Міністерство освіти і науки України
Одеський національний політехнічний університет

Інститут штучного інтелекту і робототехніки
**Кафедра «Комп'ютерні системи»**

**Кір'як Ярослав Ігорович,**
студент групи АК-151

# КВАЛІФІКАЦІЙНА РОБОТА МАГІСТРА

Дослідження методів підвищення ефективності систем моніторінгу соціальної дистанції

Напрям підготовки: 123 – "Комп'ютерна інженерія"

Спеціалізація: Спеціалізовані комп'ютерні системи

**Керівник:**
Стрельцов Олег Васильович,
к.т.н., доцент

Одеса — 2020

Одеський національний політехнічний університет

Інститут, факультет _____Інститут штучного інтелекту і робототехніки_____
Кафедра_____Комп'ютерних систем_____
Рівень вищої освіти_____другий, магістр_____
Освітньо-кваліфікаційний рівень___бакалавр_____
Спеціальність__123 Комп'ютерна інженерія_____
(шифр і назва)
Спеціалізація ___Спеціалізовані ком'ютерні системи_____

**ЗАТВЕРДЖУЮ**
**Завідувач кафедри**
_____

“____” _____20___року

# З А В Д А Н Н Я
## НА КВАЛІФІКАЦІЙНУ РОБОТУ

_____Кір’як Ярослав Ігорович_____
(**прізвище, ім’я, по батькові)**

1. Тема роботи___Дослідження методів підвищення ефективності систем моніторингу соціальної дистанції_____
_____
Керівник роботи__Стрельцов Олег Васильович, к.т.н., доцент_____,
( прізвище, ім’я, по батькові, науковий ступінь, вчене звання)
затверджені наказом ректора ОНПУ від “_16_” __листопада____2020 року № 467-в
2. Строк подання студентом роботи_____
3. Вихідні дані до роботи_____
_____
_____
_____
_____
4. Зміст роботи
_____
_____
_____
_____
_____
_____
5. Перелік люстративного матеріалу
_____
_____
_____
_____

6. Консультанти розділів роботи

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---|---|---|---|
| | | завдання видав | завдання прийняв |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

7. Дата видачі завдання_____

## КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапів кваліфікаційної роботи | Строк виконання етапів роботи | Примітка |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

**Студент**_____ _____Кір'як **Я.** І._____
( підпис ) (прізвище та ініціали)

**Керівник роботи** _____ __Стрельцов О. В.___
( підпис ) (прізвище та ініціали)

# ВСТУП

Нове покоління коронавірусної хвороби (COVID-19) було зареєстровано наприкінці грудня 2019 року в місті Ухань, Китай. Буквально через кілька місяців вірус став глобальним спалахом у 2020 році. У травні 2020 року Всесвітня організація охорони здоров'я (ВООЗ) оголосила ситуацію пандемією.

З огляду на тенденцію збільшення кількості пацієнтів, досі немає ефективних ліків або доступного лікування вірусу. Ці жорсткі умови змусили світову спільноту шукати альтернативні шляхи зменшення поширення вірусу.

Соціальне дистанціювання відноситься до запобіжних заходів щодо запобігання поширенню хвороби шляхом мінімізації близькості фізичного контакту в закритих або переповнених громадських місцях (наприклад, школах, робочих місцях, тренажерних залах, лекційних залах тощо), щоб зупинити широке накопичення ризику зараження.

Протягом декількох місяців Всесвітня організація охорони здоров'я вважала, що COVID-19 передається тільки через краплі, які вивільняються при чхання або кашлі, а вірус не затримується в повітрі. Однак 8 липня 2020 року ВООЗ оголосила: "Є нові докази того, що COVID-19 - це повітряно-крапельне захворювання, яке може поширюватися через крихітні частинки, підвішені в повітрі після того, як люди розмовляють або дихають, особливо в переповнених, закритих приміщеннях або під поганою вентиляцією". Таким чином, соціальне дистанціювання зараз вважається ще більш важливим, ніж вважалося раніше, і є одним з кращих способів зупинити поширення хвороби на додаток до носіння масок. Майже всі країни зараз бачать в цьому обов'язкову практику.

Згідно з певними вимогами ВООЗ, мінімальна відстань між людьми повинна бути не менше 6 футів (1,8 м) для того, щоб відповідати прийнятній соціальній

дистанції між людьми. Недавні дослідження підтвердили, що люди з легкими симптомами або без них також можуть бути носіями нової коронавірусної інфекції. Тому важливо, щоб всі люди підтримували контрольовану поведінку і спостерігали соціальне дистанціювання. Багато наукових робіт, таких як дослідження "5-7", показали, що дистанціювання від суспільства є ефективним не фармакологічним підходом і важливим інгібітором для обмеження передачі інфекційних захворювань, таких як H1N1, SARS і COVID-19.

Під час пандемії COVID-19 уряду намагалися впровадити різні методи соціального дистанціювання, такі як обмеження поїздок, контроль кордонів, закриття пабів і барів, а також оповіщення суспільства про необхідність підтримувати відстань від 1,6 до 2 м один від одного [1]. Однак відстежувати масштаби поширення інфекції і обсяг обмежень - непросте завдання. Людям необхідно виходити на вулицю для задоволення основних потреб, таких як: їжа, медичне обслуговування та інші необхідні справи. Тому багато інші технологічні рішення, такі як [2,3], і дослідження, пов'язані зі штучним інтелектом, такі як [4-6], намагалися втрутитися, щоб допомогти медичному співтовариству в рішенні проблем COVID-19, практикуючи соціальне дистанціювання. Ці роботи варіюються від визначення місця розташування і відстеження пацієнтів на основі GPS до сегментації і моніторингу натовпу.

У таких ситуаціях Штучний інтелект може зіграти важливу роль у забезпеченні моніторингу соціального дистанціювання. Комп'ютерний зір, як частина Штучного інтелекту, був дуже успішним у вирішенні різних складних проблем в сфері охорони здоров'я і продемонстрував свій потенціал в комп'ютерній томографії грудної клітини або розпізнаванні COVID-19 на основі рентгенівських променів [7-8], яке може внести свій вклад і в моніторинг соціального дистанціювання. Крім того, глибокі нейронні мережі дозволяють нам отримувати складні функції з даних, щоб ми могли забезпечити більш точне розуміння зображень шляхом аналізу і класифікації цих функцій. Приклади включають діагностику, клінічний ведення і лікування, а також профілактику і контроль COVID-19 [9-10].

Можливі проблеми в цій галузі полягають у важливості досягнення високого рівня точності, роботи з різними умовами освітлення, оклюзії і продуктивності в реальному часі. У цій роботі також прагнемо знайти рішення для згаданих складнощів.

