

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

Факультет інформатики та обчислювальної техніки

Кафедра технічної кібернетики

«На правах рукопису»
УДК 004.093

«До захисту допущено»

Завідувач кафедри
_____ І.Р. Пархомей
(підпис)

“ _____ ” _____ 2019 р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 126 «Інформаційні системи та технології»

на тему: Система відеоспостереження розумного дому з інтелектуальним розпізнаванням загроз

Виконав: студент другого курсу, групи ІК-81мп
(шифр групи)

_____ Петрунів Орест Романович _____

(прізвище, ім'я, по батькові)

(підпис)

Науковий керівник д.т.н., доцент Борукаєв З.Х.

(посада, науковий ступінь, вчене звання, прізвище та ініціали)

(підпис)

Консультант

(назва розділу)

(науковий ступінь, вчене звання, прізвище, ініціали)

(підпис)

Рецензент

(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали)

(підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.

Студент _____
(підпис)

Київ – 2019 року

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

Факультет інформатики та обчислювальної техніки

Кафедра технічної кібернетики

Рівень вищої освіти – другий (магістерський)

Спеціальність 126 «Інформаційні системи та технології»

ЗАТВЕРДЖУЮ

Завідувач кафедри

І.Р. Пархомей

(підпис)

« ___ » _____ 2019 р.

ЗАВДАННЯ

на магістерську дисертацію студенту

Петруніву Оресту Романовичу

1. Тема дисертації Система відеоспостереження розумного дому з інтелектуальним розпізнаванням загроз, науковий керівник дисертації Борукаєв З.Х. доцент, к.т.н., с.н.с. ,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від « ___ » _____ 20__ р. № _____

2. Термін подання студентом дисертації

3. Об'єкт дослідження - процес автоматизованого розпізнавання загроз на відео.

4. Предмет дослідження - моделі та методи інформаційних технологій для розпізнавання загроз.

5. Перелік завдань, які потрібно розробити - аналіз проблеми та існуючих рішень; аналіз і реалізація методу; розробка додатку; дослідження ефективності розробленого додатку.

6. Орієнтовний перелік ілюстративного матеріалу -

7. Орієнтовний перелік публікацій - одна публікація

8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Перевірка на співпадіння	Лісовиченко О.І., доцент		
Н.К.	Пасько В.П., доцент		

9. Дата видачі завдання

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Аналіз предметної області	13.09.2019 р.	
2	Постановка задачі	15.09.2019 р.	
3	Аналіз інформаційного забезпечення	20.09.2019 р.	
4	Аналіз алгоритмічного забезпечення	25.09.2019 р.	
5	Розробка алгоритмічного забезпечення	15.10.2019 р.	
6	Розробка програмного забезпечення	01.11.2019 р.	
7	Маркетинговий аналіз стартап-проекту	10.11.2019 р.	
8	Висновки	15.11.2019 р.	

Студент

_____ (підпис)

О.Р. Петрунів

(ініціали, прізвище)

Науковий керівник дисертації

_____ (підпис)

З.Х. Борукаєв

(ініціали, прізвище)

АНОТАЦІЯ

У роботі розглянуто проблему в області автоматизованого розпізнавання загроз, показано основні особливості існуючих рішень та додатків, їх переваги та недоліки.

При забезпеченні певного об'єкту відеоспостереженням, недостатньо просто зберігати отримані дані, необхідно активно розпізнавати ситуації які розгортаються перед камерами. Людина схильна до помилок, стомлення, втрати концентрації, а коли мова йде про розподілену систему камер - задача розпізнавання стає непосильною. Тут на допомогу приходять інтелектуальні системи, здатні забезпечувати автоматичне розпізнавання потенціальних загроз.

Визначено завдання для системи інтелектуального розпізнавання загроз у відеоряді за допомогою технологій нейронних мереж, відібрано мережі для виконання основних задач такої системи та способи їх навчання, які найбільш підходять для даної задачі. Описано структури нейронних мереж та проведено експерименти по навчанню та їх роботі.

Система забезпечує автоматичне розпізнавання ситуацій на основі відеоряду в режимі реального часу. Дозволяє зменшити витрати коштів на працівників та збільшити ефективність роботи охоронної системи.

Ключові слова: розпізнавання діяльності, відео аналітична система, нейронна мережа.

Розмір пояснювальної записки – 85 аркушів, містить 15 ілюстрацій, 25 таблиць, 2 додатки.

ABSTRACT

The paper considers the problem in the field of automated threat recognition, shows the main features of existing solutions and applications, their advantages and disadvantages.

When providing video surveillance for some areas, it is not enough to simply store the data received, it is necessary to actively recognize the situations that unfold in front of the cameras. A person is prone to mistakes, fatigue, loss of concentration, and when it comes to a distributed camera system, the task of recognition becomes impossible for a human. This is where intelligent systems capable of automatically detecting potential threats come to the rescue.

The tasks for the system of intelligent detection of threats in the video series with the help of neural network technologies are determined, the networks for tackling major tasks of such system are selected, ways of their training explain. Architecture of neural networks are described and experiments on training and their work are conducted.

The system provides automatic recognition of situations on the basis of video sequence in real time.

Key words: activity recognition, video analytics, neural network.

Explanatory note size – 85 pages, contains 15 illustrations, 25 tables, 2 applications.

**Пояснювальна записка
до магістерської дисертації**

на тему: Система відеоспостереження розумного дому з
інтелектуальним розпізнаванням загроз

Київ – 2019 року

ЗМІСТ

ВСТУП	4
РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ	6
1.1 Об'єкт та предмет дослідження	6
1.2 Огляд існуючих рішень	6
1.2.1. Комплексні системи відеоспостереження	7
1.2.2. Розумні камери	9
1.3 Постановка задачі	11
Висновки по розділу	12
РОЗДІЛ 2. АНАЛІЗ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ	13
2.1 Задачі та алгоритми	13
2.2 Виявлення об'єктів	13
2.2.1 Гаусова суміш розподілень (GMM)	16
2.2.2 Приблизний медіанний фільтр (AMF)	16
2.2.3 Метод оптичного потоку (OF)	17
2.3. Розпізнавання та відслідковування об'єктів	17
2.3.1 Нейронні мережі	20
2.3.2 Згорткові нейронні мережі (CNN)	20
2.3.3 Regional CNN (R-CNN)	22
2.3.4 Швидка R-CNN	24
2.3.5 Швидша R-CNN	25
2.3.6 You Only Look Once (YOLO)	26
2.3.7 Рекурентні нейронні мережі (RNN)	27
2.3.8 Мережі із довгою короткочасною пам'яттю (LSTMs)	29
2.4. Аналітика поведінки	31
2.5 Обґрунтування вибору програмних засобів	38
2.5.1 Python	38
2.5.2 Tensorflow	41
Висновки по розділу	44

РОЗДІЛ 3. РОЗРОБКА АЛГОРИТМІЧНОГО ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ	45
3.1 Двопотоківна архітектура для розпізнавання відео	45
3.2 Застосування оптичного потоку для CNN	46
3.2.1 Налаштування вхідного потоку для CNN	47
3.3 Багатозадачне навчання	49
3.4 Деталі реалізації	50
3.5 Оцінка	53
Висновки по розділу	55
РОЗДІЛ 4. МАРКЕТИНГОВИЙ АНАЛІЗ СТАРТАП-ПРОЕКТУ	56
4.1 Опис ідеї проекту	56
4.2 Технологічний аудит ідеї проекту	57
4.3 Аналіз ринкових можливостей запуску стартап-проекту	57
4.4 Розроблення ринкової стратегії проекту	65
4.5 Розроблення маркетингової програми стартап-проекту	67
Висновки по розділу	70
ВИСНОВКИ	71
ПЕРЕЛІК ПОСИЛАНЬ	74
ДОДАТКИ	75

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БД – База даних

СУБД – Система управління базами даних

HTML – HyperText Markup Language

ПЗ – Програмне забезпечення

API – Application programming interface

SQL – Structured Query Language

UML – Unified Modeling Language

CPU – Central Processing Unit

FC – Fully Connected

GPU – Graphics Processing Unit

RoI – Агрегування областей інтересу

ВСТУП

У області забезпечення безпеки критично важливих об'єктів від терористичних загроз існує потреба в поліпшенні показників застосовуваних для цих цілей систем [1]. За наявними прогнозами слід очікувати посилення інтенсивності протиправних дій в техносфері в найближчій перспективі [2].

На сьогоднішній день ефективним заходом протидії несанкціонованому доступу до охоронюваним об'єктам вважається цілодобова озброєна охорона спеціально створеними підрозділами служб безпеки [3]. Спостерігається інтенсивне наукове опрацювання напрямів з модернізації інженерно-технічних засобів запобігання несанкціонованому проникненню осіб до захищених об'єктів. За минулі 4 роки в базі даних Web of Science зареєстровано 80 публікацій, присвячених цьому питанню, у Scopus – 55, в РИНЦ – 66. Патентними відомствами Європи, США видано відповідно 40, 13 охоронних документа, націлених на забезпечення безпеки таких об'єктів. Сформувалася точка зору, що успіх у вирішенні означеної проблеми буде досягнуто при переході на мультисенсорні системи [4], так як в кожному із запропонованих підходів існують принципові обмеження щодо досягнення прийнятної надійності виявлення цікавить категорії осіб і розпізнавання намірів вторгся суб'єкта – необхідні умови для забезпечення безпеки критично важливих об'єктів [5]. Зазначені особливості завдання ініціювали дослідження щодо включення до складу мультисенсорних комплексів відеоаналітичних систем [6].

Відеоаналітикою називають технологію комп'ютерного аналізу відеоданих з метою прийняття рішень про стан об'єкта спостереження без участі

оператора. Її розвиток стимулюється результатами неефективної роботи людини в режимі безперервного дешифрування динамічних зображень з метою виявлення на них вакансій (як правило, незначних) змін [7, 8, 9]. В експерименті після 12 хвилин роботи з пошуку локальних об'єктів на тлі перешкод оператор пропустив 45% з числа пред'явлених [10].

З огляду на представлені дані, стає очевидною необхідність аналіз існуючих систем інтелектуального розпізнавання загроз, пошук способів збільшення їх ефективності і розробка нових, покращених рішень.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ

1.1. Об'єкт та предмет дослідження

Забезпечення безпеки людей і матеріальних цінностей – проблема яка була і буде актуальною для людства. Відеоспостереження – це відносно новий інструмент, використовуваний для її вирішення, та темпи його розвитку не можуть не вражати. Один із ключових факторів його розвитку – інтеграція методів інформаційних технологій напрямлених на розпізнавання образів. Результат їх об'єднання можна побачити вже сьогодні: системи розпізнавання облич та дій, ідентифікація окремих осіб у натовпах, робота розподіленої системи камер як одного цілого. Об'єктом цієї роботи є процес автоматизованого розпізнавання загроз на відео, а предметом – моделі та методи інформаційних технологій для розпізнавання загроз.

1.2. Огляд існуючих рішень

Загальною ідеєю існуючих рішень є обробка відеопотоку шляхом застосування набору алгоритмів, які можуть вирішувати наступні задачі: збільшення відношення сигнал/шум, визначення відстані до об'єктів, розпізнавання об'єктів, визначення здійснюваних особами дій.

Пропоновані ринком продукти можна розділити у дві категорії, комплексні системи відеоспостереження і розумні камери, розглянемо їх детальніше і наведемо описи типових представників.

1.2.1. Комплексні системи відеоспостереження

Рішення, описані в цьому розділі, присвячені комерційним системам, які передбачають автоматичне виявлення визначених подій на основі обробки відео. Це може бути придбано або програмним забезпеченням, яке постачається, або підтримувати його деяким обладнанням. Загалом такі системи включають кілька основних пристроїв, тобто для збору, перетворення та зберігання даних. Додатково програмне забезпечення кінцевого користувача постачається або як веб-інтерфейс, або, як правило, як спеціалізована програма, що спрощує обслуговування та / або управління системою. Обробка відео реалізована на основі передових технологій, розроблених кожною компанією. У деяких випадках пропонуються пристрої з чорним ящиком, де підключається кожна камера (одна або більше). Як альтернатива, виділяються спеціалізовані високопродуктивні сервери для підтримки аналізу декількох відеопотоків. Описи декількох обраних продуктів наведені в наступних підрозділах. Вони включають загальні характеристики системи, а також різні конфігурації системи.

Типовими недоліками систем цього типу є вартість, необхідність встановлення великої кількості додаткового обладнання, неможливість додавання компонентів від сторонніх розробників, закритий програмний код.

NVUO

NVUO PC NVR (IP +) – це гнучке та економічно вигідне рішення, яке може керувати камерами IP / Megapixel. NVUO NVR використовує технологію відкритої платформи, яка підтримує до 52 брендів IP / Megapixel камер з більш ніж 1100 моделями. Інші зручні інструменти включають інтуїтивно зрозумілий графік запису GUI, Е-карту, засоби покращення відео та мобільну підтримку (3G).

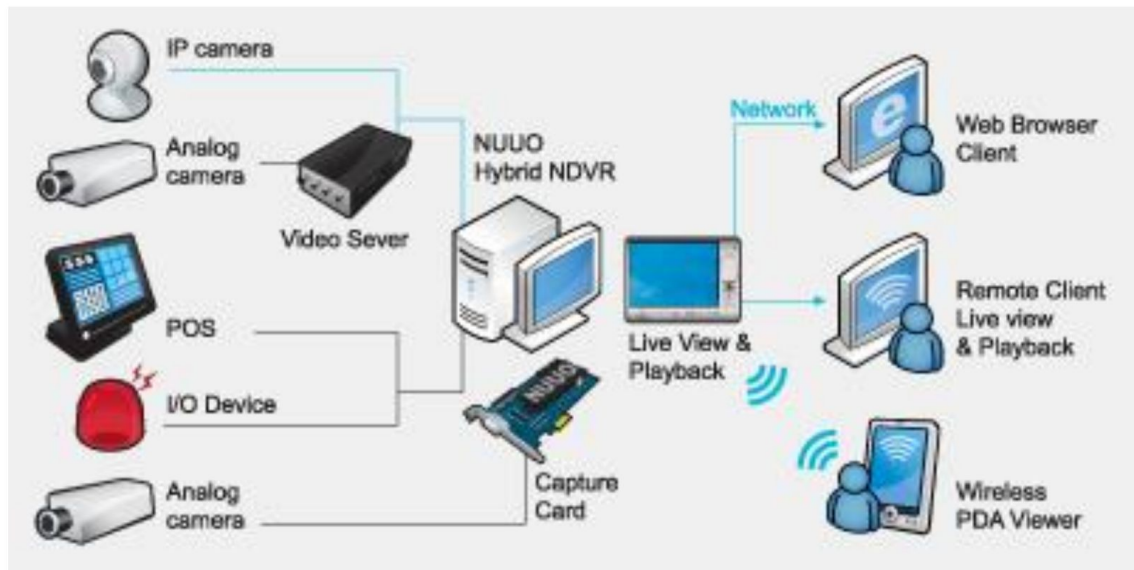


Рисунок 1.1 – Архітектура системи NUUO

Інтеграція пристроїв POS та I / O також не вистачає поставок. NUUO NVR підтримується автентичною центральною системою управління NUUO, яка є справжнім рішенням для моніторингу та управління, яке підтримує необмежену кількість камер. Основна консоль – це сервер запису NUUO NVR. Він може відображати відео в прямому ефірі та налаштовувати систему. NUUO NVR має інтуїтивний дизайн, який можна дізнатися дуже швидко. Основні функції включають наступні:

- Потокове передавання декількох відео в реальному часі з IP-камер та IP-відеосерверів.
- Унікальний графік запису графічного інтерфейсу.
- 6 виявлень подій та 10 миттєвих реакцій тривоги (наприклад, повідомлення електронною поштою та спливаюче вікно E-map).
- Повторне відображення каналу з цифровим PTZ.

Основна консоль NUUO може відображати відео в реальному часі та налаштовувати диктофон. Система відтворення NUUO робить інтелектуальний і швидший перегляд записів за допомогою інтелектуального пошуку. Збереження

відео перетворює зображення у формат ASF / AVI. Інструмент покращення відео можна використовувати для різкості, яскравості або навіть сірого масштабу зображення. Різні журнали ведуть облік усіх подій. Проста у використанні багатоканальна програма резервного копіювання для архівації аудіо та відео локально або віддалено. Система резервного копіювання NUUO також може робити знімки записів і зберігати їх в окремому місці.

1.2.2. Розумні камери

Іншим підходом до застосування відеоаналітики в системах спостереження є використання так званих смарт-камер. Розумна камера – це звичайна камера, оснащена додатковим модулем DSP, де виконується обробка відео. Ці два елементи, як правило, інтегровані всередині одного корпусу. Такі камери можна використовувати у вже існуючих системах спостереження, що пропонують оператору людини додаткову інформацію, отриману з модуля відеоаналітики. Часто поряд з цим рішенням виробники постачають спеціальне програмне забезпечення для камер з інтерфейсом, що дозволяє спілкуватися з користувачем. Таким чином оператор може бути адекватно оповіщений, відображаючи додаткову інформацію про поточний відеопотік.

Типовими недоліками систем даного типу є обмежений функціонал, неможливість розширення програмних і апаратних засобів, обробка даних у корпоративних “хмарах”

Samsung Techwin

Samsung Techwin, провідна у світі технологія візуалізації, відіграє важливу роль у захисті безпеки та щастя людей, забезпечуючи комплексний асортимент продуктів та комплексних рішень, починаючи від спостереження за містом,

захищаючи вулиці, аеропорти, порти, промислові об'єкти, військові установки та В2С.



Рисунок 1.2 – Екран модулю відслідковування Samsung Techwin

Samsung Techwin sets – новий орієнтир на внутрішньому та міжнародному ринку безпеки, забезпечуючи більш високу якість, більш чисті зображення та найсучасніші мережеві функції. Samsung Techwin має на меті забезпечити безпечне рішення one.stop, яке задовольнятиме потреби користувачів і зараз, і в майбутньому, і прагне стати світовим провідним постачальником професійних рішень з безпеки.

