови, системи рекомендацій та аналіз великих даних у реальному часі.

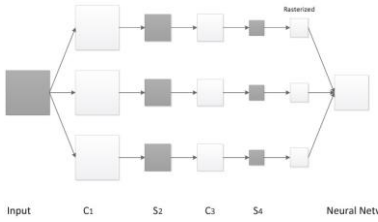


Рисунок 11 - Концептуальна структура CNN

Аналіз алгоритмів виявлення та відстеження об'єктів у системах відеоспостереження

Слайд 9

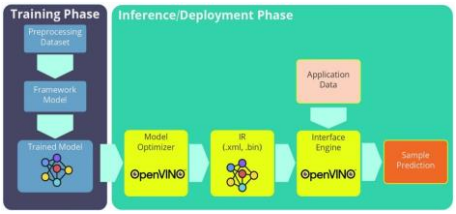


Рисунок 12 – Структура OpenVino

Вибір платформи UpSquared2 із Myriad X VPU забезпечує високу продуктивність і низьке енергоспоживання для обробки алгоритмів ШІ. Фреймворк OpenVino спрощує оптимізацію та розгортання нейронних мереж на різних апаратних платформах. Це рішення забезпечує гнучкість, портативність та ефективність у впровадженні інтелектуальних систем.

- Вбудованою платформою для інтелектуального вузла обрано UpSquared2, що оснащена процесоромIntel Atom x7-E3950, 8 ГБ RAM, 64 ГБeMMC та модулем Intel Movidius Myriad X VPU.
- Myriad X VPU забезпечує обробку висновків глибоких нейронних мереж (DNN) з мінімальним енергоспоживанням, виконуючи понад 4 трильйони операцій за секунду (TOPS).
- Архітектура VPU включає 16 128-бітних VLIW SHAVE процесорів та два LEON4 Core CPUs, що забезпечує ефективність на граничних вузлах та у батарейних системах.
- Для реалізації алгоритмів ШІ використовується фреймворк OpenVino, який оптимізує моделі CNN для різних платформ (CPU, VPU, GPU).
- Оптимізатор моделей OpenVino створює файли IR (.xml для архітектури та .bin для ваг) для прискорення висновків на системах.
- Inference Engine завантажує IR-файли, виконує висновки та балансує навантаження між ядрами або VPUs для оптимальної роботи.

Принципи впровадження моделі системи відеоспостереження на основі штучного інтелекту

Слайд 10

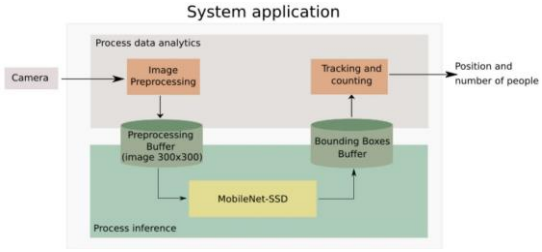


Рисунок 13 – Потік даних і обробка за допомогою багатопоточності

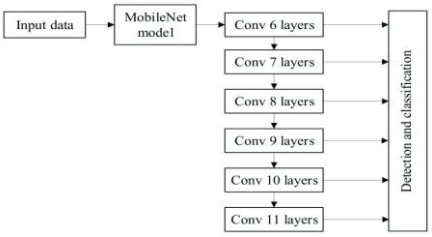


Рисунок 15– Масштабована обробка кількох відеопотоків у MobileNet-SSD на VPU

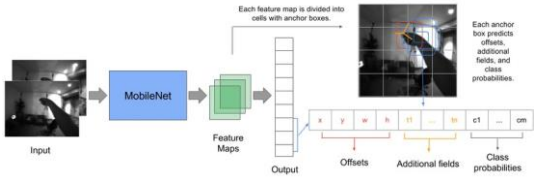


Рисунок 14– Архітектура MobileNet-SSD із глибинно-сепарабельними згортками

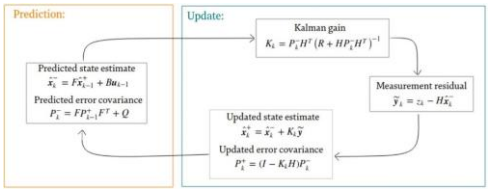


Рисунок 16– Схематична структура фільтра Калмана, що включає етапи прогнозування та оновлення.

Результати роботи системи тестування

Слайд 11

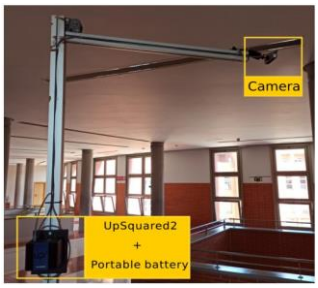


Рисунок 17– Встановлена тестова система

Система на базі UpSquared2 із MobileNet-SSD і Myriad-X VPU забезпечує високу продуктивність, обробляючи до 12 відеопотоків у реальному часі (30 fps) із мінімальною затримкою (8.71 мс на кадр).
Використання VPU дозволяє знизити енергоспоживання на 25% порівняно з виконанням виключно на CPU, забезпечуючи енергоефективність 17.08 W·ms на кадр.
У порівнянні з конкурентами, система демонструє значну перевагу у співвідношенні продуктивності до енерговитрат, роблячи її ідеальною для портативних відеосистем із низькими ресурсами.

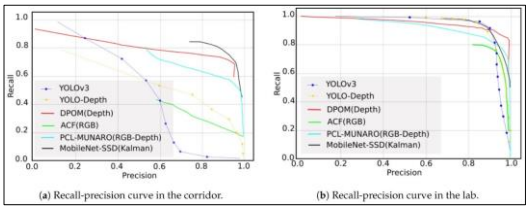


Рисунок 18– Криві точності та відгуку для різних середовищ у наборі EPFL

Рисунок 19– Обробка двох відеопотоків за допомогою CPU

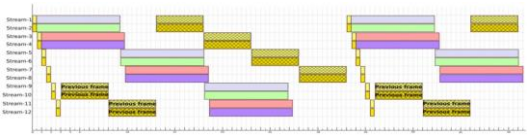


Рисунок 20– Обробка 12 відеопотоків із використанням CPU та VPU