**Основний внесок та задачі** цього дослідження можна позначити в наступному:

1. Дослідження спрямовано на підтримку скорочення поширення коронавируса і його економічних витрат, за рахунок надання рішення на основі штучного інтелекту для автоматичного моніторингу та виявлення порушень соціального дистанціювання між людьми.
2. Розробляється надійна модель глибокої нейронної мережі (DNN) для виявлення, відстеження і оцінки людей, якій даємо назву KeepDistanse. У порівнянні з деякими недавніми роботами в цій області, такими як [5], пропонуємо більш швидкі і точні результати.
3. Підтверджуємо достовірність наших експериментальних результатів, виконуючи великі тести і оцінки в різних наборах даних всередині і поза приміщеннями, які перевершують сучасні (Таблиця 3.1).
4. Розроблена модель може використовуватися як звичайна система виявлення та відстеження людей, не обмежуючись моніторингом соціального дистанціювання, і її можна застосовувати для різних додатків, таких як виявлення пішоходів в автономних транспортних засобах, розпізнавання дій людини, виявлення аномалій, системи безпеки.

**Мета роботи** – це удосконалення системи моніторингу соціальної дистанції за допомогою дослідження та моделювання методів підвищення точності та скорості виявлення на основі Комп'ютерного зору, а саме покращення моделі виявлення людини у натовпі.

**Практична цінність** даної роботи розкривається саме у необхідності контролю та витримуванні соціальної дистанції, як найголовнішому способі зупинки розповсюдження захворювання COVID-19. Представлена у даній роботі модель, є

точнішою та швидшою у виявленні людини за аналогічні системи. Розроблена модель пропонує незалежний від точки зору алгоритм класифікації людини. Тому, незалежно від кута та положення камери, результат цього дослідження є прямим застосуванням для більш широкого кола дослідників, не тільки в галузі комп'ютерного зору, штучного інтелекту та охорони здоров'я, також і в інших галузях промисловості, включаючи виявлення пішоходів.

**Об'єктом дослідження** являються системи моніторингу соціальної дистанції на основі різних методів, особливо включаючи Комп'ютерний зір, тому що на сьогоднішній день такі передови технології найвищого рівня, як Artificial Intelligence, Machine Learning, Deep Learning, Computer Vision проявили себе у найрізноманітніших сферах, особливо у медичній, та з кожним днем все більше інтегруються у всі види систем, від стартапів до глобальних корпораційних систем.

**Предметом дослідження** є підвищення ефективності системи моніторингу соціальної дистанції в різних середовищах за рахунок підвищення точності та скорості визначення.

Більш детальна і додаткова інформація будуть надані в наступних розділах. У Розділі 1 обговоримо інші роботи, пов'язані з технічною стороною, існуючі проблеми та прогалини в дослідженнях цієї області. Пропонована методологія, включаючи архітектуру моделі і наші методи виявлення і стеження об'єктів, буде запропонована в Розділі 2. У Розділі 3 експериментальні результати і продуктивність системи будуть досліджені в порівнянні з аналогами, після чого слід обговорення і заключні зауваження.

Отже, спостерігаючи ситуацію із пандемією COVID-19, що відбувається у всьому світі, було прийнято рішення удосконалити систему контролю соціальної дистанції використовуючи Штучний інтелект. У наступних розділах представлено, як саме була покращена система контролю, на чому базується загальна система та чим особлива розроблена модель KeepDistance.

# 2 УДОСКОНАЛЕННЯ МОДЕЛІ

Пропонується трьохступенева модель, що включає виявлення людей, відстеження та оцінку відстані як загальне рішення для моніторингу соціальних дистанцій та аналізу ризику зараження на основі зони. Система може бути інтегрована та застосована до всіх типів камер відеоспостереження з будь-якою роздільною здатністю від VGA до Full-HD, з продуктивністю в режимі реального часу.

## 2.1. Детект людей

На рисунку 2.1 показана загальна структура етапу 1.

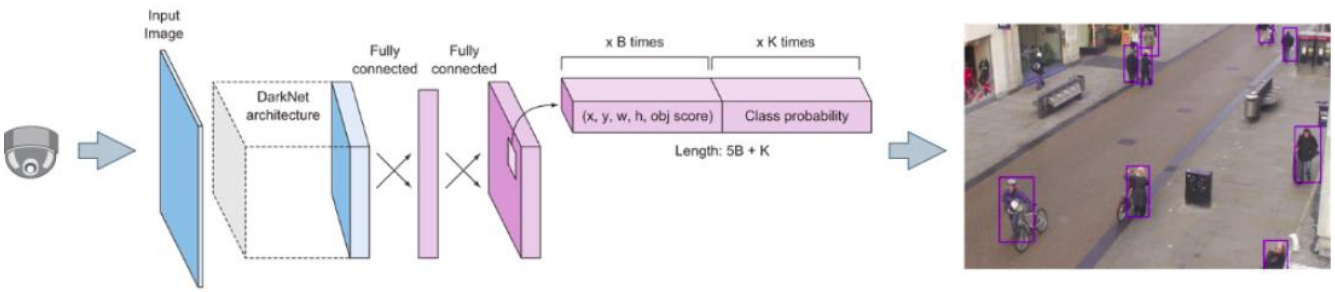


Рисунок 2.1 – Етап 1. Загальна структура модуля виявлення людей.

Камера відеоспостереження збирає вхідні відео послідовності та передає їх нашій моделі глибокої нейронної мережі. Результатом роботи моделі будуть виявлені люди на моніторі з їхніми унікальними обмежувальними рамками для локалізації. Метою є розробка надійної моделі виявлення людини (людей), здатної вирішувати різні типи проблем, таких як варіації одягу, пози, на далекій та близькій відстані, з/без оклюзії та за різних умов освітлення.

Сучасні детектори об’єктів на основі DNN (такі, як перераховані на рисунку 1.2), складаються з трьох розділів: модуля введення та пов’язаних з ним операцій, наприклад збільшення; магістраль для вилучення функцій і потім для прогнозування класів та розташування об’єктів на виході.

У таблиці 2.1 узагальнено вичерпний перелік варіантів дизайну моделі, включаючи доповнення вхідних даних, найсучасніші модулі виявлення основних об’єктів (тобто функції активації, екстрактори магістральних функцій, neck та head). Таблиця пропонує різні варіанти вибору neck, head та інших підмодулів (залежно від вимог моделі). Однак в основному зосереджуємось на вимогах нашого дослідження.