Переваги:

- Відео у розширенні 1080р.
- Миттєві сповіщення про активність.
- Двостороннє аудіо.
- Робота у темряві.
- Можливість налаштування зон активності.

1.3 Постановка задачі

Метою роботи є забезпечення безпеки середовища розумного дому шляхом впровадження інтелектуального розпізнавання загроз при відеоспостереженні.

Для досягнення мети необхідно вирішити наступні задачі:

1. визначення типів задач, які повинна виконувати система, аналіз методів їх вирішення;
2. розробка системи розпізнавання загроз шляхом інтеграції компонентів які реалізують окремі задачі в єдине інтелектуальне рішення ;
3. реалізація інтелектуальної системи розпізнавання загроз на основі відеоспостереження.

Створений продукт повинен відповідати наступним вимогам:

- розпізнавання загроз у режимі реального часу;
- відкритість рішення, можливість його видозміни, розширення/доповнення сторонніми продуктами;
- багатоплатформність;
- простота у налагодженні.

Висновки по розділу

У першому розділі проведено аналіз існуючих рішень та предметної області, проаналізовані основні сучасні розробки, визначено переваги та недоліки цих рішень. В результаті проведеного аналізу сформульована постановка задачі, наведене призначення, цілі та задачі розробки.

РОЗДІЛ 2. АНАЛІЗ ІНФОРМАЦІЙНОГО ЗАБЕЗПЕЧЕННЯ

2.1 Задачі та алгоритми

Перед системою інтелектуального розпізнавання загроз стоїть декілька комплексних задач, для вирішення яких необхідне застосування різноманітних методів та алгоритмів. Розглянемо їх детальніше.



Рисунок 2.1 – Традиційний етапи обробки у інтелектуальній системі відеоспостереження.

2.2 Виявлення об'єктів

Ідентифікація рухомого об'єкта є основоположним і критичним завданням у різних програмах комп'ютерного зору. Сегментація зображень із відеопотоку – один із найважливіших кроків у візуальній обробці, кодуванні відео та застосуванні відеоспостереження. Загальний підхід до витягування рухомих об'єктів із відеопотоку здійснюється за допомогою техніки віднімання фону. Для реалізації алгоритму віднімання фону фонове моделювання завжди є найважливішим кроком, який створює опорний кадр, містить лише фонові об'єкти. Однак завдання стає складним, коли фон містить тіні і рухомі об'єкти. У цьому випадку нам потрібно розробити фонову модель, яка приймає будь-які зміни, що відбулися на тлі сцени. Фонова модель повинна бути адаптивною для будь-яких варіантів, що відбуваються у фоновому режимі. Перед побудовою

гарної моделі фону виникло багато проблем. Вона повинен враховувати зміни, що відбуваються на тлі, наприклад, руху дерев і переміщення прапорів і т. Д. У зовнішніх умовах техніка віднімання фону повинна набувати змін, що відбулися при освітленні світла в різний час доби, і обробляти несприятливу погоду, наприклад дощ, сніг тощо. Інакше продуктивність фонового моделювання буде погіршуватися. Щоб цього уникнути, нам потрібно побудувати техніку адаптивного фонового моделювання. Метод віднімання фону, який використовується в системі відеоспостереження в реальному часі, повинен бути обчислювально недорогим, низьким рівнем використання пам'яті та здатним точно виявляти переміщення об'єкта зі сцени. Алгоритм віднімання фонового зображення може бути виконаний за допомогою чотирьох основних етапів, таких як передстадійна обробка, фонове моделювання, об'єкт виявлення та перевірка даних. Необхідна обробка попередньої стадії для зменшення шуму камери та шуму навколишнього середовища за допомогою ефекту згладжування. Другий крок – фонове моделювання, при якому новий кадр створюється за допомогою перших кількох кадрів відеопослідовності для представлення лише фонових об'єктів. Дослідження технології комп'ютерного зору продовжують розробляти фонову модель, більш адаптивну до змін, що відбулися в положенні камери, а також у навколишньому середовищі. Досить чутливим є визначення всіх відповідних пікселів, які представляють собою повний об'єкт переднього плану, а нестационарні фонові об'єкти не можуть бути виявлені як об'єкти переднього плану. На етапі виявлення об'єкта всі вхідні кадри з відеопослідовності віднімаються з фонові моделі. Потім різницю порівнюють з пороговим значенням для отримання бінарного зображення на виході. Під час перевірки даних можна проводити морфологічну операцію для видалення всіх випадкових пікселів шуму та заповнення проміжку між виявленими об'єктами

для підвищення точності. Дві основні категорії методів фонового моделювання є рекурсивними та не рекурсивними. У нерекурсивній техніці фоновий кадр оцінюється за допомогою підходу з розсувним вікном. Перш за все, декілька кадрів з відеопослідовності зберігаються в буфері та оцінюють фонову модель на основі часової зміни кожного пікселя в буфері. Він є дуже адаптивним як фоновий кадр і залежить лише від збережених зображень у буфері, але ціною великої потреби в пам'яті. Деякі популярні нерекурсивні методи – це медіаний фільтр, розрізнення кадру, непараметрична модель та лінійний прогностичний фільтр. У рекурсивних методах замість підтримки буфера для фонової оцінки вони рекурсивно оновлюють єдину фонову модель на основі кожного вхідного кадру. Ці методи вимагають менше місця для зберігання порівняно з нерекурсивними методами.



Рисунок 2.2 – Приклад техніки віднімання фону у виявленні руху

Усі кроки попередньої обробки, такі як виявлення об'єктів та класифікація для аналізу поведінки, важливі у застосуванні відеоспостереження. Виявлення об'єкта переднього плану за допомогою модельного фонового моделювання

здійснюється за допомогою методу віднімання фону. Віднімання фону порівнює поточний кадр відео з кадром відліку. Різниця пікселів порівнюється із визначеним користувачем пороговим значенням для вилучення об'єкта переднього плану, що рухається. Якщо різниця більша за поріг, то відповідний піксель належить об'єкту переднього плану, а якщо він менший за поріг, то пікселі належать до фонового об'єкта. Таким чином, все зображення з відео сегментується на передній план та фонові об'єкти у вигляді двійкової маски на виході.

Розглянемо найбільш популярні методи видалення фону.

2.2.1 Гаусова суміш розподілень (GMM)

GMM – це один з дуже популярних і широко використовуваних підходів для віднімання фону у програмах візуального спостереження у випадку нерухомих камер. Для всіх пікселів у кадрі підтримується суміш Гаусса. З плином часу нові значення пікселів оновлюють суміш гауссів, використовуючи підхід К-засобів. Оцінка фону оновлюється з урахуванням змін освітленості, шуму навколишнього середовища, незначних рухів датчика та сильної оклюзії. Проте дослідники з нагляду продовжують наголошувати на важливості адаптивних та надійних методів моделювання фону. Складаючи комбінацію різних методик моделювання, точність виявлення можна вдосконалити.

2.2.2 Приблизний медіанний фільтр (AMF)

У цьому способі перші кілька кадрів послідовності відеозаписів зберігаються в буфері. Тоді ми можемо оцінити фонову модель з медіани буферних кадрів. Матриця 3x3 часто використовується для заміни центрального пікселя медіаною всіх пікселів сусідства. Шаблон вікна ковзає по всьому

зображенню. Центральний піксель представляє оброблений піксель. Цей підготовлений фоновий кадр віднімається від поточного поточного кадру відеопослідовності, щоб отримати піксель переднього плану. Він може порівняти значення пікселя з поточного кадру з значенням відповідного пікселя з фонового кадру. Якщо значення пікселя з поточного кадру більше, ніж відповідний фоновий піксель, воно збільшується на одиницю. Цей метод може призначити різну вагу для кожного пікселя залежно від зміни інтенсивності пікселя. Таким чином, наблизений середній фільтр дає більш точний вихід у порівнянні з простим медіанним фільтром у разі динамічної сцени.

2.2.3 Метод оптичного потоку (OF)

Це векторний підхід, який оцінює рух, порівнюючи точку на об'єктах у кількох кадрах. Через складність алгоритму для реальної реалізації оптичного потоку в реальному часі необхідно використовувати спеціалізоване обладнання. Перевага методики оптичного потоку полягає в тому, що вона є надійною для багаторазового та одночасного руху камери, що робить її ідеальною для аналізу натовпу та щільного руху.

2.3. Розпізнавання та відслідковування об'єктів

Технології відстеження можна розділити на два основні підходи: 2-D моделі з чіткими фігурами або без моделей і 3-D моделі. Наприклад, у [11] 3-D геометричні моделі автомобіля, фургона та вантажного автомобіля використовуються для відстеження транспортні засоби на шосе. На основі модельного підходу використовуються явні апріорні геометричні знання про об'єкти, які слід слідкувати, які зазвичай у програмах спостереження - це люди, транспортні засоби або обидва. У [12] автор використовує дві двовимірні моделі

для відстеження автомобілів: прямокутну модель для проїзного автомобіля, близьку до камери, і модель U-образної форми для задньої частини автомобіля на відстані або просто перед камери. Система складається з модуля збору зображень, детектора смуг руху та автомобіля, координатора доступу та декількох автомобільних трекерів. У деяких системах камер на зразок [13] основна увага приділяється вилученню траєкторій, які використовуються для побудови геометричної та ймовірнісної моделі для довгострокового прогнозування, а не самого об'єкта. Апріорні знання можна отримати, обчисливши вигляд об'єкта як функцію його розташування відносно камери. Геометрія сцени виходить таким же чином. Для побудови моделей фігур важливе значення має використання методів калібрування камери. Огляд різних методик калібрування камери можна знайти в [14]. Після того, як пріоритетне знання буде доступне, воно може бути використане в надійному алгоритмі відстеження, що стосується різноманітних умов, таких як зміна освітленості, пропонуючи більш високу продуктивність при вирішенні (само) оклюзій або (само) зіткнень. Створити обмеження в моделі появи об'єктів порівняно просто, використовуючи підходи, засновані на моделі; напр. обмеження в тому, що люди здаються вертикально і контактують із землею, зазвичай застосовуються для внутрішніх та зовнішніх середовищ.

Завдання розпізнавання об'єктів потім стає процесом використання методів, заснованих на моделі, у спробі використання таких знань. Для класифікації нових виявлених об'єктів можна застосувати ряд підходів. Інтегрована система, представлена в [15] та [16], може розпізнавати та відслідковувати транспортні засоби, використовуючи визначену 3-D модель транспортного засобу, задаючи його положення в земній площині та орієнтацію. Він також може розпізнавати та відстежувати пішоходів, використовуючи

попередню 2-D модель силуету моделі, засновану на контурах B-сплайна. Поширений метод відстеження – використовувати механізм фільтрації для прогнозування кожного руху розпізнаного об'єкта. Фільтр, який найчастіше використовується в системах спостереження, – це фільтр Кальмана [53, 73]. Встановлення обмежувальних коробок або еліпсів, які зазвичай називають «краплями», для зображень областей з максимальною ймовірністю – це ще один підхід відстеження, заснований на статистичних моделях. У роботі [16] автор моделює та відстежує різні частини людського тіла за допомогою крапель, які статистично описуються просторовим та кольоровим гауссовим розподілом.

У деяких ситуаціях, що представляють інтерес, припущення щодо застосування лінійних або гауссових фільтрів не виконують, і тоді були запропоновані нелінійні фільтри Байєса, такі як розширені фільтри Калмана (ЕКФ) або фільтри для частинок. Робота, описана в роботі [15], ілюструє, що у сильно нелінійних середовищах фільтри для частинок дають кращі показники, ніж ЕКФ. Фільтр частинок – це числовий метод, який зважає (або "частинку") подання задньої щільності ймовірності шляхом перекомпонування набору випадкових вибірок, пов'язаних з вагою, та обчислення ймовірностей оцінки на основі цих ваг. Потім критичне дизайнерське рішення з використанням фільтрів частинок покладається на вибір важливості (початкової ваги) функції щільності.

Інший підхід відстеження полягає у використанні підключених компонентів [17] для сегментації змін сцени на різні об'єкти без будь-якого попереднього знання. Підхід дає хороші показники, коли об'єкт невеликий, з наближенням низької роздільної здатності, і розміщення камери вибирається ретельно. ГММ (приховані марковмоделі) також використовувались для відстеження цілей, представлених у [18], де автори використовують розширення НММ для прогнозування та відстеження траєкторій об'єктів. Хоча фільтри НММ

підходять для динамічних середовищ (оскільки немає припущення в моделі або в характеристиці типу шуму, як це потрібно при використанні фільтрів Кальмана), необхідні дані офлайн-навчання. створення напівавтоматичних інструментів, які можуть допомогти створити великий набір даних основної істини, необхідних для оцінки продуктивності алгоритмів відстеження.

2.3.1 Нейронні мережі

Нейронні мережі – це набір алгоритмів, які дуже нагадують людський мозок і призначені для розпізнавання патернів. Вони інтерпретують сенсорні дані за допомогою машинного сприйняття, маркування або кластеризації вихідних даних. Вони можуть розпізнати числові зразки, що містяться у векторах, в які повинні бути переведені всі дані реального світу (зображення, звук, текст чи часовий ряд). Штучні нейронні мережі складаються з великої кількості високо взаємопов'язаних обробних елементів (нейронів), що працюють разом для вирішення проблеми

ANN, як правило, включає велику кількість процесорів, які працюють паралельно та розташовані в ярусах. Перший рівень отримує необроблену вхідну інформацію – аналогічну зоровим нервам при зоровій обробці людини. Кожен наступний ярус отримує вихід із рівня, що передує йому, а не від необробленого входу – таким же чином нейрони, що знаходяться далі від зорового нерва, отримують сигнали від тих, хто до нього ближче. Останній ярус виробляє вихід системи.

2.3.2 Згорткові нейронні мережі (CNN)

Згорткові нейронні мережі (CNN) – це найпопулярніша модель нейронної мережі, яка використовується для проблеми класифікації зображень. Головною

ідеєю CNN є те, що місцеве розуміння зображення є досить хорошим. Практична перевага полягає в тому, що менша кількість параметрів значно покращує час, необхідний для навчання, а також зменшує кількість даних, необхідних для навчання моделі. Замість повністю пов'язаної мережі ваг від кожного пікселя, CNN має достатньо ваг для перегляду невеликого виправлення зображення. Це як читати книгу за допомогою лупи; Врешті-решт, ви читаєте всю сторінку, але ви переглядаєте лише невелику частину сторінки в будь-який момент часу.

CNN-архітектура натхненна організацією та функціональністю зорової кори та створена для імітації схеми зв'язку нейронів всередині людського мозку. Нейрони всередині CNN розбиті на тривимірну структуру, при цьому кожен набір нейронів аналізує невелику область або особливість зображення. Іншими словами, кожна група нейронів спеціалізується на виявленні однієї частини зображення. CNN використовують прогнози від шарів для отримання остаточного результату, який представляє вектор балів ймовірності, щоб представити ймовірність того, що певна особливість належить певному класу.

Різниця між алгоритмами виявлення об'єктів і алгоритмами класифікації полягає в тому, що в алгоритмах виявлення ми намагаємося намалювати обмежувальне поле навколо об'єкта, що цікавить, щоб знайти його в межах зображення. Крім того, вам може не обов'язково намалювати лише одне обмежувальне поле у випадку виявлення об'єктів, у зображенні може бути багато обмежувальних полів, що представляють різні об'єкти, що представляють інтерес, і ви не знаєте, скільки заздалегідь. Основна причина, чому ви не можете продовжувати цю проблему, будуючи стандартну згорткову мережу, за якою повністю з'єднаний шар, полягає в тому, що довжина вихідного шару є змінною – не постійною, це тому, що кількість зустрічаються об'єктів, що цікавлять, становить не виправлено.

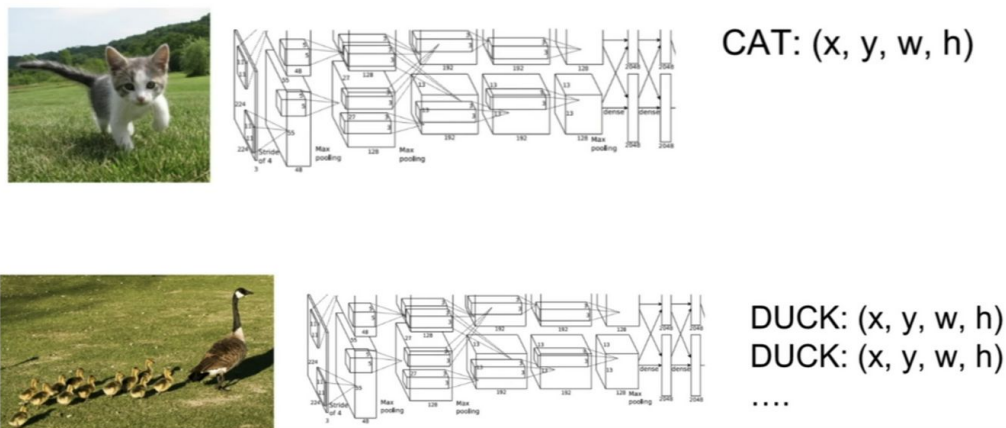


Рисунок 2.3 – Приклади CNN для задач класифікації і виявлення об'єктів

Наївним підходом до вирішення цієї проблеми було б відібрання різних регіонів, що цікавлять зображення, та використання CNN для класифікації присутності об'єкта в цьому регіоні. Проблема такого підходу полягає в тому, що об'єкти, що цікавлять, можуть мати різні просторові розташування в межах зображення та різні співвідношення сторін. Отже, вам доведеться вибрати величезну кількість регіонів, і це може обчислювально підірватись. Тому такі алгоритми, як R-CNN, YOLO тощо, були розроблені для пошуку цих подій та їх швидкого пошуку.

