

Міністерство освіти і науки України  
Львівський національний університет імені Івана Франка  
Факультет прикладної математики та інформатики  
Кафедра прикладної математики

## Курсова робота

### **Застосування 3D реконструкції для розвідання об'єктів**

Студент групи ПМПМ-22:  
**Слободянюк Дмитро Романович**,  
спеціальність 113-прикладна математика

Науковий керівник:  
кандидат фізико-математичних наук,  
доцент  
**Дяконюк Лілія Миколаївна**

Рецензент:

---

Львів–2023

# Зміст

<b>Вступ</b>	<b>3</b>
<b>1 Загальні відомості</b>	<b>4</b>
1.1. Шлях напрямку . . . . .	4
1.2. Приклади задач комп'ютероного зору . . . . .	5
<b>2 CNN</b>	<b>8</b>
2.1. Історія згорткової нейронної мережі . . . . .	8
2.2. Концепції згорткової нейронної мережі . . . . .	10
<b>3 YOLO</b>	<b>12</b>
3.1. Що таке YOLO . . . . .	12
3.2. Архітектура . . . . .	12
3.3. Розпізнання . . . . .	13
<b>4 Дані</b>	<b>16</b>
<b>5 Тренування та приклади</b>	<b>18</b>
5.1. Технічна база . . . . .	18
5.2. Приклади . . . . .	19
<b>Висновки</b>	<b>26</b>
<b>Список використаних джерел</b>	<b>27</b>

# Вступ

У сучасному світі, завдяки стрімкому розвитку технологій, виникають нові підходи до аналізу та обробки великих обсягів інформації. Однією з таких інноваційних технологій є 3D реконструкція, яка знайшла широке застосування у багатьох галузях, таких як військову справу, медицину, інженерію, архітектуру, а також у сфері візуального сприйняття систем штучного інтелекту.

Зокрема, одним із важливих напрямків застосування 3D реконструкції є розпізнавання об'єктів. Відтворення тривимірної моделі об'єкта дозволяє отримати більш повну та деталізовану інформацію про його форму, розміри та просторове розташування. Але для того щоб перенести об'єкт з 2D площини у 3D його спочатку треба вміти розпізнати, з'ясувати його характеристики: довжину, ширину, висоту та інші.

У даній курсовій роботі ми дослідимо як найкраще виявити об'єкт на зображенні, класифікувати його ознаки, розглянемо вже існуючі рішення і спробуємо запропонуємо власне.

# Розділ 1

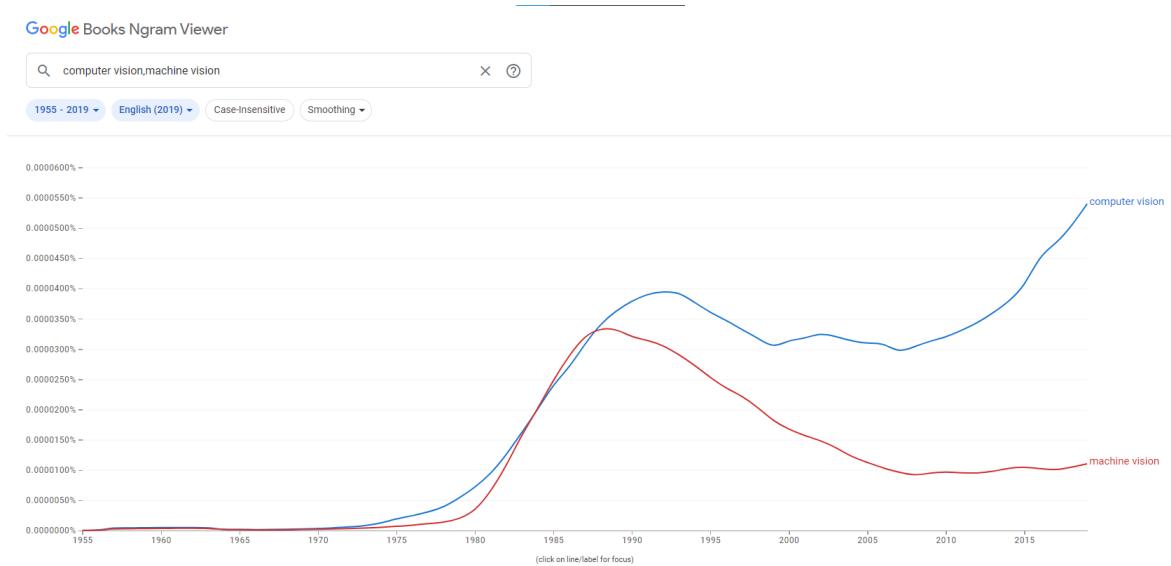
## Загальні відомості

У цьому розділі приведено загальні відомості про комп’ютерний зір, принципи його функціонування, основні методи та особливості.