Таблиця 2.1 – Сучасні варіанти та методи проектування моделі на основі згорткової нейронної мережі (CNN). Зліва направо: введення до виходу.

| **Вхідні дані** | **Detection Core →** | | | | **Head → Вихідні дані** | |
|---|---|---|---|---|---|---|
| **Augmentation** | **Активація** | **Backbone** | **Neck** | **Regularization** | **Dense** | **Sparse** |
| CutOu | ReLU | VGG | SPP | L1, L2 | YOLO* | R-CNN* |
| MixUp | Leaky-ReLU | ResNet | ASPP | DropOut (DO) | SSD* | Fast-RCNN* |
| CutMix | Param-ReLU | SpineNet | PAN | DropPath | RetinaNet* | Faster-RCNN* |
| Mosaic | ReLU6 | CSPResNeXt50 | FPN | Spatial DO | RPN* | Libra R-CNN* |
| | SELU | CSPDarknet53 | BiFPN | DropBlock | CornerNet+ | Mask R-CNN* |
| | Swish | EfficientNet | ASFF | | MatrixNet+ | CenterNet+ |
| | Mish | Darknet53 | RFB | | R-FCN* | RepPoints+ |
| | | Inception | SFAM | | FCOS+ | |
| | | | NAS-FPN | | | |
| | | **Anchor-based*** | | **Anchor-free +** | | |

2.1.1. Вхідні дані та навчальні набори даних

Для того, щоб мати надійний детектор, потрібен набір багатих навчальних наборів даних. Сюди повинні входити люди з різною статтю та віком (чоловік,

жінка, хлопчик, дівчинка) з мільйонами точних анотацій та маркування. Відібрали два великі набори даних MS COCO та набір даних Google Open Image, які задовольняють вищезазначені очікування, надавши понад 3,7 мільйона анотованих людей. Далі, детальна інформація буде надана у Розділі 4.

Набір можливих функцій активації для BoF наведено в таблиці 1. Також було досліджена ефективність цієї моделі щодо ReLU, Leaky ReLU, SELU, Swish, Parametric RELU та Mish. Попередні оцінки підтвердили ті самі результати, що надані Misra [18] для нашого застосування для виявлення людини. Функція активації Mish (рівняння (1)) сходилася до мінімальних втрат, швидше, ніж Swish та ReLU, з більшою точністю. Результат був послідовним, особливо для різноманітності ініціалізаторів параметрів, методів регуляризації та нижчих значень швидкості навчання.

Mish:

$$f(x) = x.\tanh(\text{softplus}(x)) = x.\tanh(ln(1+e^x)) \quad (2.1)$$

з похідними:

$$f`(x) = e^x\omega / \delta^2 \quad (2.2)$$

як саморегульовану не монотонну функцію активації, де:

$$\omega = 4\,(x + 1) + 4e^{2x} + e^{3x} + (4x + 6) \quad (2.3)$$

та

$$\delta = 2e^x + e^{2x} + 2. \quad (2.4)$$

### 2.1.2. Архітектура backbone

Як показано на рисунку 1.3, YOLOv4 пропонує найкращий компроміс щодо швидкості та точності для багатокласного виявлення об'єктів; однак, оскільки YOLOv4 - це сукупність різних методів, провели поглиблене вивчення кожної з підтехнік, щоб досягти найкращих результатів для однокласної моделі виявлення людей та перевершити сучасний рівень.

Основним способом підвищення точності детекторів на базі CNN є розширення сприйнятливого поля та підвищення складності моделі за допомогою більшої кількості шарів; однак використання цієї техніки ускладнює навчання моделі. Натомість, пропонується використовувати техніку пропуску з'єднань для полегшення тренувань.

Різні моделі використовують подібну політику для встановлення зв’язків між шарами, таких як між ступеневі часткові (CSP) зв’язки [19] або щільні блоки (що складаються із пакетної нормалізації, ReLU, згортки тощо) у DenseNet. Такі моделі також використовувались при розробці деяких останніх магістральних архітектур, таких як CSPResNeXt50, CSPDarknet53 [20] та EfficientNet-B3, які є підтримуваними варіантами архітектур для YOLOv4.

Модель Yolo була розроблена для нейронної мережі на основі DarkNet. DarkNet зберігає навчені коефіцієнти (ваги) в форматі, який може бути розпізнаний за допомогою різних методів на різних платформах. Ця проблема може бути каменем спотикання, тому що може знадобитися навчити модель на потужному обладнанні, а потім використовувати її на іншому обладнанні. DarkNet написаний на C і не має іншого програмного інтерфейсу, тому, якщо вимоги платформи або власні переваги змусять звернутися до іншої мови програмування, то доведеться додатково попрацювати над його інтеграцією. Також він поширюється тільки в форматі

вихідного коду, і процес компіляції на деяких платформах може бути досить проблематичним.

Таблиця 2.2 узагальнює короткий звіт наших досліджень для вищезазначених магістральних архітектур з точки зору кількості параметрів та швидкості обробки (у fps) для одного і того ж розміру входу 512 × 512.

Таблиця 2.2 – Порівняння трьох магістральних моделей за кількістю параметрів та швидкістю (кадрів в секунду) з використанням графічного процесора RTX 2070.

| **Backbone модель** | **Роздільна здатність** | **Кількість Параметрів** | **Швидкість (fps)** |
|---|---|---|---|
| CSPResNeXt50 | 512 ×512 | 20.6 M | 62 |
| CSPDarknet53 | 512 ×512 | 27.6 M | 66 |
| EfficientNet-B3 | 512 ×512 | 12.0 M | 26 |

На основі теоретичних обґрунтувань та кількох проведених експериментів, дійшли висновку, що CSDDarknet53 є найбільш оптимальною магістральною моделлю для нашого застосування, незважаючи на вищу складність (через більшу кількість параметрів). Тут більша кількість параметрів призводить до збільшення можливостей моделі виявляти кілька об’єктів, в той же час можемо підтримувати ефективність у реальному часі.

### 2.1.3. Neck модуль

Нещодавно, деякі із запропонованих сучасних моделей розмістили додаткові шари між backbone і head, які називаються neck, що розглядається для збору функцій на різних етапах магістральної мережі.

Секція відділу neck складається з декількох шляхів зверху вниз і знизу вгору для збору та об'єднання параметрів мережі в різні шари, щоб забезпечити більш точні характеристики зображення для головного відділу.

Багато моделей, заснованих на CNN, використовують повністю пов'язані шари для класифікаційної частини, і, отже, вони можуть приймати лише фіксовані розміри зображень як вхідні дані. Це може призвести до двох типів проблем: по-перше, не можемо мати справу із зображеннями з низькою роздільною здатністю, а по-друге, виявлення дрібних предметів може бути важким. Це суперечить цілям дослідження, коли навпаки прагнемо застосувати обговорювану модель у будь-яких камерах спостереження з будь-якими розмірами та роздільною здатністю вхідного зображення. Для того, щоб вирішити перше питання, можемо звернутися до існуючих методологій, таких як Повністю згорткові мережі (Fully Convolutional Networks - FCN). Такі моделі, включаючи YOLO (в останніх версіях), не мають FC-шарів і тому можуть працювати з зображеннями різних розмірів. Однак, щоб впоратися з другою проблемою (тобто, маючи справу з дрібними предметами), виконали пірамідальну техніку для посилення сприйнятливого поля та вилучення різних масштабів зображення з магістралі і, нарешті, виконавши багатомасштабне виявлення в секції head.