2.3.3 Regional CNN (R-CNN)

Щоб обійти проблему вибору величезної кількості регіонів, Росс Гіршик та ін. запропонував метод, коли ми використовуємо вибіркового пошук, щоб витягти з зображення лише 2000 регіонів, і він назвав їх регіональними пропозиціями. Тому зараз, замість того, щоб намагатися класифікувати величезну кількість регіонів, можна просто працювати з 2000 регіонами. Ці 2000 регіональних пропозицій формуються за допомогою алгоритму вибіркового пошуку, який написано нижче.

Вибірковий пошук:

1. Створюємо початкову підсегментацію, ми генеруємо багато регіонів-кандидатів
2. Використовуйте жадібний алгоритм для рекурсивного поєднання подібних регіонів у більші
3. Використовуйте створені регіони для створення остаточних пропозицій регіону-кандидата

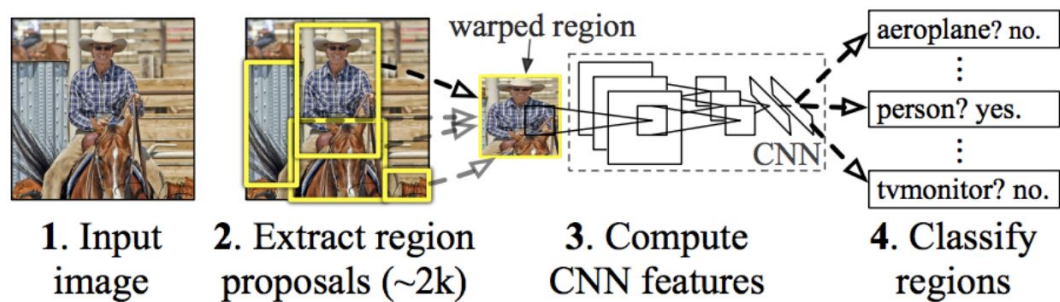


Рисунок 2.4 – Кроки обробки алгоритмом R-CNN

Ці 2000 регіонів-кандидатів пропонуються в квадрат і подаються в згорнуту нейронну мережу, яка виробляє 4096-мірний векторний вектор як вихід. CNN виступає в ролі екстрактора функцій, і вихідний щільний шар складається з особливостей, витягнутих із зображення, а витягнуті функції подаються у SVM для класифікації присутності об'єкта в рамках програми кандидата. На додаток до прогнозування наявності об'єкта в регіонах пропозицій, алгоритм також передбачає чотири значення, які є зміщеними значеннями для підвищення точності обмежувального поля. Наприклад, враховуючи пропозицію щодо регіону, алгоритм передбачав би присутність людини, але обличчя цієї людини в межах цієї пропозиції могло бути розрізане навпіл. Отже, значення зміщення допомагають у налаштуванні обмежувальної рамки пропозиції регіону.

До недоліків R-CNN можна віднести:

- Навчання мережі потребує значної кількості часу, оскільки вам доведеться класифікувати 2000 пропозицій регіону на зображення.
- Він не може бути реалізований в режимі реального часу, оскільки це займає близько 47 секунд для кожного тестового зображення.
- Селективний алгоритм пошуку – це фіксований алгоритм. Тому на цьому етапі ніякого навчання не відбувається. Це може призвести до породження поганих пропозицій регіону-кандидата.

2.3.4 Швидка R-CNN

Той же автор попередньої роботи (R-CNN) вирішив деякі недоліки R-CNN для побудови більш швидкого алгоритму виявлення об'єктів, і він отримав назву Fast R-CNN. Підхід аналогічний алгоритму R-CNN. Але, замість подачі пропозицій регіону до CNN, ми подаємо вхідне зображення до CNN, щоб створити згорнуту карту характеристик. Зі згорнутою картою ознак ми визначаємо область пропозицій і перев'язуємо їх у квадрати, а за допомогою шару об'єднання RoI перетворюємо їх у фіксований розмір, щоб він міг подаватись у повністю пов'язаний шар. Із вектора функцій RoI ми використовуємо шар softmax для прогнозування класу запропонованої області, а також значення зміщення для обмежувального поля.

Причина чому "Швидка R-CNN" швидша, ніж R-CNN, це тому, що вам не потрібно щоразу подавати 2000 пропозицій регіону в конволюційну нейронну мережу. Натомість операція згортання виконується лише один раз на зображення і з неї формується карта функції.

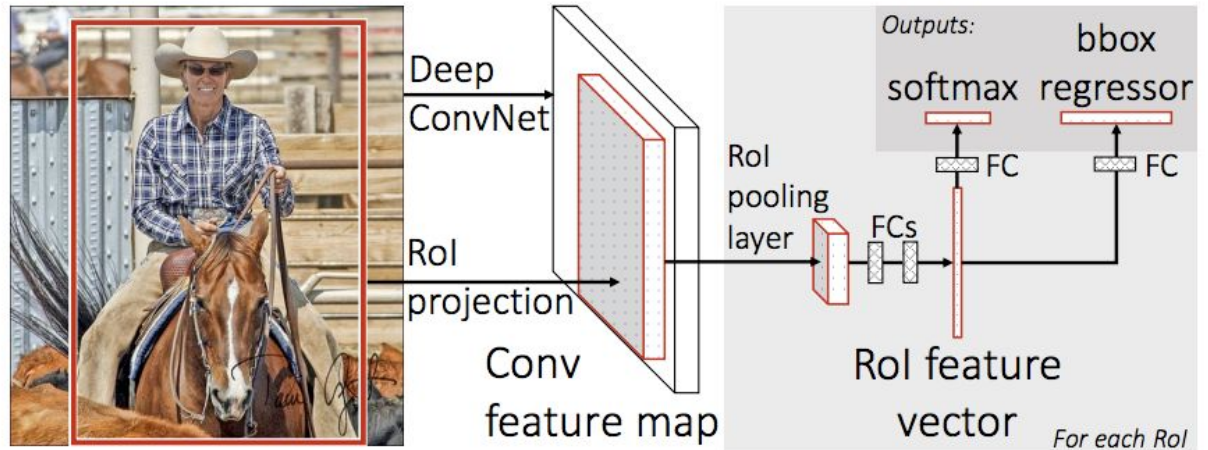


Рисунок 2.5 – Схема обробки зображення R-CNN

2.3.5 Швидша R-CNN

Обидва вищевказані алгоритми (R-CNN та швидкий R-CNN) використовують вибірковий пошук для з'ясування пропозицій регіону. Вибірковий пошук – це повільний і трудомісткий процес, що впливає на продуктивність мережі. Тому Shaoqing Ren та ін. придумав алгоритм виявлення об'єктів, який виключає алгоритм вибіркового пошуку та дозволяє мережі вивчати пропозиції регіону.

Подібно до Швидкого R-CNN, зображення надається як вхід до згорткової мережі, яка забезпечує згорткову мапу ознак. Замість використання алгоритму вибіркового пошуку на карті функцій для ідентифікації пропозицій регіону використовується окрема мережа для прогнозування пропозицій регіону. Прогнозовані пропозиції регіону потім переробляються за допомогою шару об'єднання RoI, який потім використовується для класифікації зображення в запропонованій області та прогнозування значень зміщення для обмежувальних полів.

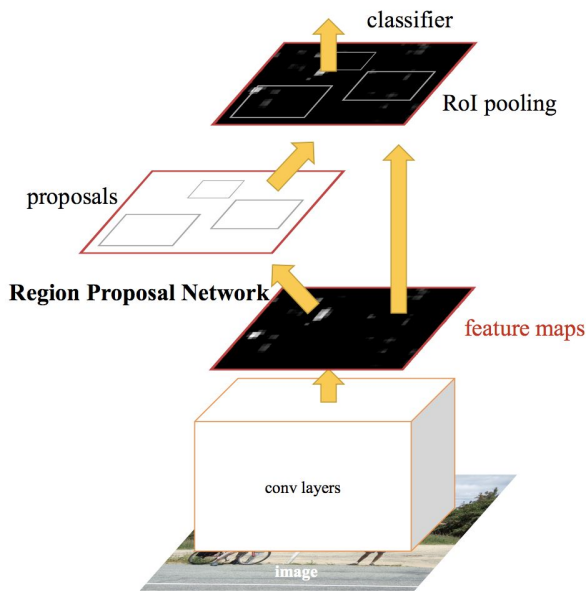


Рисунок 2.6 – Швидша R-CNN

Згідно тестів[15], Швидшу R-CNN можна використовувати для обробки в режимі реального часу.

2.3.6 You Only Look Once (YOLO)

Усі попередні алгоритми виявлення об'єктів використовують регіони для локалізації об'єкта в зображенні. Мережа не дивиться на повне зображення. Натомість частини зображення, які мають велику ймовірність вмісту об'єкта. YOLO або You Only Look Once – алгоритм виявлення об'єктів, що сильно відрізняється від алгоритмів на основі регіону, які бачили вище. У YOLO єдина згорнута мережа прогнозує обмежувальні поля та ймовірності класу для цих полів.

Як працює YOLO, це те, що ми робимо зображення та розбиваємо його на сітку $S \times S$, у кожній із сіток ми беремо m обмежувальних коробок. Для кожного з обмежувальних вікон мережа виводить значення ймовірності класу та зсуву для обмежувального поля. Обмежувальні поля, що мають ймовірність класу вище

порогового значення, вибираються та використовуються для пошуку об'єкта в зображенні.

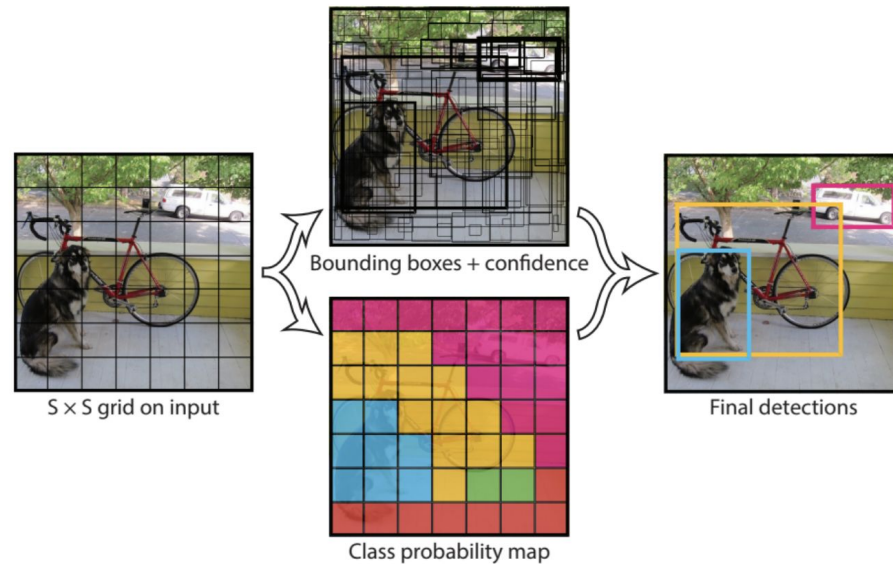


Рисунок 2.7 – Етапи виділення об'єктів алгоритмом YOLO

YOLO на порядок швидший (45 кадрів в секунду), ніж інші алгоритми виявлення об'єктів. Обмеження алгоритму YOLO полягає в тому, що він бореться з дрібними предметами всередині зображення, наприклад, у нього можуть виникнути труднощі при виявленні зграї птахів. Це пов'язано з просторовими обмеженнями алгоритму.

2.3.7 Рекурентні нейронні мережі (RNN)

Рекурентна нейронна мережа – це узагальнення подаючої нейронної мережі, яка має внутрішню пам'ять. RNN є рекурентним за своєю суттю, оскільки він виконує однакову функцію для кожного введення даних, тоді як вихід поточного вводу залежить від попереднього обчислення. Після отримання результату він копіюється та відправляється назад у мережу, що повторюється.

Для прийняття рішення він враховує поточний вхід і вихід, який він отримав з попереднього вводу.

На відміну від нейронних мереж прямого поширення, RNN можуть використовувати свій внутрішній стан (пам'ять) для обробки послідовностей входів. Це робить їх застосовними до таких завдань, як несегментоване, пов'язане розпізнавання рукописного тексту або розпізнавання мовлення. В інших нейронних мережах всі входи незалежні один від одного. Але в RNN всі входи пов'язані один з одним.

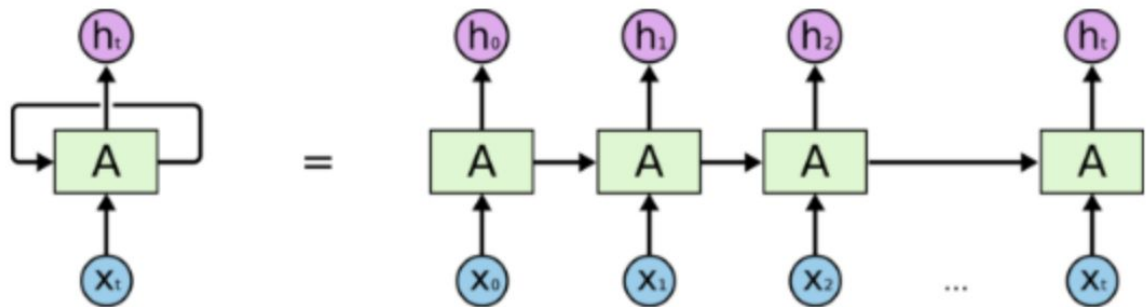


Рисунок 2.8 – розгорнута рекурентна нейронна мережа

Спочатку він бере $X(0)$ з послідовності введення, а потім виводить $h(0)$, що разом з $X(1)$ є входом для наступного кроку. Отже, $h(0)$ і $X(1)$ – вхід для наступного кроку. Аналогічно, $h(1)$ з наступного – це вхід з $X(2)$ для наступного кроку тощо. Таким чином, він постійно пам'ятає контекст під час тренувань.

Формула поточного стану:

$$h_t = f(h_{t-1}, x_t).$$

Застосовуючи функцію активації:

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t),$$

де W – вага, h – одиничний прихований вектор, W_{hh} – вага у попередньому прихованому стані, W_{hx} – вага у поточному входному стані, \tanh – функція активації, яка реалізує нелінійність, що притискає активації до діапазону $[-1;1]$

Переваги рекурентних нейронних мереж:

- RNN може моделювати послідовність даних так, що кожний зразок можна вважати залежним від попередніх.
- Рецидивуюча нейронна мережа використовується навіть із звивистими шарами для розширення ефективного сусідства пікселів.

Недоліки рекурентної нейронної мережі:

- Проблеми зникнення градієнту.
- Навчання RNN – дуже складне завдання.
- Він не може обробляти дуже довгі послідовності, якщо використовувати \tanh або relu в якості функції активації.

2.3.8 Мережі із довгою короткочасною пам'яттю (LSTMs)

У середині 90-х рр. Німецькі дослідники Сепп Хохрейтер та Юрген Шмідхубер запропонували зміну періодичної мережі з так званими одиницями довготривалої пам'яті як рішення вирішальної проблеми градієнта.

LSTM допомагають зберегти помилку, яка може бути розповсюджена через час та шари. Підтримуючи більш постійну помилку, вони дозволяють повторюваним мережам продовжувати вчитися протягом багатьох етапів часу (понад 1000), тим самим відкриваючи канал для віддаленого зв'язку причин та наслідків. Це одна з головних проблем машинного навчання та ШІ, оскільки алгоритми часто стикаються в середовищах, коли сигнали про винагороду рідкісні та затримані, наприклад, саме життя.

LSTM містять інформацію за межами нормального потоку періодичної мережі в осередку, що закривається. Інформація може зберігатися, записуватися або читатися з комірки, подібно до даних у пам'яті комп'ютера. Камера приймає рішення про те, що зберігати та коли дозволяти читати, писати та стирати, через ворота, які відкриваються та закриваються. Однак, на відміну від цифрового накопичувача на комп'ютерах, ці ворота є аналоговими, реалізованими з елементарним множенням на сигмоїди, які знаходяться в межах 0-1. Аналог має перевагу перед цифровим у тому, що він є диференційованим, а тому придатний для зворотного розповсюдження.

Ці ворота діють на сигнали, які вони отримують, і подібні до вузлів нейронної мережі, вони блокують або передають інформацію на основі її сили та імпорту, яку вони фільтрують за допомогою власних наборів ваг. Ці ваги, як ваги, які модулюють вхідні та приховані стани, регулюються за допомогою процесу навчання мереж, що повторюються. Тобто, комірки дізнаються, коли дозволити введення даних, залишення або видалення за допомогою ітераційного процесу вгадування, помилки зворотного розповсюдження та регулювання ваги за допомогою градієнтного спуску.

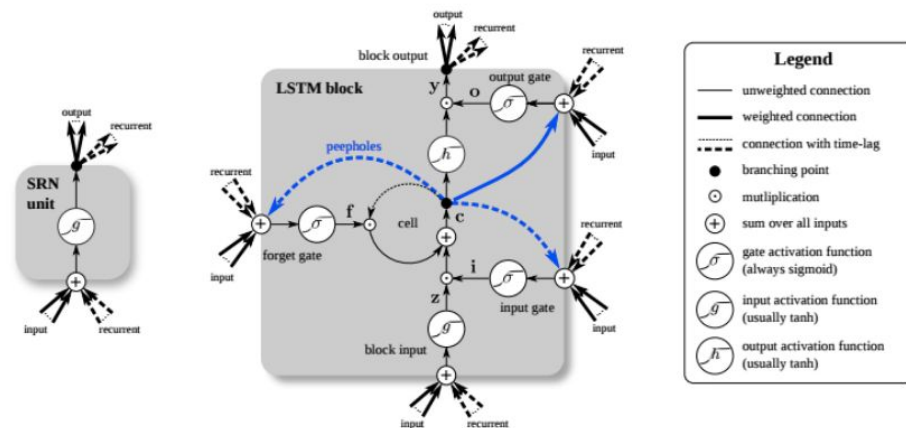


Рисунок 2.9 – Схеми простої рекурентної мережі та LSTM.