### 1.1. Шлях напрямку

З кінця 1960-х починається розробка алгоритмів для аналізу та візуалізації даних, але ця технологія була обмежена обчислювальною потужністю. До 1970-х років дослідники розробили більш складні алгоритми для обробки зображень і розпізнавання образів. В 80-90х більший акцент вже було приділено розробці алгоритмів для комп’ютерного бачення. Ці алгоритми дозволили комп’ютерам навчатися на основі даних і підвищувати точність з часом. Проривним став алгоритм розпізнавання обличчя Віоли-Джонса, розроблений у 2001 році. У нульових та десятих роках 21 сторіччя алгоритми глибокого навчання зробили революцію в комп’ютерному зорі, дозволивши комп’ютерам вивчати ієрархічні представлення візуальних даних. Розробка згорткових нейронних мереж(Convolutive Neural Network(CNN)) та інших алгоритмів глибокого навчання дозволила комп’юетарм розпізнавати об’єкти, відстежувати рух і виконувати інші складні завдання з більшою точністю, ніж будь-кол раніше. Графік нижче ілюструє інтенсивність згадок

про комп'ютерний або машинний зір в англомовних джерелах.



## 1.2. Приклади задач комп'ютероного зору

У таблиці нижче подані приклади задач комп'ютерного зору

Галузь	Об'єкти	Модальності	Задачі	Джерела інформації
Робототехніка	Тривимірні зовнішні та внутрішні сцени Механічні частини	Світло X-променів	Індентифікація чи опис об'єктів на сцені Промислові задачі	Моделі об'єктів Моделі відбиття світла від об'єктів

Розділ 1. Постановка задачі

6

Галузь	Об'єкти	Модальність	Задачі	Джерела інформації
Аерозйомка	Будова Землі тощо	Світло Інфрачервоне випромінювання Радари	Покращення зображень Аналіз ресурсів Прогнозування погоди Шпіонаж Супровід запусків Тактичний аналіз	Карти Геометричні моделі форм Моделі формування зображень
Астрономія	Зорі Планети	Світло	Хімічний склад Покращення Зображень	Геометричні моделі форм
Медицина (макро)	Частини тіла	X-промені Ультразвук Ізотопи Тепло	Діагностика відхилень Планування операцій та лікування	Анатомічні моделі Моделі формування зображень
Медицина (мікро)	Клітини Ланцюги протеїнів Хромосовим	Електронна мікроскопія Світло	Патологія цитологія Каріотипування	Моделі форм
Хімія	Молекули	Електронна густина	Аналіз молекулярного складу	Хімічні моделі Структурні моделі

## Розділ 1. Постановка задачі

7

Галузь	Об'єкти	Модальність	Задачі	Джерела інформації
Нейроанатомія	Нейрони	Світло Електронна мікроскопія	Визначення просторової орієнтації	З'єднання нейронів
Фізика	Треки частинок	Світло	Знаходження нових частинок Ідентифікація треків	Атомна фізика

Отже, за більш ніж 60 років розвитку індустрії системи комп'ютерного зору проникли у значну кількість галузей, де змогли привнести чи малий прогрес та розвиток у вирішенні певних задач різного характеру та рівня складності.

# Розділ 2

## CNN

У цьому розділі приведено огляд структури CNN моделей на основі яких ґрунтуються більша частина моделей для задач коп'ютерного зору.

### 2.1. Історія згорткової нейронної мережі

В 1959р. , два нейрофізіологи Девід Хубель і Торстен Візель проводили досліди, а пізніше на їх основі опублікували статтю “Receptive fields of single neurons in cat’s striate cortex описуючи що нейрони всередині мозку кота організовані у формі шарів. Ці шари вчаться розпізнавати візуальні моделі, спочатку витягуючи локальні об’єкти, потім об’єднує витягнуті об’єкти для представлення більш високого рівня. З часом цей підхід став одним із основних принципів глибокого навчання.

Натхнений їх працею Куніхіко Фукусіма запропонував в 1980 році Neocognitron, який є самоорганізованою нейронною мережею з кількома рівнями, здатними розпізнавати ієрархічно візуальні шаблони через навчання і ця архітектура стала першою теоретичною моделлю CNN.