У DNN нижні шари (тобто перші кілька шарів) витягують локалізовану інформацію про візерунок та текстуру для поступового накопичення семантичної інформації, яка потрібна у верхніх шарах. Однак під час процесу вилучення особливостей деякі частини локальної інформації, які можуть знадобитися для остаточного налаштування моделі, можуть бути втрачені. У підході PANet інформація про нижні шари буде додана до верхніх шарів для посилення локалізованої інформації; отже, можна очікувати кращого налаштування та прогнозування. В нещодавньому дослідженні Бочковського [22] показано, що оператор конкатенації виконує ефективніше оператора додавання для збереження локалізованої інформації та передачі їх на верхні шари.

З метою подальшого вдосконалення сприйнятливих полів та досягнення кращої потужності виявлення малих об'єктів, розглядаємо модуль мережевих пірамід функцій YOLO (FPN) для багатомасштабного виявлення. Модуль витягує функції в різних масштабах з backbone. Посилання [23] покращив YOLOv3 за допомогою модуля просторового об'єднання пірамід (SPP) замість FPN, що призводить до збільшення $AP_{50}$ на 2,7% щодо виявлення об'єкта MS COCO. Вдосконалений SPP використовує операцію максимального об'єднання замість операції "Bag of Words" для вирішення проблеми просторових розмірів та для вирішення багатомасштабного виявлення в head розділі. Метод застосовує ядро $k \times k$ ядро максимального пулу, де k = {1, 5, 9, 13}, а крок дорівнює 1.

У розділі 3 розглядається ефективність цього підходу для підвищення точності моделі на основі YOLOv4. На рисунку 2.2 показані багато масштабні head розділи, які використовували в цій мережі для виявлення об’єктів різного розміру.

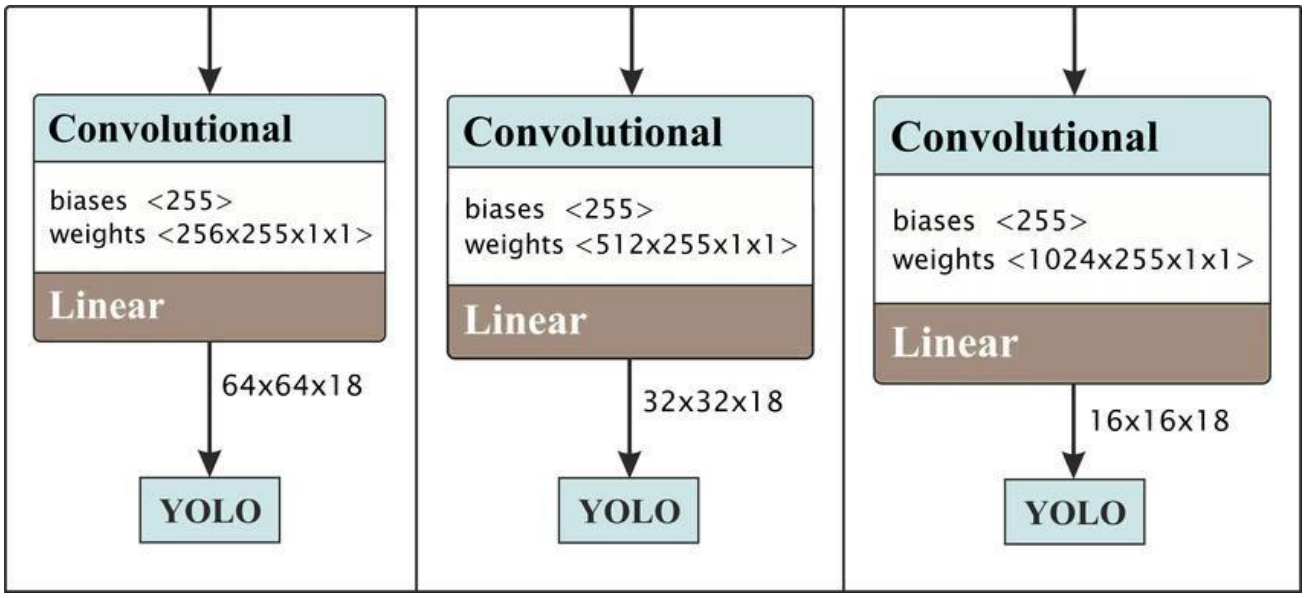


Рисунок 2.2 – Head розділі на основі YOLO, застосовувані в різних масштабах.

Експериментуючи з різними раціональними конфігураціями для neck модуля моделі, використовується (Spatial Pyramid Pooling - SPP) і PAN, а також модуль просторової уваги (Spatial Attention Module - SAM), що разом зробило один з

найбільш ефективних, послідовних і надійних компонентів для зосередження моделі на оптимізації параметрів.

### 2.1.4. Head модуль

Модуль head в DNN відповідає за класифікацію об'єктів (наприклад, людей, велосипедів, стільців тощо), а також за обчислення розміру об'єктів та координат відповідних обмежувальних рамок.

Зазвичай є два типи head секцій: одноступінчаста (щільна) та двоступінчаста (розріджена). Двоступеневі детектори використовують пропозицію регіону перед застосуванням класифікації. Спочатку детектор виділяє набір пропозицій об’єктів (обмежувальні рамки для кандидатів) шляхом вибіркового пошуку. Потім він змінює їх розмір до фіксованого розміру, перш ніж подавати їх на модель CNN. Це схоже на детектори на основі R-CNN.

Незважаючи на точність двоступеневих детекторів, такі методи не підходять для систем з обмеженими обчислювальними ресурсами.

З іншого боку, одноступінчасті детектори виконують уніфікований процес виявлення. Вони прив'язують пікселі зображення до укладених сіток і перевіряють ймовірність існування об'єкта в кожній клітинці сіток, подібно до роботи, виконаної Лю "Multi Shot Multibox Detector" (відомий як SSD), або іншими роботами, виконаними Redmon [23], і Бочковський детектори YOLO. Такі детектори використовують регресійний аналіз для обчислення розмірів обмежувальних рамок та інтерпретації їх ймовірностей класу. Цей підхід пропонує чудові вдосконалення з точки зору швидкості та ефективності.