Ці вентиля можуть дізнатися, які дані в послідовності важливо зберегти або викинути. Роблячи це, він може передавати відповідну інформацію по довгому ланцюгу послідовностей, щоб робити прогнози. Практично всі сучасні результати, засновані на періодичних нейронних мережах, досягаються цими двома мережами. LSTM та GRU можна знайти в розпізнаванні мовлення, синтезі мовлення та генеруванні тексту. Ви навіть можете використовувати їх для створення підписів для відео.

2.4. Аналітика поведінки

Однією з найскладніших проблем у сфері комп'ютерного зору та штучного інтелекту є семантичне навчання поведінці та розуміння з боку спостереження за діяльністю відео (візуального) спостереження. Дослідження в цій галузі зосереджуються головним чином на розробці методів аналізу візуальних даних з метою вилучення та обробки інформації про поведінку фізичних об'єктів (наприклад, людей) на сцені.

В автоматизованих системах візуального спостереження важливе практичне значення має надійне виявлення підозрілих чи загрожуючих поведінці людей [12]. Автоматизована система візуального спостереження, як правило, вимагає надійного поєднання зображень та методів штучного інтелекту. Методи обробки зображень використовуються для забезпечення функцій зображення низького рівня. Методи штучного інтелекту використовуються для надання експертних рішень. Було проведено широкі дослідження щодо таких методів обробки зображень низького рівня, як виявлення, розпізнавання об'єктів та відстеження об'єктів; однак порівняно небагато досліджень повідомляло про надійну класифікацію та розуміння людської діяльності із послідовностей відеозображень [14].

Виявлення підозрілої поведінки людини включає моделювання та класифікацію людської діяльності за певними правилами [5]. Моделювання та класифікація людської діяльності не тривіальна через випадковість і складний характер руху людини. Ідея полягає у розподілі спостережуваних рухів людини на деякі дискретні стани, а потім їх відповідній класифікації. Мабуть, розподіл спостережуваних рухів дуже специфічний для застосування та загалом важко передбачити, що буде представляти собою підозрілу чи загрожуючу поведінку [5, 12].

Більшість підходів у галузі відео розуміння – об'єднані методи виявлення подій, що стосуються домену [4]. Приклади таких систем використовують динамічну деформацію часу для розпізнавання жестів [18] або самоорганізуючі мережі для класифікації траєкторій. Основним недоліком цих підходів є використання методів, специфічних лише для певної сфери застосування, що спричиняє труднощі при застосуванні цих методів до інших областей. Тому деякі дослідники [4] запропонували та прийняли двоступеневий підхід до проблеми розуміння відео:

- Візуальний модуль обробки зображень нижчого рівня використовується для вилучення візуальних сигналів та примітивних подій.
- Ця зібрана інформація використовується в модулі штучного інтелекту вищого рівня для виявлення більш складних і абстрактних моделей поведінки

Розділивши проблему на дві або три [2] підпроблеми, дослідники можуть використовувати більш прості та більш доменні незалежні методи на кожному етапі. Перший етап зазвичай включає і використовує методи обробки зображень та стохастичні методи аналізу даних, а другий етап проводить структурний аналіз символічних даних, зібраних на попередньому кроці.

У загальних рамках візуального спостереження, виявлення / сегментація руху та класифікація об'єктів зазвичай групуються як завдання зору нижнього рівня. Розпізнавання поведінки людини базується на успішному відстеженні людського суб'єкта через послідовності зображень і вважається завданням зору високого рівня. Процес відстеження, як обговорюється в роботі, можна вважати завданням зору середнього рівня, або він може бути розділений на нижчі та вищі два етапи, як було запропоновано в роботі [4] та показано на рис. 2.3.

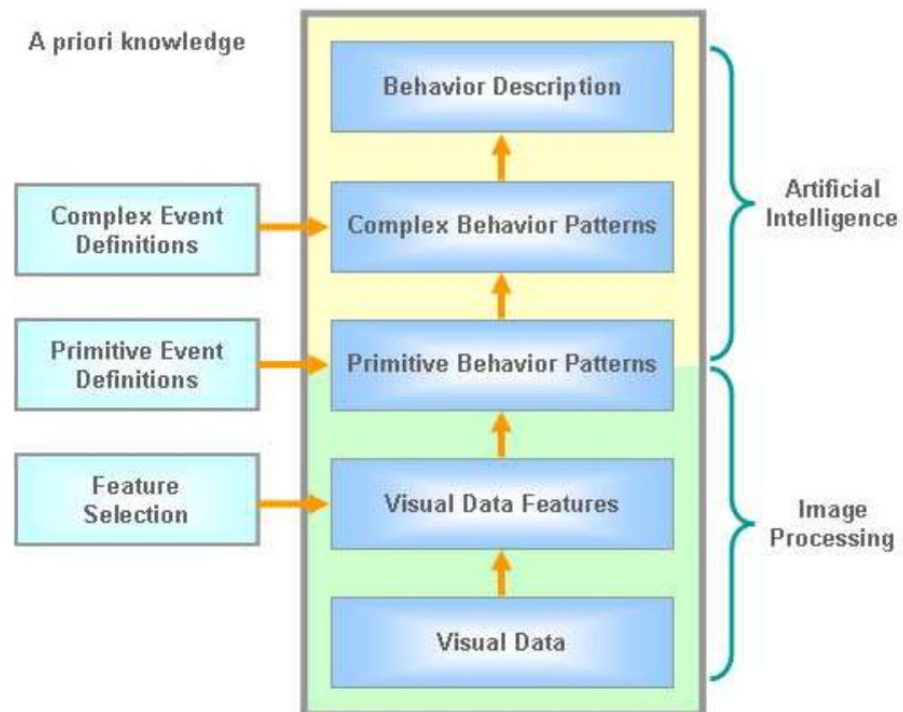


Рисунок 2.3 – Загальна архітектура системи “розуміння” відео

Як показано на рис. 2.3, на першому рівні загальної системи відеоспостереження витягуються геометричні особливості, подібно областям руху. На основі цих вилучень об'єкти розпізнаються та відстежуються. На другому рівні розпізнаються події, в яких беруть участь виявлені об'єкти. Для виконання цього завдання використовується обране подання подій, яке визначає поняття та відносини у сфері моніторингу діяльності людини.

Для спільноти комп'ютерного зору природний підхід до розпізнавання сценаріїв полягає у використанні ймовірнісної чи нейронної мережі. Вузли цієї мережі зазвичай відповідають сценаріям, які розпізнаються в даному екземплярі з обчисленою ймовірністю.

Для спільноти штучного інтелекту природним способом розпізнавання сценарію є використання символічної мережі, де вузли зазвичай відповідають булевим розпізнаванням сценаріїв. Загальною характеристикою цих підходів є те, що всі повністю визнані форми поведінки зберігаються.

Іншою розробкою, яка привернула увагу дослідників, є непідконтрольне навчання та розпізнаванню поведінки, що полягає у можливості системи інтерпретації бачення навчання та виявлення частих сценаріїв сцени, не вимагаючи попередніх визначень поведінки користувачем.

Будь-який об'єкт сцени, що бере участь у поведінці / дії, також повинен включати інших людей, групи людей, натовп чи статичні об'єкти (наприклад, обладнання). Діяльність передбачає регулярно повторювану послідовність рухових подій. Автоматичне розуміння та тлумачення відео потрібно знати, як це зробити

представляють та розпізнають поведінку, що відповідає різним типам понять, до яких належать [4, 7]:

- Основні властивості – це характеристики об'єкта, така як його траєкторія чи швидкість.
- Стани описують ситуацію, що характеризує один або кілька об'єктів (дійових осіб), визначених у даний момент часу (наприклад, суб'єкт збуджується) або стабільну ситуацію, визначену за часовий інтервал. Що стосується держави: "фізична особа залишається поруч з торговим автоматом", залучені два суб'єкти (актори): людина та обладнання.

- Події – це зміна стану у два рази поспіль (наприклад, суб'єкт потрапляє у зону інтересу).
- Сценарії – це комбінація станів, подій або під-сценаріїв. Поведінки – це конкретні сценарії, залежні від програми, визначеної користувачами. Наприклад, для спостереження за станціями метро кінцеві користувачі могли визначити цілеспрямоване поведінку: "засмічення", "покинутий багаж", "вандалізм", "переповненість", "боротьба" тощо

Здатність добувати смислову інформацію з біологічного руху людини відома вже давно. У своїй наглядній роботі Йохансон [11] показав, що подання скоординованого руху людських суглобів є достатнім для створення враження про людину, що ходить чи бігає через космос.

Що стосується виявлення ворожого наміру [5], у кожній точці ходу точкового світла (PLW) може бути свій власний рух жесту, який при дослідженні по відношенню до ланок об'єкта може бути використаний для визначення загального стану система. Незвичайні події, такі як вандалізм або переповненість районів, можуть бути виявлені незвичайними рухами, а також малоймовірними положеннями об'єктів.

Люди мали вроджену здатність розпізнавати чужі емоційні настрої на основі інтуїції; ця вродженість повинна проявлятися і фізично. Наприклад, коли хтось переживає емоції, які візуальні підказки, які передають це? Вираз обличчя – це безпосередній показник, а як щодо їх поведінки? Чи позначають це також постава, жести або конкретні частини тіла? Система зможе вивчити візуальні підказки, які мають певне значення для виявлення емоції [11] шляхом визначення конкретних областей тіла, що ідентифікують емоції. Дослідники виявлять, що рухи певних частин тіла можуть ідентифікувати емоцію більше,

ніж інші [5, 11]. Наприклад, дослідники можуть виявити, що в гніві тулуб найбільше викликає цю емоцію.

Огляд наявних та сучасних методик показує велику різноманітність прийомів розуміння відео в автоматичному розпізнаванні поведінки. Завдання полягає в ефективному поєднанні цих методів для вирішення великої різноманітності реального світу. Навчання та розуміння моделі поведінки можна розглядати як класифікацію даних про особливості, що змінюються за часом, тобто узгодження невідомої послідовності тесту з групою мічених опорних послідовностей, що представляють типову або вивчену поведінку [2]. Основна проблема розуміння поведінки полягає в тому, щоб вивчити послідовності еталонної поведінки з навчальних зразків, а також розробити як методи навчання, так і відповідність для ефективного впорядкування з невеликими варіаціями даних про особливості в кожному класі руху руху. Основні існуючі методи розуміння поведінки включають наступні:

а) Моделі прихованих марків (НММ): НММ – це статистичний інструмент, що використовується для моделювання генеративних послідовностей, що характеризується набором спостережуваних послідовностей [18].

б) Динамічне викривлення в часі (DTM): DTW – це техніка, яка обчислює нелінійну функцію деформації, яка оптимально вирівнює дві часові послідовності змінної довжини [18]. Функцію деформації можна використовувати для обчислення подібності між двома часовими рядами або для пошуку відповідних областей між двома часовими рядами.

в) Машина кінцевих станів (FSM): FSM або кінцевий стан автомата або просто стан машина, це модель поведінки, що складається з кінцевої кількості станів, переходів між цими станами та дій. Машина з кінцевим станом – це абстрактна модель машини з примітивною внутрішньою пам'яттю.

г) Недетерміновано-кінцевий стан автоматики (NFA): NFA або недетермінований кінцевий стан машини – це машина кінцевого стану, де для кожної пари символів стану та вводу може бути кілька можливих наступних станів. Це відрізняє його від детермінованого кінцевого автомата (DFA), де наступний можливий стан однозначно визначається. Хоча DFA та NFA мають різні визначення, у формальній теорії може бути показано, що вони еквівалентні, оскільки для будь-якого даного NFA можна побудувати еквівалентний DFA, і навпаки.

д) Нейронна мережа із затримкою часу (TDNN): TDNN – це підхід до аналізу даних, що змінюються часом. У TDNN одиниці затримки додаються до загальної статичної мережі, а деякі попередні значення в послідовності, що змінюється за часом, використовуються для прогнозування наступного значення. По мірі того, як стають доступні більші набори даних, більше уваги приділяється нейронним мережам для представлення тимчасової інформації. Методи TDNN успішно застосовуються у таких програмах, як розпізнавання жестів рукою та читання губ.

е) Синтаксичні/граматичні методи: Основна ідея цього підходу – розділити проблему розпізнавання на два рівні. Нижчий рівень виконується за допомогою стандартних незалежних імовірнісних тимчасових детекторів поведінки, таких як НММ, для виведення можливих часових характеристик низького рівня. Ці виходи забезпечують вхідний потік для стохастичного без контексту граматичного аналізатора. Граматика та аналізатор забезпечують тимчасові обмеження більшого діапазону, розмежовують невизначені виявлення низького рівня та дозволяють включати апіорні знання про структуру тимчасової поведінки [22].

з) Самоорганізуюча нейромережа: методи, обговорені в пунктах (а) – (е), включають навчання під контролем. Вони застосовні для відомих сцен, де типи рухів об'єкта вже відомі. Самоорганізуючі нейронні мережі підходять для розуміння поведінки, коли рухи об'єкта необмежені.

и) Методи, засновані на агентах: Замість вивчення великої кількості моделей поведінки за допомогою централізованого підходу методи, засновані на агентах, розкладають навчання на взаємодії агентів із значно простішими способами поведінки та правил.

і) Штучні імунні системи: Деякі дослідники використали можливість дотримання моделей поведінки та ворожих намірів на рівні оптичного потоку, використовуючи підходи штучної імунної системи.

2.5 Обґрунтування вибору програмних засобів

2.5.1 Python

AI-проекти відрізняються від традиційних програмних проектів. Відмінності полягають у технологічному стеку, навичках, необхідних для проекту на базі AI, та необхідності глибоких досліджень. Для реалізації ваших прагнень AI слід використовувати стабільну, гнучку та доступну інструментальну мову програмування. Python пропонує все це, саме тому ми сьогодні бачимо багато проектів AI які його використовують.

Від розробки до розгортання та обслуговування, Python допомагає розробникам бути продуктивними та впевненими у створенні програмного забезпечення. Переваги, завдяки яким Python найкраще підходить для машинного навчання та проектів на основі AI, включають простоту та послідовність, доступ до великих бібліотек та рамок для AI та машинного

навчання (ML), гнучкість, незалежність платформи та широке співтовариство. Вони сприяють загальній популярності мови.

Простий і послідовний

Python пропонує стислий і читабельний код. Хоча складні алгоритми та універсальні робочі процеси стоять за машинним навчанням та AI, простота Python дозволяє розробникам писати надійні системи. Розробники змушені докладати усіх зусиль для вирішення проблеми ML, а не зосереджуватися на технічних нюансах мови.

Широкий вибір бібліотек та рамок

Впровадження алгоритмів AI та ML може бути складним і вимагає багато часу. Важливо створити добре структуровану та перевірену обстановку, щоб розробники могли розробити найкращі рішення щодо кодування.

Щоб скоротити час розробки, програмісти звертаються до ряду фреймворків і бібліотек Python. Бібліотека програмного забезпечення – це попередньо написаний код, який розробники використовують для вирішення загальних завдань програмування. Python, має багатий набір технологій, має широкий набір бібліотек для штучного інтелекту та машинного навчання. Ось деякі з них:

- Keras, TensorFlow та Scikit-learning для машинного навчання
- NumPy для високоефективних наукових обчислень та аналізу даних
- SciPy для комплексних обчислень
- Pandas для аналізу даних загального призначення
- Seaborn для візуалізації даних

Scikit-learn має різні алгоритми класифікації, регресії та кластеризації, включаючи машини підтримки векторів, випадкові ліси, підвищення градієнта,

k-засоби та DBSCAN, і призначений для роботи з числовими та науковими бібліотеками Python NumPy та SciPy.

За допомогою цих рішень ви можете швидше розвивати свій продукт. Вашій команді розробників не доведеться винаходити колесо і може використовувати наявну бібліотеку для впровадження необхідних функцій.

Таблиця 2.1 – Типові застосування AI і найбільш популярні технології

Аналіз даних та візуалізація	NumPy, SciPy, Pandas, Seaborn
Машинне навчання	TensorFlow, Keras, Scikit-learn
Комп'ютерний зір	OpenCV
Обробка природної мови	NLTK, spaCy

Платформонезалежність

Незалежність платформи відноситься до мови програмування або рамки, що дозволяє розробникам реалізовувати речі на одній машині та використовувати їх на іншій машині без будь-яких (або лише з мінімальними) змінами. Одним із ключових чинників популярності Python є те, що мова є платформою незалежною. Python підтримується багатьма платформами, включаючи Linux, Windows та macOS. Код Python може використовуватися для створення автономних виконуваних програм для найбільш поширених операційних систем, а це означає, що програмне забезпечення Python можна легко поширювати та використовувати в цих операційних системах без інтерпретатора Python.

Більше того, розробники зазвичай використовують такі сервіси, як Google або Amazon для своїх обчислювальних потреб. Однак ви часто можете знайти компанії та науковців, які використовують власні машини з потужними графічними процесорами (GPU) для навчання своїх моделей ML. А той факт, що Python не залежить від платформи, робить це навчання набагато дешевшим та простішим.

2.5.2 Tensorflow

TensorFlow – популярна основа машинного навчання та глибокого навчання. Це безкоштовна бібліотека з відкритим кодом, яка виходить 9 листопада 2015 року та розроблена командою Google Brain Team. Він повністю заснований на мові програмування Python і використовується для чисельних обчислень та потоку даних, що робить машинне навчання швидшим та простішим.