Значне вдосконалення архітектури Neocognitron було зроблено LeCun в 1989 році, розробивши сучасну структуру згорткової нейронної мережі під назвою LeNet-5, яка успішно розпізнала набір рукописних цифр MNIST.

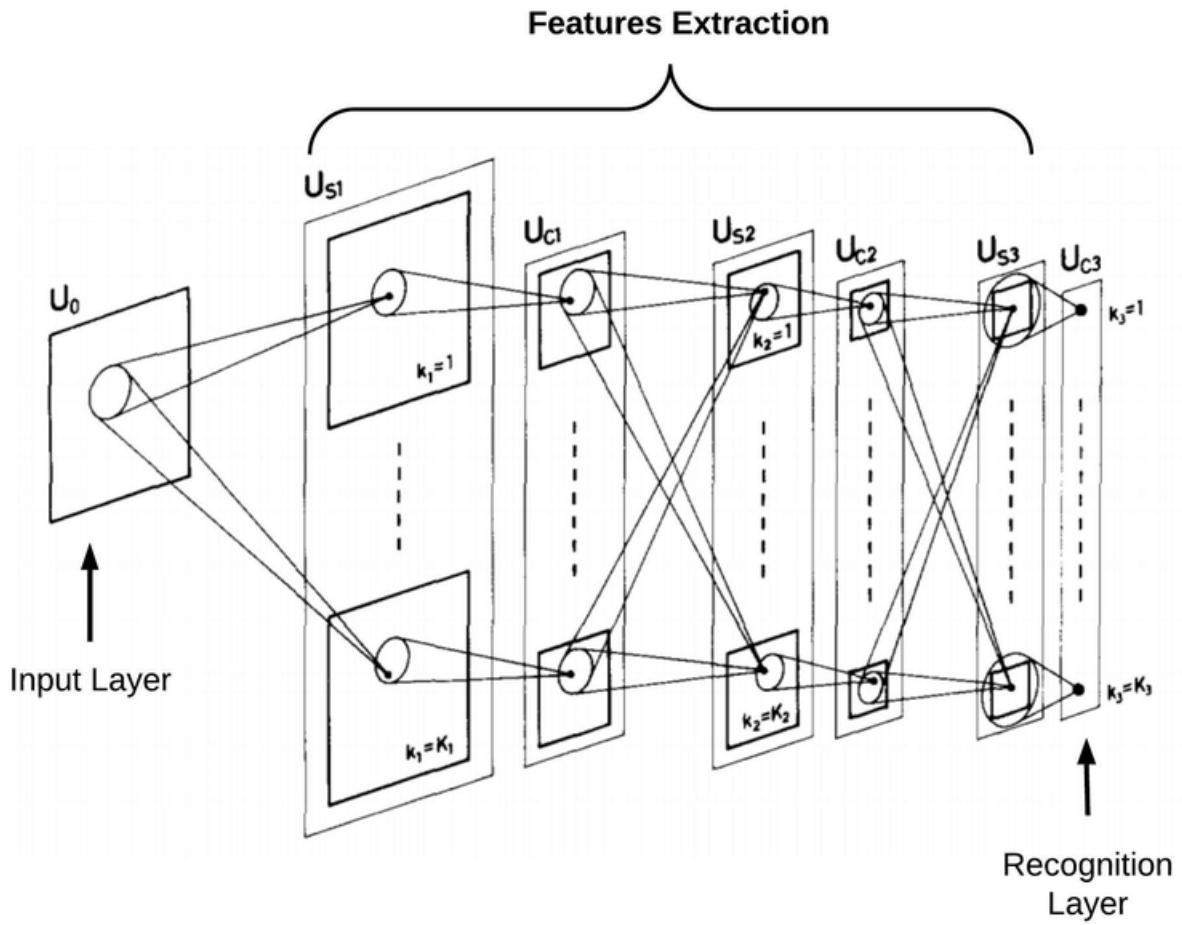


Рис. 2.1: Схематична діаграма, що ілюструє взаємозв'язки між шарами в неокогнітроні, Куніхіко Фукусіма

LeNet-5 був навчений за допомогою алгоритму зворотного поширення помилок, і він може розрізнавати візуальні шаблони безпосередньо з необроблених входних зображень, без використання будь-якого окремого механізму розробки функцій. Після відкриття LeNet-5 через певні обмеження, наприклад відсутність великих масивів навчальних даних, інновацій в алгоритмі та недостатню обчислювальну потужність, CNN не показав належних результатів у різних складних задачах. Але сьогодні, в час Big Data, нам доступні величезні означені масиви даних, більш сучасніший алгоритм та

потужніше залізо. З цими покращеннями, в 2012 році Олексій Крижевський розробив AlexNet, який досягнув високої точності на ImageNet Large Scale Visual Recognition Challenge. Перемога AlexNet проклала шлях до винаходу кількох моделей CNN, також до застосування цих моделей у різних сферах комп’ютерного зору та обробки природної мови.