Модель Single Shot Detector (SSD) реалізує ідею використання пірамідальної ієрархії виходів згорткової мережі для ефективного виявлення об'єктів різних

розмірів. Зображення послідовно передається на шари згорткової мережі, які зменшуються в розмірах. Вихід з останнього шару кожної розмірності бере участь в ухваленні рішення по детекції об'єктів, таким чином, складається "пірамідальна характеристика" зображення. Це дозволяє виявляти об'єкти різних масштабів, так як розмірність виходів перших шарів сильно корелює з обмежуючими рамками для великих об'єктів, а останніх - для невеликих. На відміну від YOLO, SSD не розбивати зображення на сітку довільного розміру, а пророкує зміщення ключових рамок. Ключові рамки на різних рівнях масштабуються так, що одна розмірність вихідного шару відповідає за об'єкти свого масштабу. В результаті, великі об'єкти можуть бути виявлені тільки на більш високому рівні, а маленькі об'єкти - на низьких рівнях. Як і в інших алгоритмах, функція втрат забезпечує спільний внесок як втрат локалізації, так і втрат класифікації (рисунок 2.3).

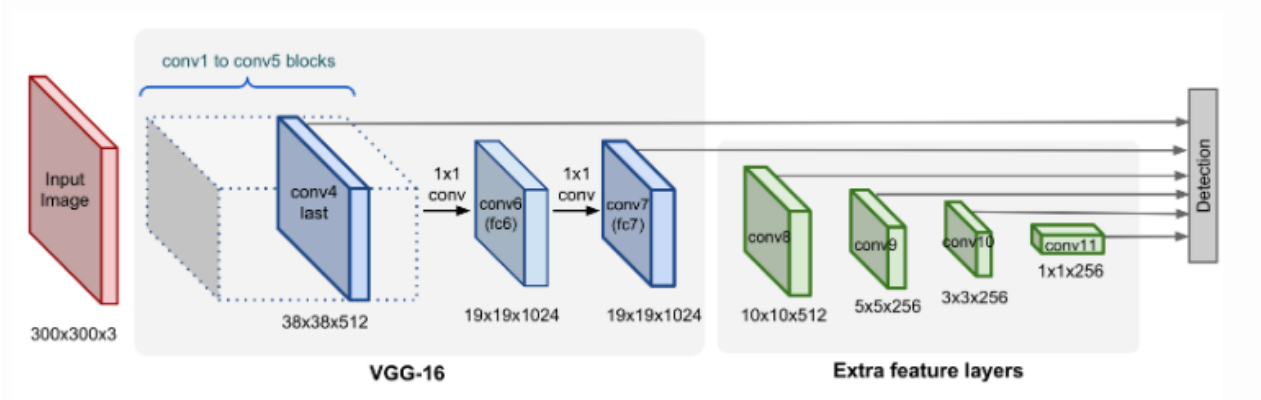


Рисунок 2.3 – Архітектура нейронної мережі для алгоритма SSD.

В секції моделі head використовуємо ту ж конфігурацію, що і YOLOv3. Подібно до багатьох інших моделей на основі якоря, YOLO використовує заздалегідь визначені поля для виявлення декількох об'єктів. Тоді модель виявлення об’єкта буде навчена передбачати кожну згенеровану прив'язку, яка належить до певного класу. Після цього буде використано зсув для регулювання розмірів анкерної коробки з метою кращого збігу даних із земними реаліями на основі класифікації та втрат регресії.

Припускаючи вихідну точку комірки сітки ($c_x$, $c_y$) у верхньому лівому куті зображення об'єкта та обмежувальну рамку перед шириною та висотою ($p_w$, $p_h$), мережа передбачає обмежувальну рамку в центрі ($x\hat{}$, $y\hat{}$) і розмір ($w\hat{}$, $h\hat{}$) з відповідним зміщенням та масштабами ($b_x$, $b_y$, $b_w$, $b_h$) наступним чином:

$$\begin{aligned} x\hat{} &= \sigma(b_x) + c_x \\ y\hat{} &= \sigma(b_y) + c_y \\ w\hat{} &= p_w e^{bw} \\ h\hat{} &= p_h e^{bh} \end{aligned} \qquad (2.5)$$

де σ - функція оцінки надійності сигмоїдної області в межах 0 та 1.

Представляємо клас "людина" з 4-ма кортежами (x, y, w, h), де (x, y) - центр обмежувальної рамки, а w, h - ширина та висота відповідно.

Використовується три якірні коробки, щоб знайти максимум трьох людей у кожній комірці сітки. Отже, загальна кількість каналів - 18: (1 клас + 1 об’єкт + 4 координати) × 3 якоря.

Оскільки для кожного просторового розташування маємо декілька опорних коробок, один об’єкт може бути пов’язаний з декількома опорними коробками. Цю проблему можна вирішити за допомогою техніки не максимального придушення (Non-maximal Suppression - NMS) та обчислення перетину над об'єднанням (Intersection Over Union - IoU) для обмеження асоціації опорних коробок.

В рамках операції коригування ваги та мінімізації втрат використовується повний IoU (CIoU) (у рівнянні (7)) замість базового IoU (рівняння (6)). CIoU не лише порівнює місце розташування та відстань обмежувальних коробок для кандидатів до обмежувальних коробок, що стосуються істини, але також порівнює співвідношення сторін розміру створених обмежувальних коробок з розміром обмежувального ящика для правдивої землі.

$$IoU = |B \cap B^{gt}| / |B \cup B^{gt}| \quad (2.6)$$

$B^{gt} = (x^{gt}, y^{gt}, w^{gt}, h^{gt})$ - це вікно з основною істиною, а $B = (x, y, w, h)$ - прогнозоване поле. Використовуємо CIoU не тільки як метрику виявлення, а й як функцію втрат:

$$L_{CIoU} = 1 - IoU + |\rho^2 (B, B^{gt})| / c^2 + \alpha v \quad (2.7)$$

, де $\rho$ - евклідова відстань між великою істиною $B^{gt}$ та передбачуваним $B$ обмежувальним полем. Довжина діагоналі найменшої обмежувальної коробки, що охоплює обидві коробки $B$ і $B^{gt}$, представлена через $c$ а $\alpha$ - позитивний параметр компромісу:

$$\alpha = v / (1 - IoU) + v \quad (2.8)$$

і $v$ вимірює узгодженість пропорцій, як показано нижче:

$$\square = \frac{4}{\square^2}(\square\square\square\square\square\square \frac{\square^{\square\square}}{\square^{\square\square}} - \square\square\square\square\square\square \frac{\square}{\square})^2 \quad (2.9)$$

Навіть у випадку нульового відсотка перекриття, функції втрат все одно дають нам вказівку про те, як відрегулювати ваги для першого зближення розміру аспекту до 1, а по-друге, як зменшити відстань похибки обмежувальних полів кандидатів до центру обмежувальної рамки. Подібний підхід, який називається Distance-IoU, використовується в [24] для іншої програми.