TensorFlow може тренувати та запускати глибокі нейронні мережі для розпізнавання зображень, рукописної класифікації цифр, періодичної нейромережі, вбудовування слів, обробки природних мов, виявлення відео та багато іншого. TensorFlow працює на декількох процесорах або графічних процесорах, а також на мобільних операційних системах.

Слово TensorFlow складається з двох слів, тобто Tensor і Flow:

1. Тензор – багатовимірний масив.
2. Потік використовується для визначення потоку даних в роботі.

TensorFlow використовується для визначення потоку даних, що функціонують на багатовимірному масиві або тензорі

Пару років тому глибоке навчання почало перевершувати всі інші алгоритми машинного навчання при наданні величезної кількості даних. Google

побачив, що може використовувати ці глибокі нейронні мережі для покращення своїх послуг:

- Gmail.
- Фото.
- Пошукова система Google.

Вони будують основу під назвою Tensorflow, щоб дозволити дослідникам та розробникам спільно працювати над моделлю AI. Розроблений та розширений, він дозволяє багатьом людям користуватися ним.

Вперше він був оприлюднений наприкінці 2015 року, тоді як перша стабільна версія з'явилася у 2017 році. Він є відкритим кодом за ліцензією Apache Open Source. Ви можете користуватися ним, змінювати його та розповсюджувати модифіковану версію за окрему плату, не платячи нічого Google.

Архітектура Tensorflow складається з трьох частин:

- Попередня обробка даних.
- Побудова модель.
- Тренування та оцінка моделі.

Його називають Tensorflow, оскільки він приймає вхід як багатовимірний масив, також відомий як тензори. Ви можете побудувати своєрідну блок-схему операцій (звану графік), яку ви хочете виконати на цьому вході. Вхід надходить з одного кінця, а потім він протікає через цю систему декількох операцій і виходить з іншого кінця як вихід.

Ім'я Tensorflow прямо походить від основної основи: Tensor. У Tensorflow всі обчислення включають тензори. Тензор – це вектор або матриця n-розмірів, що представляє всі типи даних. Усі значення в тензорі містять однаковий тип

даних з відомою (або частково відомою) формою. Форма даних – це розмірність матриці або масиву.

Тензор може виникнути з вхідних даних або результату обчислення. У TensorFlow всі операції проводяться всередині графіка. Графік – це набір обчислень, який відбувається послідовно. Кожна операція називається оп-вузлом і з'єднана між собою.

На графіку окреслюються операції та з'єднання між вузлами. Однак значення не відображаються. Край вузлів – це тензор, тобто спосіб заповнити операцію даними.

Ось чому його називають TensorFlow, оскільки тензор йде в ньому, протікає через список операцій, а потім виходить інша сторона.

TensorFlow використовує фреймворк графік. Граф збирає та описує всі розрахунки серій, зроблені під час тренінгу. Граф має безліч переваг:

- Це було зроблено для роботи на декількох процесорах або графічних процесорах і навіть мобільній операційній системі.
- Переносимість графіка дозволяє зберегти обчислення для негайного чи пізнішого використання. Графік можна зберегти для виконання у майбутньому.
- Усі обчислення на графіку проводяться шляхом з'єднання тензорів разом.
- У тензора є вузол і край. Вузол здійснює математичну операцію і виробляє виходи кінцевих точок. По краях ребра пояснюють співвідношення вводу / виводу між вузлами.

TensorFlow – найкраща бібліотека з усіх, оскільки вона створена для того, щоб бути доступною для всіх. Бібліотека Tensorflow містить різні API, побудовані на масштабній архітектурі глибокого навчання, такі як CNN або

RNN. TensorFlow заснований на обчисленні графіка; це дозволяє розробнику візуалізувати побудову нейронної мережі за допомогою Tensorboard. Цей інструмент корисний для налагодження програми. Нарешті, Tensorflow побудований для розміщення в масштабі. Він працює на процесорі та GPU.

Tensorflow привертає найбільшу популярність на GitHub порівняно з іншими рамками глибокого навчання.

Висновки по розділу

В даному розділі подано основні етапи алгоритму розуміння відео, розглянуто і проаналізовано способи та технології реалізації задач кожного з етапів, наведено опис і характеристики методів, їх основні переваги та недоліки.

РОЗДІЛ 3. РОЗРОБКА АЛГОРИТМІЧНОГО ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

3.1. Двопоточкова архітектура для розпізнавання відео

Відео природно може бути розкладене на просторові та часові компоненти. Просторова частина, у вигляді зовнішнього вигляду кадру, містить інформацію про сцени та предмети, зображені на відео. Часова частина, у формі руху по кадрах, передає рух спостерігача (камери) та предметів. Відповідно ми розробили нашу архітектуру розпізнавання відео, розділивши її на два потоки, як показано на рис. 3.1. Кожен потік реалізований за допомогою глибокого ConvNet, бали програмного забезпечення якого поєднуються пізнім синтезом. Ми розглядаємо два методи злиття: усереднення та навчання багатокласного лінійного SVM на складених L2-нормалізованих балів softmax як функції.

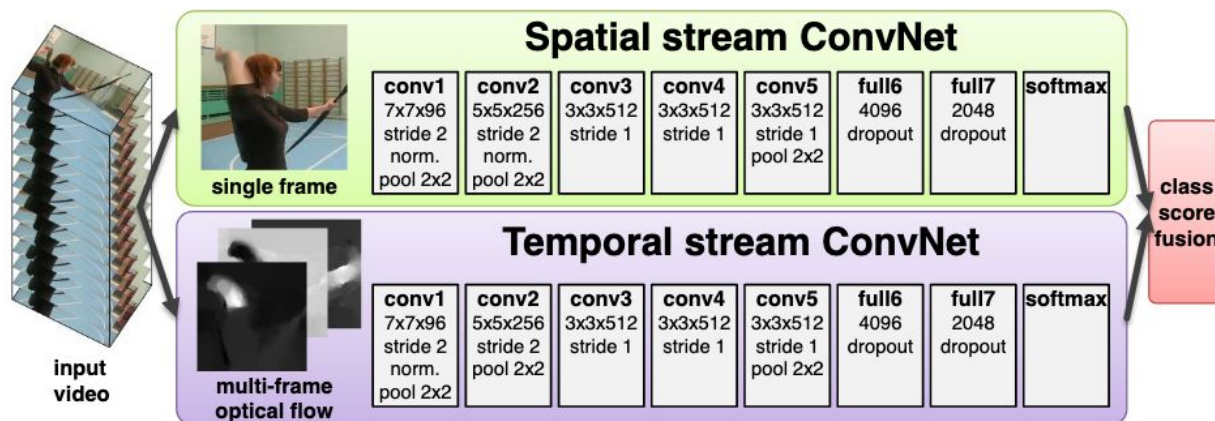


Рисунок 3.1 – Двопоточкова архітектура для розпізнавання відео

Просторовий потік ConvNet працює на окремих відеокадрах, ефективно виконуючи розпізнавання дій із нерухомих зображень. Статичний вигляд сам по собі є корисною підказкою, оскільки деякі дії сильно пов'язані з певними об'єктами. Насправді, як буде показано в секті. 6, класифікація дій з нерухомих

кадрів (потік просторового розпізнавання) є досить конкурентоспроможною. Оскільки просторовий ConvNet по суті є архітектурою класифікації зображень, ми можемо спиратися на останні досягнення в масштабних методах розпізнавання зображень [15] та попередньо тренувати мережу на великому наборі даних класифікації зображень, наприклад наборі даних викликів ImageNet. Деталі представлені в розділі 3.4. Далі ми описуємо тимчасовий потік ConvNet, який експлуатує рух і значно підвищує точність.

3.2. Застосування оптичного потоку для CNN

У цьому розділі ми описуємо модель ConvNet, яка формує потік тимчасового розпізнавання нашої архітектури (розділ 3.1). На відміну від моделей ConvNet, розглянутих у розділі 2.3, вхід до нашої моделі формується шляхом укладання полів оптичного переміщення потоку між декількома послідовними кадрами. Такий вхід явно описує рух між відеокадрами, що полегшує розпізнавання, оскільки мережі не потрібно неявно оцінювати рух. Ми розглядаємо кілька варіантів входу на основі оптичного потоку, які ми описуємо нижче.

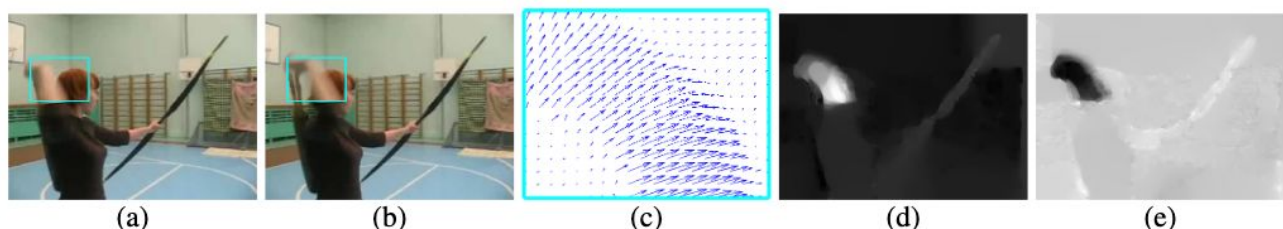


Рисунок 3.2 – Представлення кадру відеовходу оптичним потоком

На рис. 3.2 – (a), (b): пара послідовних відеокадрів із площею навколо рухомої руки, окресленої блакитним прямокутником. (c): крупним планом густого оптичного потоку в окресленій області; (d): горизонтальний компонент дорівнює вектору переміщення (більша інтенсивність відповідає позитивним

значенням, менша інтенсивність до негативних значень). (д): вертикальний компонент. Зверніть увагу, як (d) та (e) виділяють рухому руку та лук. Вхід до ConvNet містить кілька потоків.

3.2.1. Налаштування вхідного потоку для CNN

Накладання оптичних потоків. Густих оптичний потік можна розглядати як набір полів вектора переміщення між парами послідовних кадрів t і $t + 1$. За допомогою $dt(u, v)$ позначаємо вектором переміщення точку (u, v) у кадрі, яка переміщує точку до відповідної точки у наступному кадрі $t + 1$. Горизонтальну та вертикальну компоненти векторного поля, dx_t та dy_t можна розглядати як канали зображення (показані на рис. 3.2), добре підходять для розпізнавання за допомогою згорткової мережі. Щоб представити рух по послідовності кадрів, ми укладаємо канали потоку dx, y т L послідовних кадрів, щоб сформувати загальну кількість вхідних каналів $2L$. Більш офіційно, нехай w і h – ширина і висота відео; а вхідний об'єм ConvNet $I_\tau \in \mathbb{R}^{w \times h \times 2L}$ для довільного кадру τ будується наступним чином:

$$\begin{aligned} I_\tau(u, v, 2k - 1) &= d_{\tau+k-1}^x(u, v), \\ I_\tau(u, v, 2k) &= d_{\tau+k-1}^y(u, v), \quad u = [1; w], v = [1; h], k = [1; L]. \end{aligned} \quad (3.1)$$

Для довільної точки (u, v) канали $I_\tau(u, v, c)$, $c = [1; 2L]$ кодують рух у цій точці через послідовність L кадрів (як показано на рис. 3.2 зліва).

Складання траєкторій. Альтернативне представлення руху, натхнене дескрипторами, заснованими на траєкторії, замінює оптичний потік, відібраний в одних і тих же місцях через кілька кадрів, потік, відібраний по траєкторіях руху. У цьому випадку вхідний об'єм I_τ , що відповідає кадру τ , приймає таку форму:

$$\begin{aligned}
 I_\tau(u, v, 2k - 1) &= d_{\tau+k-1}^x(\mathbf{p}_k), \\
 I_\tau(u, v, 2k) &= d_{\tau+k-1}^y(\mathbf{p}_k), \quad u = [1; w], v = [1; h], k = [1; L].
 \end{aligned}
 \tag{3.2}$$

p_k – k -та точка вздовж траєкторії, яка починається з розташування (u, v) у кадрі τ і визначається наступним відношенням рецидиву:

$$\mathbf{p}_1 = (u, v); \quad \mathbf{p}_k = \mathbf{p}_{k-1} + \mathbf{d}_{\tau+k-2}(\mathbf{p}_{k-1}), \quad k > 1.$$

Порівняно з поданням вхідного об'єму (формула 3.1), де канали $I_\tau(u, v, c)$ зберігають вектори переміщення у місцях розташування (u, v) , вхідний об'єм (формула 3.2) зберігає вектори, відібрані в місцях p_k уздовж траєкторія (як показано на рис. 3.2 – права)

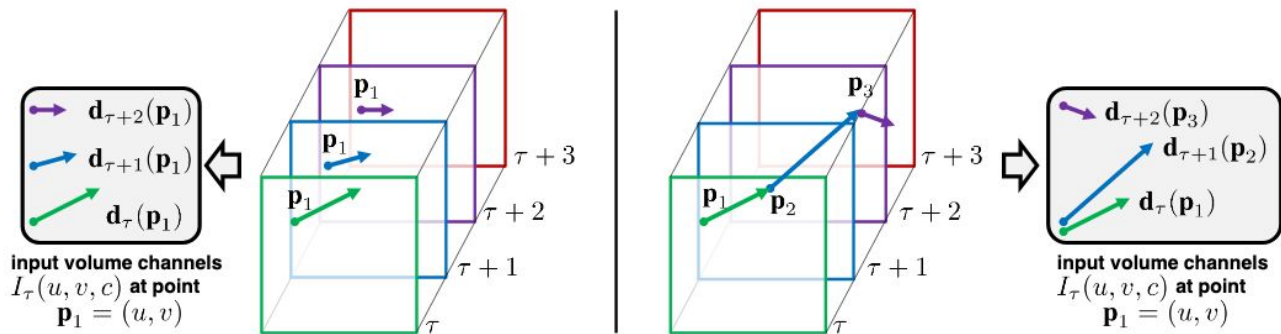


Рисунок 3.3 – Виведення входу ConvNet з багатокадрового оптичного потоку. Ліворуч: оптичне стеження потоку (1) вибірки векторів переміщення призначають одне і те ж місце розташування в декількох кадрах. Праворуч: укладка траєкторії (2) відбирає вектори вздовж траєкторії. Кадри та відповідні вектори переміщення відображаються тим самим кольором.

Двонаправлений оптичний потік. Оптичні подання потоку (1) і (2) мають справу з переднім оптичним потоком, тобто поле переміщення dt кадру t визначає розташування його пікселів у наступному кадрі $t + 1$. Це природно врахувати розширення до двонаправленого оптичного потоку, яке можна

отримати шляхом обчислення додаткового набору полів переміщення у зворотному напрямку. Потім ми побудуємо вхідний об'єм I_t , укладаючи $L / 2$ прями потоки між кадрами τ та $\tau + L / 2$ та $L / 2$ назад потоками між кадрами $\tau - L / 2$ та τ . Таким чином, вхід I_t має таку ж кількість каналів ($2L$), як і раніше. Потоки можна представити, використовуючи будь-який із двох методів (1) та (2).

Віднімання середнього потоку, як правило, вигідно виконувати нульове центрування мережевого входу, оскільки це дозволяє моделі краще використовувати нелінійності випрямлення. У нашому випадку компоненти поля вектора переміщення можуть приймати як позитивні, так і негативні значення, і, природно, зосереджені в тому сенсі, що при великій різноманітності рухів рух в одну сторону такий же ймовірний, як і рух у протилежному. Однак, даючи пару кадрів, в оптичному потоці між ними може переважати певне переміщення, наприклад, викликаний рухом камери. Важливість компенсації руху камери раніше висвітлювалася в [10, 26], де глобальна складова руху була оцінена і віднята з щільного потоку. У нашому випадку ми розглядаємо більш простий підхід: з кожного поля переміщення d віднімаємо його середній вектор.

Архітектура. Вище ми описали різні способи поєднання декількох полів зміщення оптичного потоку в єдиний об'єм $I_t \in R_w \times h \times 2L$. Враховуючи, що для ConvNet потрібен вхід фіксованого розміру, ми відбираємо підмножину $224 \times 224 \times 2L$ з I_t і передаємо її в мережу як вхід. Конфігурація прихованих шарів залишається значною мірою такою ж, як у просторовій сітці, і проілюстрована на рис. 3.1.

3.3 Багатоцільове навчання

На відміну від просторового потоку ConvNet, який можна заздалегідь підготувати на великому наборі даних про класифікацію нерухомих зображень

(наприклад, ImageNet), тимчасовий ConvNet потребує навчання за відеоданими – а наявні набори даних для класифікації відео дії все ще досить малі. У наших експериментах (розділ 6) навчання проводиться на наборах даних UCF-101 та HMDB-51, які мають лише 9,5К та 3,7К videos відповідно. Для зменшення перевиконання можна розглянути можливість поєднання двох наборів даних в один; це, однак, не є простим через перетин між множинами класів. Один варіант (який ми оцінюємо пізніше) полягає лише в тому, щоб додати зображення з класів, які не відображаються в початковому наборі даних. Це, однак, вимагає ручного пошуку таких занять і обмежує кількість додаткових даних про навчання.

Більш ґрунтовний спосіб поєднання декількох наборів даних базується на багатозадачному навчанні [16]. Її метою є вивчення (відео) подання, яке застосовується не лише до розглядуваної задачі (наприклад, класифікації HMDB-51), але й до інших завдань (наприклад, класифікація UCF-101). Додаткові завдання виконують роль регулятора та дозволяють використовувати додаткові дані про навчання. У нашому випадку архітектура ConvNet модифікована таким чином, що вона має два класифікаційних шари softmax поверх останнього повністю з'єданого шару: один рівень softmax обчислює результати класифікації HMDB-51, інший – результати оцінки THEUCF-101. Кожен із шарів оснащений власною функцією втрати, яка працює лише на відео, що надходять із відповідного набору даних. Загальна втрата тренінгу обчислюється як сума втрат окремих завдань, а похідні ваги мережі можуть бути знайдені шляхом зворотного поширення.