Для запобігання проблеми перебільшення оцінили деякі загальні методи регуляризації, як показано в таблиці 1. Подібно до наведених результатів у [25], знайшли DropBlock (DB) як один з найбільш ефективних методів регуляризації порівняно з іншими варіантами.

На рисунку 2.4 узагальнено трирівневу структуру нашого модуля детекту людини у послідовності взаємопов’язаних компонентів.

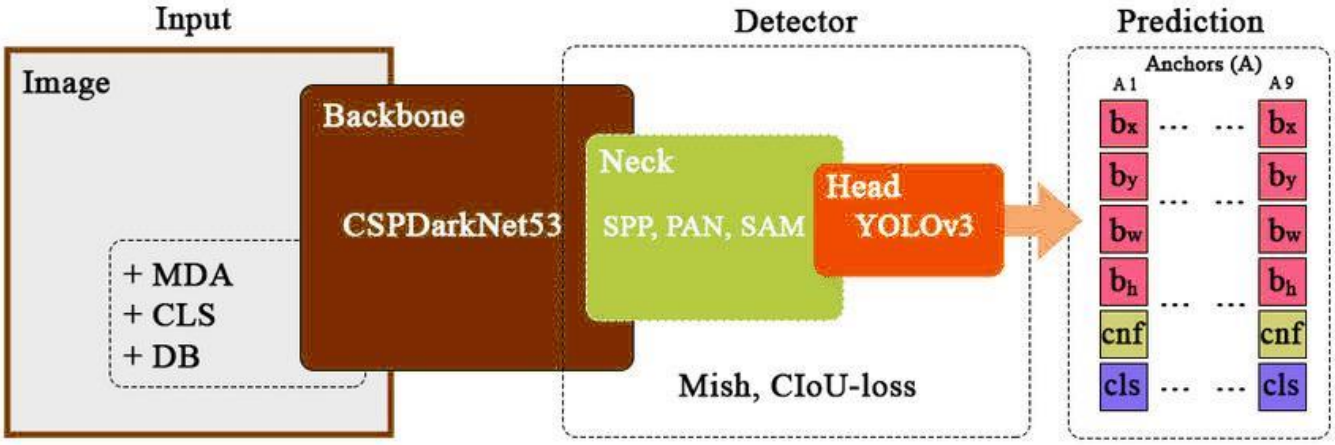


Рисунок 2.4 – Структура мережі запропонованого трирівневого модуля детекту людини.

У вхідній частині до вхідного зображення застосовується збільшення даних (Mosaic Data Augmentation - MDA), згладжування міток класів (Class Label Smoothing - CLS) та DropBlock (DB). У частині детектора була використана функція активації Mish, а метрика CIoU розглядається як функція втрат. У частині прогнозування для кожної комірки на кожному рівні анкерні поля містять інформацію, необхідну для знаходження обмежувального поля, коефіцієнт довіри об’єкта та відповідний клас об’єкта. Загалом у нас є дев’ять якірних коробок.

## 2.2. Відстеження людей

Наступним кроком після фази виявлення є відстеження людей та призначення ідентифікаторів кожної людині. Використовується проста техніка відстеження в режимі онлайн та реального часу (SORT), [22] як основу для фільтру Кальмана разом із угорською методикою оптимізації для відстеження людей. Фільтр Калмана прогнозує положення людини в момент часу $t + 1$ на основі поточного вимірювання за один раз і математичного моделювання руху людини. Це ефективний спосіб продовжити локалізацію людини у випадку оклюзії. Угорський алгоритм - це

комбінаторний алгоритм оптимізації, який допомагає призначити унікальний номер для ідентифікації даного об’єкта в наборі кадрів зображень, досліджуючи, чи є людина в поточному кадрі тією самою виявленою людиною у попередніх кадрах чи ні.

На рис. 2.5а показаний зразок виявлення людей та присвоєння ідентифікаторів, на рисунку 2.5б - шлях відстеження кожної людини, а на рисунку 2.5в - остаточне положення та статус кожного індивіда після 100 кадрів виявлення, відстеження та призначення ідентифікатора.

(а) (б) (в)

Рисунок 2.5 – Виявлення людей, присвоєння ідентифікаторів, відстеження та подання траєкторії руху. (а) - зразок виявлення людей та присвоєння ідентифікаторів, (б) - шлях відстеження кожної людини, (в) - остаточне положення та статус кожного індивіда.

Пізніше використовується така часова інформація для аналізу рівня соціальних порушень та зони високого ризику на місці події. Стан кожної людини в кадрі моделюється як:

$$\square = [\square, \square, \square, \square, \square', \square', \square']^T \qquad (2.10)$$

де *(u, v)* - горизонтальне та вертикальне положення цільової обмежувальної рамки (тобто центроїда); *s* - позначає масштаб (площа), а r - співвідношення сторін обмежувальної рамки; $\square', \square'$ *та* $\square'$- прогнозовані значення фільтра Кальмана для

горизонтального положення, вертикального положення та центроїда обмежувальної рамки відповідно.

Коли ідентифікована людина асоціюється з новим спостереженням, поточне обмежувальне поле буде оновлено до нового спостережуваного стану. Це буде розраховано на основі компонентів швидкості та прискорення, оцінених в рамках фільтру Кальмана. Якщо передбачувані ідентичності запитуваного суттєво відрізнятимуться від нового спостереження, буде використаний майже той самий стан, який передбачає фільтр Калмана, майже без корекції. В іншому випадку поправочні ваги будуть розподілені пропорційно між передбаченням фільтра Кальмана та новим спостереженням (вимірюванням).

Як згадувалося раніше, використовується угорський алгоритм для вирішення проблеми асоціації даних, обчислюючи IoU (рівняння (6)) та відстань (різницю) фактичних вхідних значень до прогнозованих значень за допомогою фільтра Кальмана.

Після процесу виявлення та відстеження для кожного вхідного кадру $I_{w\times h}$ в момент часу $t$ визначаємо матрицю $D_t$ яка включає розташування $n$ виявленої людини в сітці несучого зображення:

$$\square_{\square} = \{\square^{\square}_{(\square_{\square},\,\square_{\square})} \mid \square_{\square} \in \square,\ \square_{\square} \in \square\}$$

(2.11)

## 2.3. Оцінка між відстанями

Стерео-зір - це популярна техніка для оцінки відстані, така як у [29]; однак це не є здійсненним підходом у нашому дослідженні, коли націлені на інтеграцію ефективного рішення, застосовного у всіх громадських місцях з використанням лише базової CCTV камери відеоспостереження. Тому дотримуємося монокулярного рішення.