3.4 Деталі реалізації

Конфігурація ConvNets. Конфігурація шарів наших просторових та часових ConvNets схематично показана на рис. 3.1. Усі приховані вагові шари використовують функцію активації випрямлення (ReLU); максимальне об'єднання виконується понад 3×3 просторових вікон з кроком 2; . Єдина відмінність між просторовою та часовою конфігураціями ConvNet полягає в тому, що ми видалили другий рівень нормалізації з останнього, щоб зменшити споживання пам'яті.

Навчання. Процедура навчання може розглядатися як адаптація процедури [15] до відеокадрів, і, як правило, однакова і для просторових, і для тимчасових мереж. Мережеві ваги вивчаються за допомогою міні-серійного стохастичного градієнта спуску з імпульсом (встановлено на 0,9). На кожній ітерації побудовано міні-батч 256 зразків шляхом вибірки 256 навчальних відеороликів (рівномірно по класах), з кожного з яких вибирається один кадр випадковим чином. При просторовій сітковій підготовці підримок 224×224 випадково обрізається з обраного кадру; Потім він зазнає випадкового горизонтального перегортання та тремтіння RGB. Відео заздалегідь змінюють масштаб, щоб найменша сторона кадру дорівнювала 256. Зауважимо, що на відміну від [15], суб-зображення вибирається з усього кадру, а не лише з центру 256×256 . ми обчислюємо оптичний об'єм потоку I для обраного тренувального кадру, як описано в секті. 3. З цього об'єму вхід фіксованого розміру $224 \times 224 \times 2$ л випадковим чином обрізається та перевертається. Швидкість навчання спочатку встановлюється на 10^{-2} , а потім зменшується за встановленим графіком, який зберігається однаковим для всіх навчальних наборів. А саме, тренуючи ConvNet з нуля, частота змінювалася на 10^{-3} , після 50К ітерацій, потім на 10^{-4} після 70К

ітерацій, а навчання припиняється після ітерацій 80К. У сценарії точного налаштування швидкість змінюється на 10-3 після ітерацій 14К, а навчання припиняється після ітерацій 20К.

Тестування. Під час тестування, даючи відео, ми відбираємо фіксовану кількість кадрів (25 в наших експериментах) з рівним тимчасовим проміжком між ними. Потім з кожного з кадрів ми отримуємо 10 входів ConvNet, обрізаючи та перевернувши чотири кути та центр кадру. Оцінки класів для цілого відео отримують шляхом усереднення балів у вибіркових кадрах та посівах у них.

Підготовка на ImageNet ILSVRC-2012. При попередній підготовці просторового ConvNet ми використовуємо те саме навчання та тестування даних, як описано вище (обрізання, перегортання, тремтіння RGB). Це дає 13,5% помилки в топ-5 для набору перевірки ILSVRC-2012, що вигідно порівняно з 16,0%, повідомленими в [31] для подібної мережі. Ми вважаємо, що основною причиною вдосконалення є збиття входів ConvNet з усього зображення, а не лише з його центру.

Навчання мульти-GPU. Наша реалізація виходить із загальнодоступного пакета інструментів Caffe, але містить низку суттєвих модифікацій, включаючи паралельне навчання на декількох GPU, встановлених в одній системі. Ми використовуємо паралелізм даних і розділяємо кожну партію SGD на кілька GPU. Навчання єдиного тимчасового ConvNet займає 1 день у системі з 4 картками NVIDIA Titan, що становить швидкість в 3 рази в порівнянні з навчанням для одного GPU.

Оптичний потік обчислюється з використанням нестандартної реалізації GPU [20] від OpenCV панель інструментів. Незважаючи на швидкий час обчислення (0,06 сек на пару кадрів), все-таки буде впроваджено вузьке місце, якщо це зробити на льоту, тому ми попередньо обчислили потік перед

тренуванням. Щоб уникнути зберігання полів переміщення у вигляді поплавців, горизонтальну та вертикальну складові потоку лінійно масштабували до діапазону $[0,255]$ і стискали за допомогою JPEG (після декомпресії потік відновлюється до початкового діапазону). Це зменшило розмір потоку для набору даних UCF-101 з 1,5 ТБ до 27 Гб.

3.5 Оцінка

Набори даних та протокол оцінювання. Оцінка проводиться за показниками розпізнавання дій UCF-101 [24] та HMDB-51, які є одними з найбільших доступних наборів анотованих відеороликів. UCF-101 містить 13К відео (в середньому 180 кадрів / відео), анотоване до 101 класів дій; HMDB-51 включає 6,8К відеозаписів із 51 дії. Протокол оцінювання є однаковим для обох наборів даних: організатори забезпечують три розбиття на дані тренувань та тестів, а результативність вимірюється середньою точністю класифікації для розбиття. Кожен розділ UCF-101 містить 9,5 крес-відео; розбиття HMDB-51 містить 3,7 К навчальних відео. Почнемо з порівняння диференціальних архітектур на першому розщепленні набору даних UCF-101. Для порівняння із сучасною технікою ми пропонуємо стандартний протокол оцінювання та повідомляємо про середню точність за три розщеплення як на UCF-101, так і на HMDB-51.

Просторові ConvNets. По-перше, ми вимірюємо продуктивність просторового потоку ConvNet. Три сценарії розглядаються: (i) тренування з нуля на UCF-101, (ii) попередня підготовка до ILSVRC-2012, наступна за допомогою тонкої настройки на UCF-101, (iii) утримання попередньо підготовленої мережі фіксованою і лише тренування останній (класифікаційний) шар. Для кожного з параметрів ми експериментуємо із встановленням коефіцієнта регуляризації

випаду до 0,5 або 0,9. З результатів, представлених у таблиці 1а, видно, що підготовка ConvNet виключно на наборі даних UCF-101 призводить до надмірної підгонки (навіть при сильному відміні) та поступається перед підготовкою на великому наборі даних ILSVRC-2012. Цікаво, що точна настройка всієї мережі дає лише незначне поліпшення в порівнянні з навчанням лише останнього шару. В останній обстановці, більш високий рівень випадку нормалізує навчання та призводить до гіршої точності. У наступних експериментах ми вирішили підготувати останній шар поверх попередньо підготовленого ConvNet.

Тимчасові ConvNets. Оцінивши просторові варіанти ConvNet, ми тепер звернемося до тимчасових архітектур ConvNet та оцінимо ефект конфігурацій входу, описаних у розділі 2.6. Зокрема, ми вимірюємо ефект: використовуючи кілька ($L = \{5, 10\}$) складених оптичних потоків; траєкторія укладання; середнє віднімання зміщення; з використанням двонаправленого оптичного потоку. Архітектури навчаються на наборі даних UCF-101 з нуля, тому ми використовували агресивний коефіцієнт відсіву 0,9, щоб допомогти покращити генералізацію. Результати наведені в таблиці 3.1. По-перше, можна зробити висновок, що укладання декількох ($L > 1$) полів переміщення у вхід є дуже вигідним, оскільки забезпечує мережу інформацією про тривалий рух, яка є більш дискримінаційною, ніж потік між парою кадрів ($L = 1$ набір). Збільшення кількості вхідних потоків від 5 до 10 призводить до меншого поліпшення, тому ми тримали фіксовану L до 10 у наступних експериментах. По-друге, ми знаходимо, що середнє віднімання є корисним, оскільки зменшує ефект глобального руху між кадрами. Ми використовуємо його в наступних експериментах як за замовчуванням. Різниця між різними техніками укладання – гранична; виявляється, що оптичне укладання потоку працює краще, ніж укладання траєкторії, а використання двонаправленого оптичного потоку лише

трохи краще, ніж однонаправлений потік вперед. Нарешті, зазначимо, що тимчасові ConvNets значно перевершують просторові ConvNets (табл. 3.1), що підтверджує важливість інформації про рух для розпізнавання дій. RGB-кадрів (11 кадрів у нашому випадку). Тренуючись з нуля на UCF-101, він досяг 56,4% точності, що краще, ніж однокадрова архітектура, що навчається з нуля (52,3%), але все ще далеко від мережі, навченої з нуля на оптичному потоці. Це свідчить про те, що, хоча важлива для багатьох кадрів інформація, важливо також її представити ConvNet відповідним чином.

Таблиця 3.1 – Точність просторової мережі на UCF-101

Налаштування тренування	Швидкість dropout	
	0.5%	0.9%
З нуля	42.5%	52.3%
Претренування і finetuning	70.8%	72.8%
Претренування і останній шар	72.7%	59.9%

Таблиця 3.2 – Точність часової мережі на UCF-101

Налаштування входу	Видалення середнього	
	виключене	включене
Однофреймовий оптичний потік (L=1)	-	73.9%
Накопичення оптичного потоку (L=5)	-	80.5%
Накопичення оптичного потоку (L=10)	79.9%	81.0%
Накопичення траєкторії (L=10)	79.6%	80.2%
Накопичення оптичного потоку (L=10), двонапрявлене	-	81.2%

Висновки по розділу

В третьому розділі описано етап розробки алгоритмічного та програмного забезпечення системи. Наведено процес навчання мереж, продемонстровані результати на тестових наборах даних.

РОЗДІЛ 4. МАРКЕТИНГОВИЙ АНАЛІЗ СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

Таблиця 4.1. Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Програмний пакет на особистий комп'ютер (або сервер), який дозволяє автоматично розпізнавати визначену поведінку і об'єкти на основі даних із відеокамери.	Безпека	Користувач може без додаткових зусиль збільшити захищеність території шляхом автоматичного розпізнавання діяльності.
	Комерційна	Розпізнавання об'єктів на території, наприклад авто на парковці, повідомлення про кількість вільних місць.

Таблиця 4.2. Опис ідеї стартап-проекту

У таблиці, Н - низьке значення параметру, С - середня, В - високе

№	Техніко-економічні характеристики ідеї	Продукція конкурентів				W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	Arlo Pro 3	Wyze	Nest Hello			
1	Швидкість роботи	В	В	В	В	Довгий час навчання	Простота архітектури	Комплексне розпізнавання в реальному часу
2	Зручність використання	С	В	С	С	Потрібне вміння використання комп'ютера	Додавання зручного інтерфейсу з першим апдейтом	Автоматизована робота
3	Вимоги до системи	Н	С	С	В	Потрібне вміння використання комп'ютера	Постачається як 1 файл	Використання найновіших методів

Закінчення таблиці 4.2

4	Кросплатформеність	Присутня	Відсутня	Відсутня	Відсутня	Відсутня підтримка старих версій ОС	-	Підтримка актуальних ОС
---	--------------------	----------	----------	----------	----------	-------------------------------------	---	-------------------------

4.2 Технологічний аудит ідеї проекту

Таблиця 4.3. Технологічна здійсненність ідеї проекту

№	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Розпізнавання діяльності	Використання SVM та ConvNet	Існуючі статті по реалізації	Доступна
2	Розпізнавання об'єктів	Використання згорткової нейронної мережі	Faster R-CNN	Повністю відкрита
<i>Обрана технологія реалізації ідеї проекту: 1</i>				

Висновок: технологічна реалізація продукту –можлива, вибрана технологія No1 яка може нам допомогти розробити якісний продукт з використанням комбінації технологій, та перспектива у майбутніх ідей.

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Таблиця 4.4. Попередня характеристика потенційного ринку

№	Показники стану ринку	Характеристика
1	Кількість головних гравців, од	4
2	Загальний обсяг продаж, грн./ум.од	200
3	Динаміка ринку	Ринок розумного спостереження росте з кожним роком
4	Наявність обмежень для входу	Відсутні, відкритий ринок (і державні замовлення)

Закінчення таблиці 4.4

5	Специфічні вимоги до стандартизації та сертифікації	Відсутні
6	Середня норма рентабельності в галузі або по ринку, %	70%

Висновок: враховуючи кількість головних гравців по ринку, зростаючу динаміку ринку, невелику кількість конкурентів та середню норму рентабельності можна зробити висновок, що на даний момент, ринок для входження стартап-продукту є привабливим

Таблиця 4.5. Характеристика потенційних клієнтів стартап-проекту

	Потреба, що формує ринок	Цільова аудиторія	Відмінності у поведінці цільових груп клієнтів	Вимоги споживачів до товару
1	Потреба контролю стану прибудинкової території	Жителі міст середнього розміру	Стартап буде давати користувачам відчуття безпеки.	Зручність у використанні. Швидка робота системи. Спроможність швидко освоїти як користуватися системою. Можливість редагувати не вірні результати роботи нейронної мережі.
2	Контроль стану великих/комплексних територій	Підприємці/охоронні компанії	Збільшення ефективності відеоспостереження при зменшенні витрат на персонал	

Таблиця 4.6. Фактори загроз

	Фактор	Зміст загрози	Можлива реакція компанії
1	Конкуренти	Наявність конкурентів котрі надають схожі рішення	Зменшення ціни на поставлену послугу; Розробка унікальних характеристик товару; Надання ліцензій на обслуговування
2	Кошти на розробку та підтримку продукту	Закінчення грошей та недостатнє фінансування	Залучення додаткових інвесторів, мотивація роботи на перспективу; Ітеративна розробка продукту задля покрокового виведення продукту на ринок та отримання відповіді користувачів
3	Вихід аналогу	Вихід аналогу даного товару може призвести до знецінення та безідейності даного товару	Вихід товару на ринок в коротші строки з не повною, але достатньою, функціональністю для зацікавлення усіх цільових аудиторій; Проведення рекламної компанії

Таблиця 4.7. Фактори можливостей

	Фактор	Зміст можливості	Можлива реакція компанії
1	Новий продукт	Вихід на ринок, Зменшення монополії, Надання нових рішень у сфері	Розробка нової функціональності; Вихід нової продукції на ринок; Надання різноманітних типів ліцензій в залежності від потреб користувача \ замовника.
2	Вихід аналогу	Надати продукт з певними характеристиками та можливостями що відсутні у компаній конкурентів	Аналіз ринку та користувачів задля задоволення їх потреб та надання функціональності у найкоротші строки за ціну, котра є дешевшою ніж у продуктів-замінників.
3	Зворотній зв'язок від користувачів	Можливість отримання необхідної інформації для вдосконалення продукту	Наявність вхідних даних та реакція на них з боку команди розробників задля задоволення потреб та бажань кінцевих користувачів системи кешування даних.

Закінчення таблиці 4.17

4	Грошова винагорода за рекламу	При достатньому попиту на систему кешування даних можлива комерціалізація продукту на основі реклами задля отримання грошової винагороди для подальшого розвитку продукту та оплати заробітної плати працівникам	Точкова комерціалізація продукту; Введення реклами; Ведення додаткових коштів у проект задля його подальшого розвитку.
---	-------------------------------	--	--

Таблиця 4.9. Ступеневий аналіз конкуренції на ринку

№	Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1	Тип конкуренції: монополістична	Товар від кожної компанії на ринку, являється недосконалим замінником товару, реалізованого іншими фірмами; На ринку є умови для входу та виходу; Ціна корелює між суперниками;	Розробка продукту з характеристиками, які покривають сфери вживання що не покривають інші товари-замінники; Кореляція цін у відповідності до товарів замінників; Різні типи ліцензій.
2	Рівень конкурентної боротьби: світовий	Всі продукти замінники розроблялись інтернаціональними командами з різних куточків світу, продукти не належать до певної держави	Вихід на ринок збуту продукту з клієнто-необхідною функціональністю; Налагодження маркетингу на основних Інтернет ресурсах задля охоплення великої кількості потенційних користувачів; Надання бета-версій продукту.

Закінчення таблиці 4.9

3	Галузева ознака: внутрішньогалузева	Даний тип продукту може використовуватися тільки у сфері розробки ІТ додатків \ продуктів	Надання зручного, інтуїтивно зрозумілого інтерфейсу; Підтримка всім відомих методів взаємодії з середовищем розробки; Наявність документації та онлайн підтримки.
4	Конкуренція за видами товарів: товарно-видова	Дана конкуренція – конкуренція між товарами одного виду.	Впровадження функціональності яка відсутня у товарів-замінників; Спрощення інтерфейсів; Надання підтримки.
5	Характер конкурентних переваг: цінова та не цінова	Цінові переваги – точкова комерціалізація; Не цінова – надання функціональності, що відсутня у товарах-замінниках.	Надання платних ліцензій лише на критично важливу функціональність для клієнта з певним строком підтримки, що зазначена у відповідній ліцензії; Впровадження унікальної функціональності.
6	За інтенсивністю: марочна	Наявність унікального знаку що відрізняє даний продукт від продуктів-замінників	Впровадження власної назви та власного знаку.
7	Рівень конкурентної боротьби: світовий	Всі продукти замінники розроблялись інтернаціональними командами з різних куточків світу, продукти не належать до певної держави	Вихід на ринок збуту продукту з клієнто-необхідною функціональністю; Налагодження маркетингу на основних Інтернет ресурсах задля охоплення великої кількості потенційних користувачів; Надання бета-версій продукту.

Таблиця 4.10. Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	HUAWEI	Samsung	HUAWEI, Samsung	Форуми, точки продажу додатку	Розумні камери

Висновки	Прямі конкуренти намагаються сконцентруватися на інших напрямках своїх продуктів	На даний момент їх сервіс який може дати нам конкуренцію не дорівлений і не впроваджений	Постачальники диктують умови збереження даних, які захищають приватність. Також постачальники не дають змогу зловживати нормами.	Клієнти можуть диктувати умови на ринку через повідомлення на форумах або в полі відгуків в точках продажу додатку.	Можливість введення стандартизації роботи алгоритмів розпізнавання ситуацій
----------	--	--	--	---	---

Проаналізувавши можливості роботи на ринку з огляду на конкурентну ситуацію можна зробити висновок: оскільки кожний з існуючих продуктів не впливає у великій мірі на поточну ситуацію на ринку в цілому, кожний з існуючих продуктів має свою специфічну сферу використання та свої позитивні та негативні сторони щодо рішення певних типів задач, то робота та вихід на даний ринок є можливою і реалізованою задачею.

Для виходу на ринок продукт повинен мати функціонал що відсутній у продуктів-аналогів, повинен задовольняти потреби користувачів, мати необхідний та достатній функціонал з конфігурування, підтримку зі сторони розробників та можливість розробки спеціального функціоналу за відповідною ліцензією.

Таблиця 4.11. Обґрунтування факторів конкурентоспроможності

	Фактор конкурентоспроможності	Обґрунтування
1	Прагматичність	Через запуск стартапу система буде не дуже складно з точки зору архітектури перший час. Через певний період із додаванням функціоналу та оптимізації алгоритмів роботи програмний код буде все складнішим. Такий етап наступить не раніше одного року постійної роботи над проектом.
2	Зручність	Оскільки стартап розробляється на багатьох платформах з різною шириною екранів, то

		зручність використання системи на різних пристроях буде відігравати не малу роль у спроможності конкурувати з іншими гравцями ринку
3	Швидкість роботи	Швидкість роботи відіграє велику роль для користувачів, оскільки вони не будуть готові чекати декілька хвилин на виведення результату роботи додатку.
4	Оптимізація	Якщо додаток буде дуже часто видавати помилки при роботі, то користувачі не будуть вважати додаток надійним.
5	Приватність	В останні роки приватність людей та інформація щодо них все частіше зловживається шахраями або великими корпораціями, які потребують погодження з умовами доступу до приватної інформації та її обробки..
6	Технічна підтримка	Якщо технічна підтримка компанії буде працювати своєчасно та швидко, то це допоможе зберегти репутацію компанії на відміну від конкурентів, де їй не приділяють увагу.

Таблиця 4.12. Порівняльний аналіз сильних та слабких сторін системи кешування мало змінних даних

	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з запропонованим						
			3	2	1	0	1	2	3
1	Прагматичність	2	HH UA WEI	Sams ung	-	0			
2	Зручність					0	HH UA WEI	Sam sung	

3	Швидкість роботи	6		HUA WEI		Sam sun g			
4	Оптимізація	4			Sams ung	HU A W EI			
5	Приватність	0	HU AW EI	Sams ung		0			
6	Технічна підтримка	0		HUA WEI		0	Sam sung		

Таблиця 4.13. SWOT аналіз стартап-проекту

<p>Сильні сторони (S):</p> <ul style="list-style-type: none"> – Швидкість роботи – Автономність – Приватність – Відкритий вихідний код 	<p>Слабкі сторони (W):</p> <ul style="list-style-type: none"> – Труднощі з налаштуванням – З'єднання камери із комп'ютером – Непередбачені ситуації
<p>Можливості (O):</p> <ul style="list-style-type: none"> – Приватні контракти – Державні контракти – Розробка апаратного забезпечення 	<p>Загрози (T):</p> <ul style="list-style-type: none"> – Складність роботи алгоритму при невиявлених випадках використання додатку. – Небажання користувача розбиратись

Таблиця 4.14. Альтернативи ринкового впровадження стартап-проекту

№	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Безкоштовне надання певного функціоналу у користування споживачам на обмежений термін	Головний ресурс – люди, даний ресурс – наявний	2-3 місяці
2	Реклама	Залучення власних коштів для реклами товару	1-2 місяці

3	Написання статей та опис товару на відомих ресурсах	Головний ресурс – час, даний ресурс – наявний	2-3 тижні
4	Презентація товару на хакатонах й інших ІТ заходах	Ресурс – час та гроші для участі, наявні	1-3 місяці

4.4. Розроблення ринкової стратегії проекту

Таблиця 4.15. Вибір цільових груп потенційних споживачів

№	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Домовласники, які хочуть збільшити безпеку власної території завдяки відеоспостереженню	Присутня	Середній	Присутня	Легка

Закінчення таблиці 4.15

2	Підприємці/охорона, які бажають зменшити витрати шляхом автоматизації розпізнавання	Присутня	Середній	Присутня	Середня
Які цільові групи обрано: 1, 2					

Відповідно до проведеного аналізу можна зробити висновок, що підходящою цільовою групою для розповсюдження даного програмного продукту є домовласники схильні до тривожності та підприємці. Відповідно до стратегії охоплення ринку збуту товару обрано стратегії цільового маркетингу та особистого контакту.

Таблиця 4.16. Визначення базової стратегії розвитку

Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку
Надання функціональності що відсутня у товарів-замінників, підтримка клієнтів	Проведення реклами, освітлення унікальної функціональності через інтернет ресурси та інші канали, контакт напряду з споживачами;	Зниження ступеню заміності товару; Прихильність клієнтів; Відмітні властивості товару; Відмітні характеристики товару;	Стратегія диференціації

Таблиця 4.17. Визначення базової стратегії конкурентної поведінки

Чи є проект «першопрохідцем» на ринку	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, які?	Стратегія конкурентної поведінки
Ні, оскільки є товари-замінники, але дані товари замінники не мають	Так, ціль компанії знайти нових споживачів та, частково, забрати існуючих у	Компанія частково копіює характеристики товару конкурента, основна ціль компанії розробка нового	Стратегія заняття конкурентної ніші

деякого необхідного функціоналу	конкурентів задля задоволення потреб останніх	унікального функціоналу, з підтримкою основного функціоналу конкурентів	
---------------------------------	---	---	--

Таблиця 4.18. Визначення стратегії позиціонування

№	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту
1	Ефективність	Диференціація	Зменшення тривоги через видимість безпеки	Висока ефективність розпізнавання небезпечних ситуацій
2	Відкритість вихідного коду	Диференціація	Перспектива розвитку проекту	Розвиток в науці
3	Приватність	Заняття конкурентної ніші	Ваші дані належать тільки вам	Захищеність особистої інформації

Відповідно до проведеного аналізу можна зробити висновок, що стартап-компанія вибирає як базову стратегію розвитку – стратегію диференціації, як базову стратегію конкурентної поведінки – стратегію заняття конкурентної ніші.

4.5 Розроблення маркетингової програми стартап-проекту

Таблиця 4.19. Визначення ключових переваг концепції потенційного товару

№	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
---	---------	----------------------------	--

1	У власному домі	Забезпечення безпеки	Швидкість роботи, приватність
2	На підприємстві	Економія ресурсів	Відкритий вихідний код

Таблиця 4.20. Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
1. Товар за задумом	Система розпізнавання об'єктів і поведінки		
2. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх/Тл/Е/Ор
	Зручність	Нм	Е
	Швидкість роботи	Нм	Тх
	Оптимізація	Нм	Тх
	Технічна підтримка	Нм	Тх
	Приватність	Нм	Тх
	Програма упакована в 1 бінарний файл		

Закінчення таблиці 4.20

3. Товар із підкріпленням	До продажу: наявна повна документація, акції на придбання декількох ліцензій, знижки для певних сегментів на покупку товару
	Після продажу: додаткова підтримка спеціалістів налаштування, підтримка з боку розробника
За рахунок чого потенційний товар буде захищено від копіювання: захист інтелектуальної власності, патент	

Таблиця 4.21. Визначення меж встановлення ціни

Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
100-500\$/р	100-1000\$/р	500-5000\$/місяць	50-100\$/р

Таблиця 4.22. Формування системи збуту

Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
Всі користувачі будуть купляти товар по одинці	Можливість скачувати додаток влюбий час, у будь-якому місці	2 рівня (посередник + клієнт)	Роздріб

Таблиця 4.23. Концепція маркетингових комунікацій

№	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Домовласники	Інтернет, конференції, приватні зустрічі	Безпека	Показати можливість користування	Рекламне звернення спрямовано до потенційних клієнтів, користування системою
2	Підприємці		Економія	Показати можливість продукцією для клієнта	

Як результат було створено ринкову (маркетингову) програму, що включає в себе визначення ключових переваг концепції потенційного товару, опис моделі товару, визначення меж встановлення ціни, формування системи збуту та концепцію маркетингових комунікацій.

Висновки по розділу

В четвертому розділі описано стратегії та підходи з розроблення стартап-проекту, визначено наявність попиту, динаміку та рентабельність роботи ринку, як висновок було вказано що існує можливість ринкової комерціалізації проекту. Розглянувши потенційні групи клієнтів, бар'єри входження, стан конкуренції та конкурентоспроможність проекту було встановлено що проект є перспективним. Розглянуто та вибрано альтернативу впровадження стартап-проекту та доведено доцільність подальшої імплементації проекту.

ВИСНОВКИ

У першому розділі було проаналізовано існуючі рішення для розпізнавання загроз, вказані їх недоліки і обґрунтування необхідності створення нового рішення.

Нейронні мережі можуть дати нам певне уявлення про те як мозок працює, як інтегрує інформацію. Але справжнє питання полягає в тому, чому і як вся ця обробка в людині супроводжується досвідченим внутрішнім життям і чи може штучна нейронна мережа досягти такого рівня як біологічна нейронна мережа яка є в людині з народження?

В другому розділі ми розглянули різні нейронні мережі а саме вони нас змушують задуматися, чи можуть нейронні мережі стати пензлем для художників –новим способом створення, візуальні концепції, або, можливо, навіть пролити світло на коріння творчого процесу в цілому. Загалом, нейронні мережі зробили комп'ютерні системи більш корисними, роблячи їх більш схожими до біологічної нейронної мережі людини. Таким чином, наступного разу, коли ви думаєте, що вам може сподобатися, то ваш мозок буде настільки ж надійним, як комп'ютер, подумайте ще раз будьте вдячними, що у вас є така чудова нейронна мережа, яка вже встановлена у вашій голові!

Протягом останнього десятиліття в області машинного навчання домінують так звані глибокі нейронні мережі, які користуються перевагами покращення в обчисленні потужності та доступності даних. Підтип нейронної мережі має назву згорткова нейронна мережа, добре підходить для завдань пов'язаних з зображенням. Мережа навчається шукати різні функції, такі як краї, кути та кольорові відмінності, та об'єднати їх у більш складні фігури. Для

виявлення об'єкта система має оцінювати розташування імовірних об'єктів та їх класифікувати.

В третьому розділі ми на практиці, ми створили робочу програму MATLAB Fast R-CNN. Ми дізналися, що найбільш складною частиною впровадження глибокої системи навчання є збір даних для навчання та проведення самого навчання. Доступні набори даних є корисною відправною точкою як для дослідження, так і для практичних реалій. Тривалість навчання може бути ще скорочена за допомогою попередньо навченої мережі. Навіть якщо кінцева система не має тих же класів об'єктів, що і еталонні дані, візуальні проблеми є достатньо універсальними, аби отримати вигоду від детекторів, навчених для іншої проблеми. Оптимальні нижні шари згорткової мережі часто схожі, незалежно від проблеми, так само, як людське око використовує ті ж поля для всіх візуальних завдань. Таким чином, має сенс ініціалізувати шари за допомогою попередньо навченої мережі.

Що стосується реалізації, що не існує простих "нестандартних" рішень для ефективного впровадження згорткових мереж. Поточні програмні засоби, такі як Caffe та MatConvNet, вимагають спеціальних навичок. Якщо можна використовувати такі інструменти, створення робочої реалізації не є надто складним. Тим не менше, інструменти досить вигадливі щодо сумісності версій програмного забезпечення та апаратного забезпечення..

В четвертому розділі описано стратегії та підходи з розроблення стартап-проекту, визначено наявність попиту, динаміку та рентабельність роботи ринку, як висновок було вказано що існує можливість ринкової комерціалізації проекту. Розглянувши потенційні групи клієнтів, бар'єри входження, стан конкуренції та конкурентоспроможність проекту було встановлено що проект є

перспективним. Розглянуто та вибрано альтернативу впровадження стартап-проекту та доведено доцільність подальшої імплементації проекту.

ПЕРЕЛІК ПОСИЛАНЬ

1. Саймон Хайкин. Нейронные сети: Полный курс. Киев. 2008. 1103 с.
2. Paul Viola. Detecting Pedestrians Using Patterns of Motion and Appearance. 2005. Vol. 63(2). P. 153-161. DOI: 10.1007/s11263-005-6644-8.
3. Evgeniy Bart, Evgeny Byvatov and Shimon Ullman. View Invariant Recognition Using Corresponding Object Fragments. 2004. vol. 3022. P. 152-165.
4. Liu Y. and Goto S. An efficient and accurate approach of circular object detection in color images. 2014. vol. 40. P. 26–36.
5. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. Мир. 1992. 240 с.
6. Agoston M.K. Computer graphics and geometric modeling: Implementation and algorithms. 2005. P. 290-306. DOI: 10.1007/b138805.
7. Amosov O.S. High-speed neurofuzzy algorithms for filtering the mobile object trajectory parameters. 2016. P. 389-392.
8. Jang-Shing Roger Jang. ANFIS: adaptive-network-based fuzzy inference system. 1993. Vol. 23(3). P. 665-685. DOI: 10.1109/21.256541.
9. Ouyang W. Joint Deep Learning for Pedestrian Detection. 2013. P. 2046-2063. DOI: 10.1109/ICCV.2013.257.
10. Enzweiler M. Monocular Pedestrian Detection: Survey and Experiments. 2009. Vol. 31(12). P. 2169-2195. DOI: 10.1109/TPAMI.2008.260.
11. Dalal N., and Triggs B. Histograms of oriented gradients for human detection. In Computer Vision and Pattern Recognition. 2005. vol. 1, P.876–893.

12. Felzenszwalb P., Felzenszwalb F. and Huttenlocher D., Huttenlocher P. Efficient graphbased image segmentation. *International journal of computer vision* 59.2004. P.167–181.
13. Hoiem D., Efros A., Efros A., and Hebert, M. Putting objects in perspective. *International Journal of Computer Vision* 80.2008. P.3–35.
14. Krizhevsky A., Sutskever I. and Hinton G. E. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 2012. P.1047–1105.
15. M. Brand and V. Kettner, “Discovery and Segmentation of Activities in Video,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 844-851, 2000.
16. A. Bobick and A. Wilson, “A State-Based Approach to the Representation and Recognition of Gesture,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 12, pp. 1325-1337, 1997.
17. A. Dick and M. Brooks, “Issues in Automated Visual Surveillance,” in *Proceedings of International Conference on Digital Image Computing: Techniques and Application*, pp. 195-204, 2003.
18. R. Collins, A. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Yin, D. Tolliver, N. Enomoto, and O. Hasegawa, “A System for Video Surveillance and Monitoring,” *Technical Report CMU-RI-TR-00-12*, Carnegie Mellon University, 2000.
19. P. Kumar, A. Mittal, and P. Kumar, “Study of Robust and Intelligent Surveillance in Visible and Multi-modal Framework,” *Informatica* 32, pp. 63-77, 2008.

20. Y. Ivanov and A. Bobick, "Recognition of Visual Activities and Interactions by Stochastic Parsing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 852-872, 2000.
21. A. Bobick and J. Davis, "The Recognition of Human Movement Using Temporal Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 3, pp. 257-267, 2001.
22. Performance Evaluation and Tracking and Surveillance (PETS) 2007; Web site: <http://pets2007.net/>
23. C. Stauffer and W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 246-252, 1999.
24. L. Brown, A. Hampapur, J. Connell, M. Lu, A. Senior, C. Shu, and Y. Tian, "IBM Smart Surveillance System (S3): An open and extensible architecture for smart video surveillance," 2005.
25. O. Javed and M. Shah, "Tracking and Object Classification for Automated Surveillance," *Proceedings of the 7th European Conference on Computer Vision, Part-IV*, pp. 343-357, 2002.
26. A. Watson and A. Ahumada, Jr., "Model of Human Visual-Motion sensing," *J. Opt. Soc. Am., A* 2, pp. 322-342, 1985.