З іншого боку, за допомогою однієї камери проекція 3-D світової картини на 2-D перспективну площину зображення призводить до нереальних піксельних відстаней між об’єктами. Це називається ефектом перспективи, коли не можемо сприйняти рівномірний розподіл відстаней на всьому зображенні. Наприклад, паралельні лінії перетинаються на горизонті, і віддалені люди від камери здаються набагато коротшими, ніж ті, хто знаходиться ближче до центру координат камери.

У тривимірному просторі центр або опорна точка кожної обмежувальної рамки асоціюється з трьома параметрами *(x, y, z)*, тоді як на зображенні, отриманому від камери, початковий тривимірний простір зменшується до двовимірного *(x, y)*, а параметр глибини *(z)* недоступний. У такому зниженому розмірному просторі пряме використання критерію евклідової відстані для вимірювання оцінки відстані між людьми було б помилковим.

Для того, щоб застосувати калібрований перехід IPM, спочатку потрібно провести калібрування камери, встановивши $z = 0$, щоб усунути ефект перспективи. Також повинно знати розташування камери, її висоту, кут зору, а також специфікації оптики (тобто внутрішні параметри камери) [104].

Застосовуючи IMP, 2D-піксельні точки *(u, v)* будуть відображені у відповідні світові координатні точки ($X_w$, $Y_w$, $Z_w$):

$$[\square \ \square \ 1]^{\square} = \square\square\square[\square_{\square} \ \square_{\square} \ \square_{\square} \ 1]^{\square} \qquad (2.12)$$

де *R* - матриця обертання:

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta & 0 \\ 0 & \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (2.13)$$

*T* - матриця перекладу:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -\frac{h}{\sin\theta} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.14)$$

*K* і внутрішні параметри камери відображаються наступною матрицею:

$$K = \begin{bmatrix} f * ku & s & c_x & 0 \\ 0 & f * kv & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (2.15)$$

де *h* - висота камери, *f* - фокусна відстань, а *ku* та *kv* - виміряні коефіцієнти коефіцієнта калібрування у горизонтальних та вертикальних одиницях пікселів відповідно. ($c_x$, $c_y$) - це основний точковий зсув, який коригує оптичну вісь площини зображення.

Камера створює зображення з проекцією тривимірних точок у світовій координати, що падає на площину сітківки. За допомогою однорідних координат взаємозв'язок між тривимірними точками та результуючими точками зображення проекції можна показати наступним чином:

$$\begin{bmatrix} u \\ (v) \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} X_w \\ \begin{bmatrix} Y_w \\ Z_w \end{bmatrix} \\ 1 \end{bmatrix} \quad (2.16)$$

**де M** $\in$ $R^{3\times4}$ - матриця перетворення з елементами $m_{ij}$ у рівнянні (2.16), яка відображає точки світових координат у точки зображення на основі розташування камери та опорного кадру, наданого Власною Матрицею Камери K (Рівняння (2.15)), матриця обертання R (рівняння (2.13)) та матриця перекладу T (Рівняння (2.14)).

Враховуючи площину зображення камери, перпендикулярну до доступу $Z$ у світовій системі координат (тобто z = 0), розміри наведеного рівняння можна звести до наступного вигляду:

$$\begin{bmatrix} u \\ (v) \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \begin{bmatrix} X_w \\ (Y_w) \\ 1 \end{bmatrix} \tag{2.17}$$

і, нарешті, перехід з перспективного простору в інверсний перспективний простір (BEV) також може бути виражений у такій скалярній формі:

$$(u, v) = (\frac{m_{11} \times x_w + m_{12} \times y_w + m_{13}}{m_{31} \times x_w + m_{32} \times y_w + m_{33}}, \frac{m_{21} \times x_w + m_{22} \times y_w + m_{23}}{m_{31} \times x_w + m_{32} \times y_w + m_{33}}) \tag{2.18}$$

## 2.4 Висновки

У цьому розділі була розглянута триступенева модель, що може може бути інтегрована та застосована до всіх типів камер відеоспостереження з будь-якою роздільною здатністю від VGA до Full-HD, з продуктивністю в режимі реального часу.

Експериментуючи з різними раціональними конфігураціями для neck модуля моделі, використовується (Spatial Pyramid Pooling - SPP) і PAN, а також модуль просторової уваги (Spatial Attention Module - SAM), що разом зробило один з найбільш ефективних, послідовних і надійних компонентів для зосередження моделі на оптимізації параметрів.

# ВИСНОВОК

Підводячи підсумки виконаної роботи, можна сказати, що була запропонована модель детектора людини KeepDistance на основі нейронних мереж для виявлення та відстеження статичних та динамічних людей у громадських місцях з метою моніторингу метрик соціального дистанціювання в епоху COVID-19 та пізніше. Були оцінені та досліджені різні типи ультра сучасних backbones, necks та heads. Використані backbone CSPDarkNet53 разом із SPP/PAN та SAM neck, YOLO head, з функцією активації Mish. Та застосовано функцію повної втрати IoU та збільшення Mosaic data на наборах даних MS COCO та Google Open Image, щоб збагатити фазу навчання, що в підсумку привело до ефективного та точного детектора людини, застосовного в різних середовищах із використанням будь-якого типу камер відеоспостереження.

Запропонований метод був оцінений для набору даних Oxford Town Centre, включаючи 7530 кадрів та приблизно 150 000 виявлення людей та оцінку відстані. Система була здатна виконувати різноманітні завдання, включаючи оклюзію, варіації освітлення, відтінки та часткову видимість, і виявилася великим розвитком з точки зору точності (99,8%) та швидкості (24,1 fps) порівняно з трьома станами найсучасніші техніки. Система виконувалась у режимі реального часу за допомогою базової платформи GPU або багатоядерної/багатопотокової платформи центрального процесора 10-го покоління або вище. Також адаптували зворотне перспективне геометричне картографування та алгоритм відстеження SORT для нашого додатку для оцінки міжлюдських відстаней та для відстеження рухомих траєкторій людей, оцінки та аналізу ризику зараження на користь органів охорони здоров'я та урядів.

KeepDistance пропонує незалежний від точки зору алгоритм класифікації людини. Тому, незалежно від кута та положення камери, результат цього дослідження є прямим застосуванням для більш широкого кола дослідників, не тільки в галузі комп’ютерного зору, штучного інтелекту та охорони здоров’я, але і в інших галузях промисловості, включаючи виявлення пішоходів для систем допомоги водієві, автономні транспортні засоби, виявлення аномалій поведінки в громадських місцях та натовпі, системи охоронного спостереження, розпізнавання дій у спорті, торгових центрах, громадських місцях; і взагалі, будь-які програми, де необхідно використовувати виявлення людей.

# СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ


1. Australian Government Department of Health. Deputy chief medical officer COVID-19. Dep. Soc. distancing for coronavirus DOI: https://doi.org/10. 1136/bmj.m1845 (2020).
2. Shahidi, M. Zero-shot learning from autonomous vehicles to COVID-19 diagnosis: A review. 3, 1–27, DOI: 10.2139/ssrn.3624379 (2020).
3. Togaçar, M., Ergen, B. COVID-19 detection using deep learning models to exploit social mimic optimization chest X-ray images using color and stacking approaches. https://doi.org/10.1016/j. compbiomed.2020.103805 (2020).
4. A., Khan, Paul, M. Computer vision for COVID-19 control: A survey. Image Video Process. DOI: 10.31224/osf.io/yt9sx (2020).
5. Nguyen, T. T. Artificial intelligence in the battle against coronavirus (COVID-19): a survey and future research. DOI: 10.13140/RG.2.2.36491. (2020).
6. Choi, W. & Shim, E. Optimal strategies for vaccination and social distancing in a game-theoretic epidemiological model. J. Theor. Biol. 110422, DOI: https://doi.org/10. 1016/j.jtbi.2020.110422 (2020).
7. C, E., K, P. Systematic biases in disease forecasting–the role of behavior change. J. Epidemics 96—105, 10.1016/j.epidem.2019.02.004 (2019).
8. O, K. W. & G, M. A. A contributions to the mathematical theory of epidemics–i. The Royal Soc. publishing DOI: https://doi.org/10.1098/rspa.1927.0118 (1991).
9. Heffernan, J. M., Wahl, L. M. Perspectives on the basic reproductive ratio. J. Royal Soc. Interface 2, 281–293, DOI: 10.1098/rsif.2005.0042 (2005).
10. Gupta, R., Pandey, G., Chaudhary, P. Machine learning models for government to predict COVID-19 outbreak. DOI: 10.1145/3411761 (2020).
11. Ferguson, N. M. et al. Strategies for mitigating an influenza pandemic. Nature 442, 448–452, DOI: https: //doi.org/10.1038/nature04795 (2006).
12. Thu, T. P. B., Ngoc, P. N. H., Hai, N. M. et al. Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries. Sci. Total. Environ. 140430, DOI: https://doi.org/10.1016/j.scitotenv. 2020.140430 (2020).
13. Nguyen, C. T. et al. Enabling and emerging technologies for social distancing: A comprehensive survey. DOI: https://arxiv.org/abs/2005.02816 (2020).
14. C, R. T. Game theory of social distancing in response to an epidemic. PLoS computational biology 1–9, DOI: 10.1371/journal.pcbi.1000793 (2010).

15. Ainslie, K. E. et al. Evidence of initial success for china exiting COVID-19 social distancing policy after achieving containment, DOI: https: //doi.org/10.12688/wellcomeopenres.15843.1 (2020).
16. Vidal-Alaball, J. in the face of the COVID-19 pandemic. Atencion primaria 52, 418—422, DOI: https://doi.org/10.1016/j.aprim.2020.04.003 (2020).
17. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. & Chen, L. MobileNetV2: Inverted residuals and linear bottlenecks. CVPR, DOI: 10.1109/CVPR. 2018.00474 (2018).
18. Sonbhadra, S. K., Agarwal, P. Target specific mining of COVID-19 scholarly articles using one-class approach. J. Chaos, Solitons 140, DOI: 10.1016/j.chaos.2020.110155 (2020).
19. Punn, N. S. & Agarwal, S. Automated diagnosis of COVID-19 with limited posteroanterior chest X-ray images using fine-tuned deep neural networks. Image Video Process. DOI: arXiv:2004.11676 (2020).
20. Jhunjhunwala, A. Role of telecom network to manage COVID-19: Aarogya Setu. Transactions Indian Natl. Acad. Eng. 1–5, DOI: 10.1007/s41403-020-00109-7 (2020).
21. Robakowska, M. et al. The use of drones during mass events. Disaster Emerg. Medicine J. 2, 129–134, DOI: 10.5603/DEMJ.2017.0028 (2017)
22. Harvey, J., Adam. LaPlace. MegaPixels: Origins, ethics, and privacy implications of publicly available face recognition image datasets (2019).
23. Xin, T. et al. Freesense. Proc. ACM on Interactive, Mobile, Wearable Ubiquitous Technol. 2, 1 – 23, DOI: https://dl.acm.org/doi/10.1145/3264953 (2018).
24. Shi, F. et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19. IEEE reviews biomedical engineering DOI: 10.1109/RBME.2020.2987975 (2020).
25. Hossain, F. A., Lover, A. A., Corey, G. A., Reigh, N. G. & T, R. FluSense: a contactless syndromic surveillance platform for influenzalike illness in hospital waiting areas. In ACM Int. Joint Conference on Pervasive and Ubiquitous Computing, 1–28, DOI: https://doi.org/10.1145/3381014 (2020).
26. Polese, M. et al. Machine learning at the edge: A data-driven architecture with applications to 5G cellular networks. IEEE Transactions on Mob. Comput. DOI: 10.1109/TMC.2020.2999852 (2020).
27. Brighente, A., Formaggio, F., Di Nunzio, G. M. & Tomasin, S. Machine learning for In-Region location verification in wireless networks. IEEE J. on Sel. Areas Commun. 37, 2490–2502, DOI: 10.1109/JSAC.2019.2933970 (2019).
28. Liu, L. et al. Deep learning for generic object detection: A survey. Int. journal computer vision 128, 261–318, DOI: https://doi.org/10.1007/s11263-019-01247-4 (2020).
29. Rezaei, M., Sarshar, M. & Sanaatiyan, M. M. Toward next generation of driver assistance systems: A multimodal sensor-based platform. In 2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE), vol. 4, 62–67 (IEEE, 2010).

30. Sabzevari, R., Shahri, A., Fasih, A., Masoumzadeh, S. & Ghahroudi, M. R. Object detection and localization system based on neural networks for robo-pong. In 2008 5th International Symposium on Mechatronics and Its Applications, 1–6 (IEEE, 2008).
31. Nguyen, D. T., Li, W. & Ogunbona, P. O. Human detection from images and videos: A survey. Int. J. Pattern Recognit. 51, 148–175, DOI: https://doi.org/10.1016/j. patcog.2015.08.027 (2016).
32. Serpush, F. & Rezaei, M. Complex human action recognition in live videos using hybrid FR-DL method. arXiv preprint, Comput. Vis. Pattern Recognit. DOI: arXiv: 2007.02811 (2020).