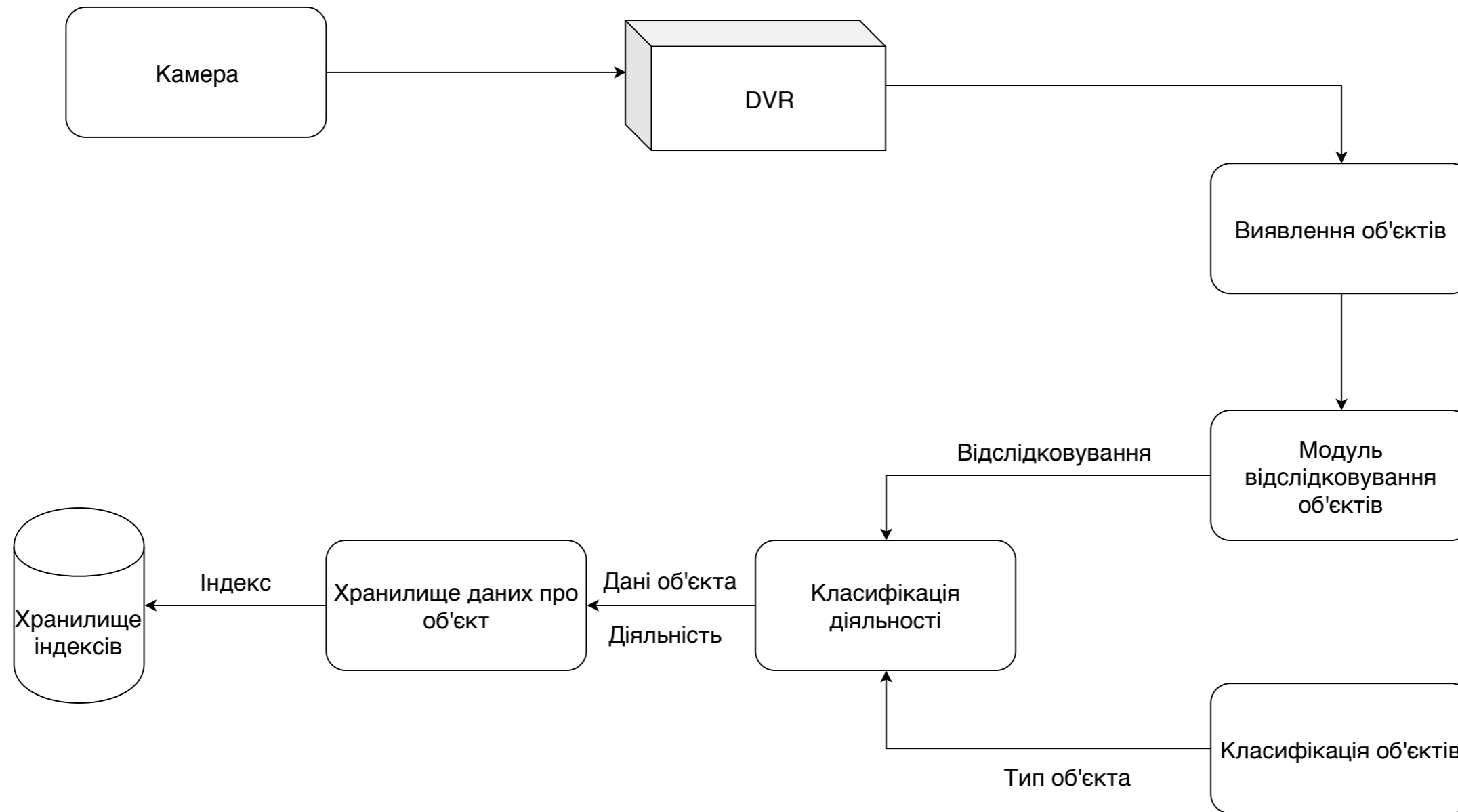
ДОДАТКИ

ДОДАТОК А

Результат перевірки на антиплагіат

ДОДАТОК Б
Ілюстраційні матеріали

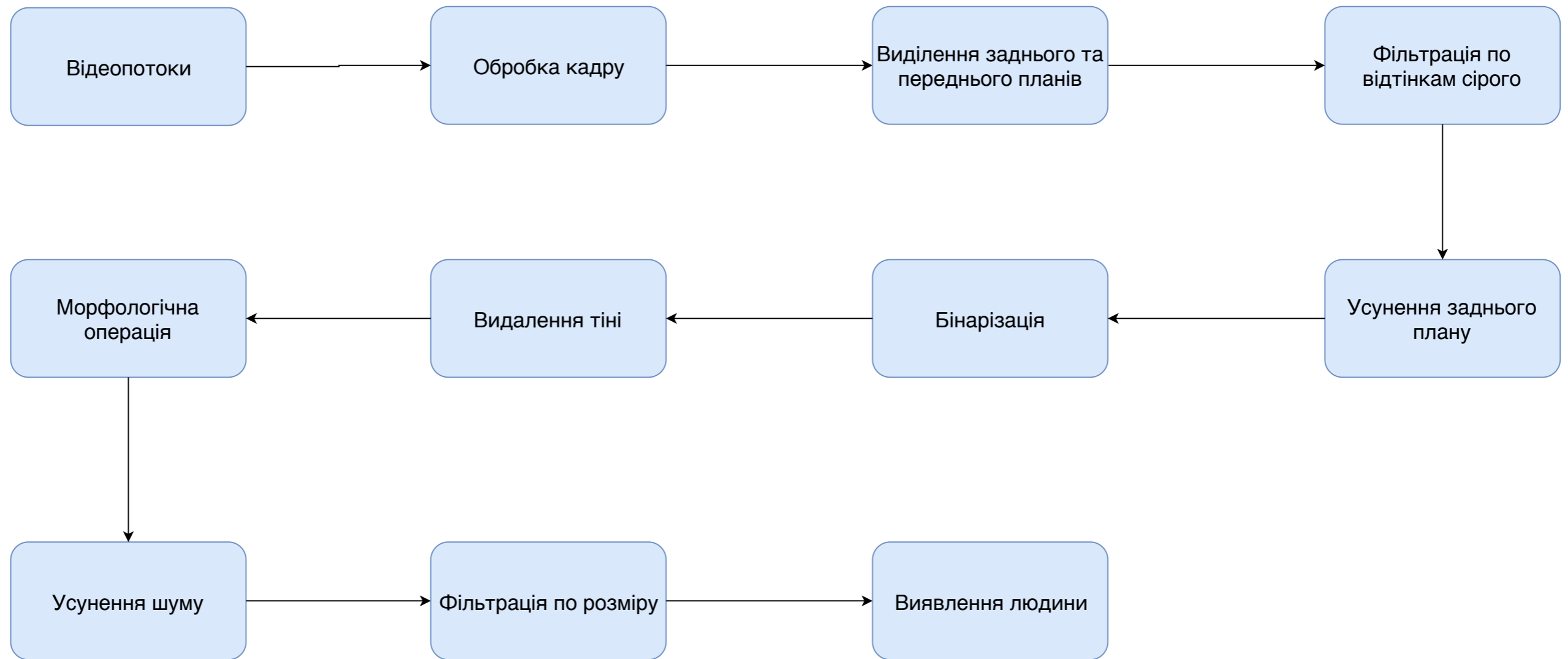
Загальна структура розумної системи відеоспостереження



Демонстраційний плакат №1
до дипломної роботи на тему
«Система відеоспостереження розумного дому з інтелектуальним
розпізнаванням загроз»»

Виконав: студент гр. ІК-81мп Петрунів О.Р.
Керівник: д.т.н., доцент Борукаєв З.Х.

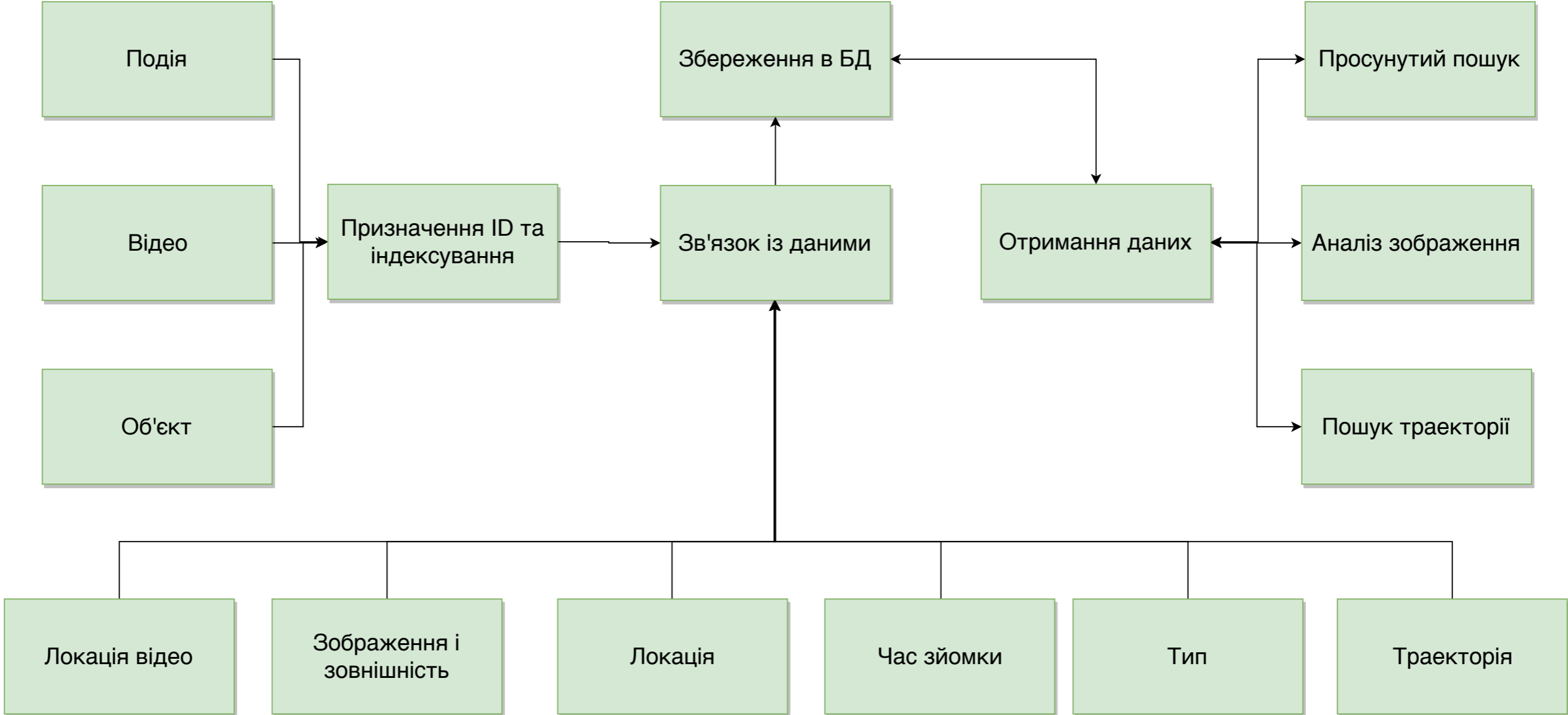
Графічне представлення процесу Виявлення людини



Демонстраційний плакат №2
до дипломної роботи на тему
«Система відеоспостереження розумного дому з інтелектуальним
розпізнаванням загроз»»

Виконав: студент гр. ІК-81мп Петрунів О.Р.
Керівник: д.т.н., доцент Борукаєв З.Х.

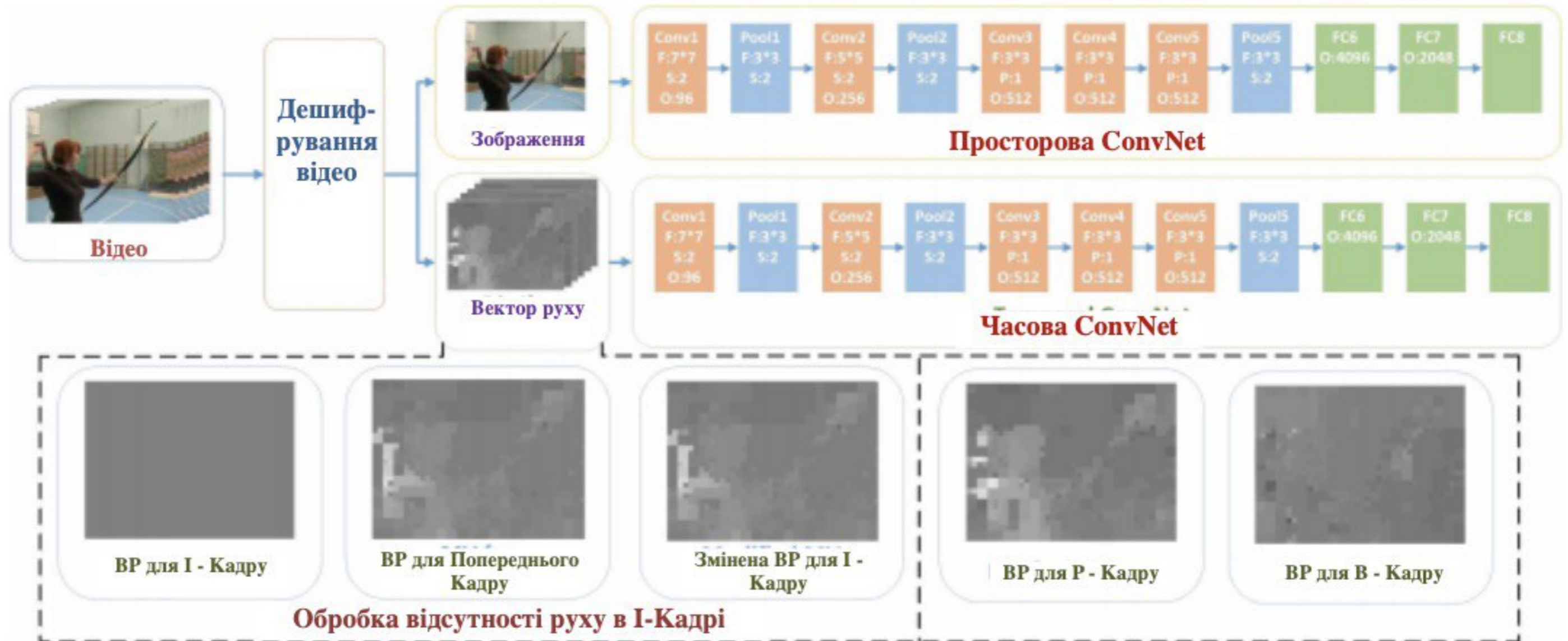
Схема індексування та отримання даних



Демонстраційний плакат №3
до дипломної роботи на тему
«Система відеоспостереження розумного дому з інтелектуальним
розпізнаванням загроз»»

Виконав: студент гр. ІК-81мп Петрунів О.Р.
Керівник: д.т.н., доцент Борукаєв З.Х.

Структура системи розпізнавання дій в режимі реального часу



Демонстраційний плакат №4
до дипломної роботи на тему
«Система відеоспостереження розумного дому з інтелектуальним
розпізнаванням загроз»

Виконав: студент гр. ІК-81мп Петрунів О.Р.
Керівник: д.т.н., доцент Борукаєв З.Х.

Алгоритм розпізнавання ситуацій і комбінування результатів

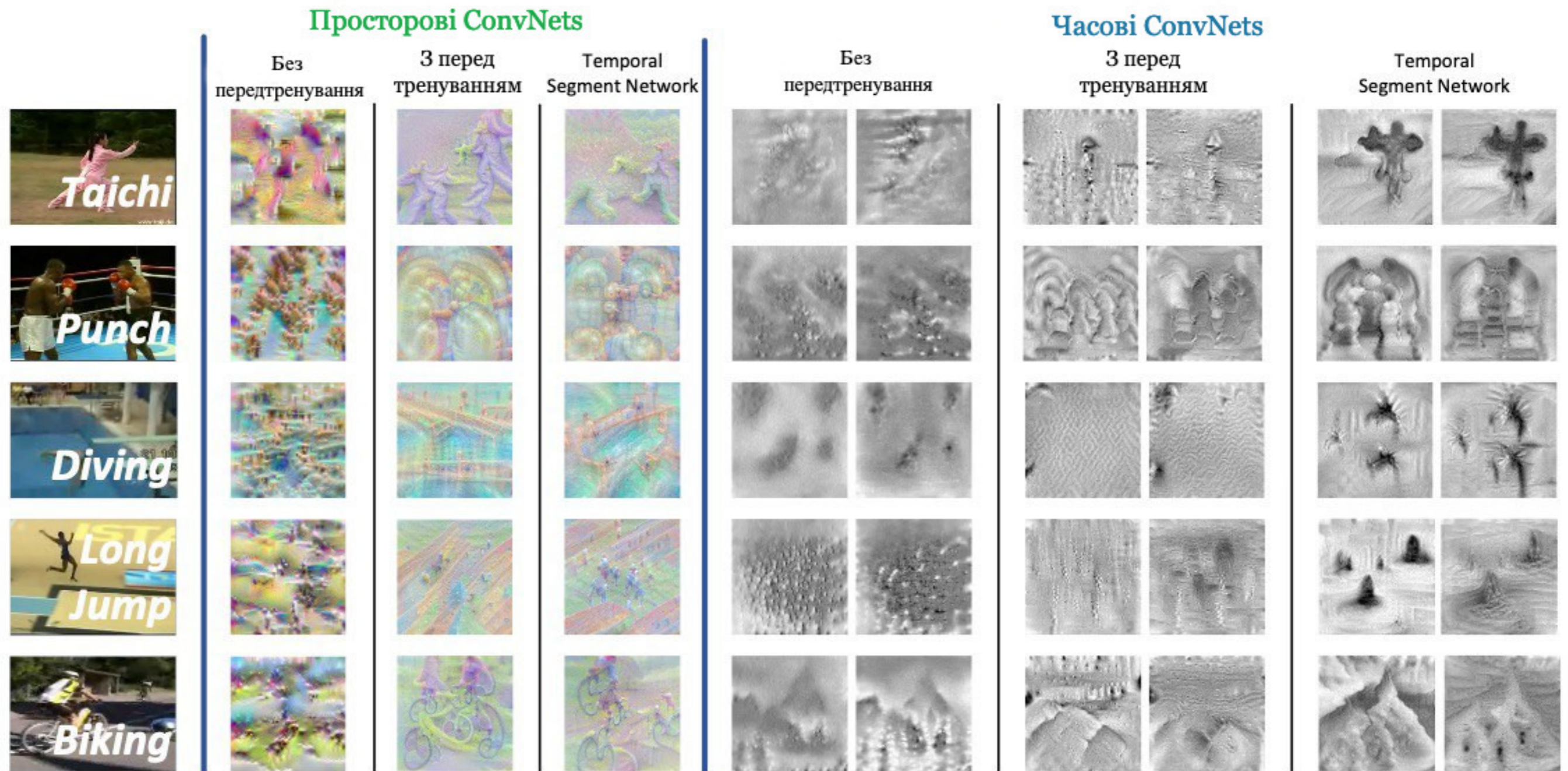


Опис вхідних даних
 А - події
 ttl - час застаріння події
 apr - апіорні ймовірності подій
 Ce - еталони ситуацій
 mp - мінімально можлива ймовірність ланцюжка подій
 tO - час на збір подій для розпізнавання ситуації

Демонстраційний плакат №5 до дипломної роботи на тему «Система відеоспостереження розумного дому з інтелектуальним розпізнаванням загроз»

Виконав: студент гр. ІК-81мп Петрунів О.Р.
 Керівник: д.т.н., доцент Борукаєв З.Х.

Візуалізація ConvNet моделей для розпізнавання дій



Демонстраційний плакат №6
до дипломної роботи на тему
«Система відеоспостереження розумного дому з інтелектуальним
розпізнаванням загроз»

Виконав: студент гр. ІК-81мп Петрунів О.Р.
Керівник: д.т.н., доцент Борукаєв З.Х.