## 2.2. Концепції згорткової нейронної мережі

Згорткова нейронна мережа(CNN), також завна ConvNet є типом штучної нейронної мережі(ANM), яка має глибоку архітектуру прямого зв’язку та здатність узагальнювати, як порівняно з іншими пережами з FC рівнями, вона може вивчати дуже абстрактні особливості та характеристики об’єктів, в тому числі просторові дані, і має здатність до їх індентифікації.

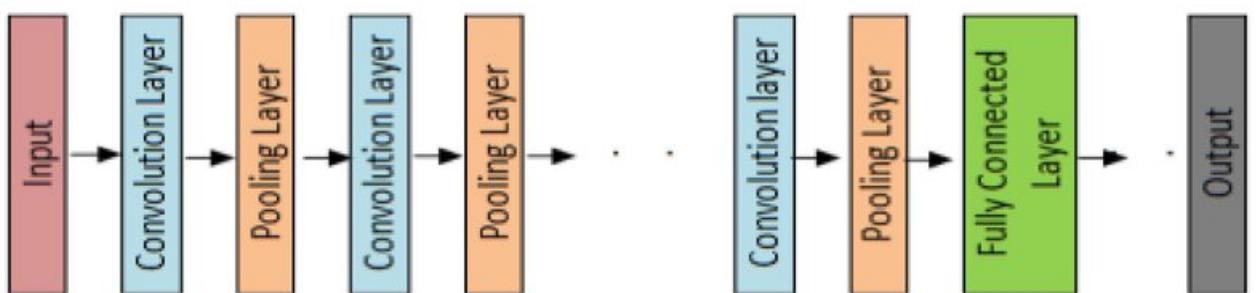


Рис. 2.2: Концептуальна модель CNN

Глибока CNN-модель складається з певного набору рівнів оброю, які можуть вивчати різні характеристики з вхідних даних(в нашому випадку зображень) з декількома рівнями абстракції. Початкові рівні вивчають і виділяють функції високого рівня(з нижчою абстракцією), а більш глибокі рівні

вивчають і виділють функції низького рівня (з вищою абстракцією). Базова концептуальна модель згорткової нейронної мережі зображена на малюнку вище.

Чому саме використовують CNN, а не інші нейронні мережі, коли говорять про задачі комп'ютерного зору?

- Одна з головних вибору згорткових в такому контексті є функція розподілу ваги CNN, яка зменшує кількість тренувальних параметрів в мережі, що допоможе моделі уникнути перенасичення і в цілому покращить узагальнення.
- В CNN рівень класифікації та рівні вилучення ознак навчаються разом, що робить вихід моделі більш упорядкованим і робить вихід більш залежним від вилучених ознак.
- Реалізація великої мережі є більш складною якщо використовувати інші типи нейронних мереж, ніж згорткові.

Сьогодні CNN використовують як механізм для досягнення перспективних результатів у різних додатках на основі комп'ютерого зору, таких як класифікація зображень, виявлення об'єктів, розпізнавання облич, розпізнавання мови, транспортних засобів, виразу обличчя, тексту тощо.

## Розділ 3

# YOLO

У цьому розділі приведено структуру YOLO.

### 3.1. Що таке YOLO

YOLO - означає "You Only Look Once", що є алгоритмом виявлення об'єктів у режимі реального часу, запропонованим Джозефом Редмондом. Він названий так тому, що система дивиться на зображення лише один раз і робить на його основі остаточні прогнози виявлення та розпізнавання. Щодо технічної складової, то базова версія працює зі швидкістю 45 кадрів в секунду, а полегшена - 150. YOLO переосмислює проблему виявлення як проблему регресії на відміну від більшості запропонованих раніше систем, які використовують класифікатори для виявлення, а також він має простий однопотоковий конвеєр, що робить його обчислювально ефективним.

### 3.2. Архітектура

- Змінює розмір вхідного зображення до 448x448 перед проходження через мережу

- Спочатку застосовується згортка  $1 \times 1$ , щоб зменшити кількість каналів, за якою потім слідує згортка  $3 \times 3$ , щоб створити прямокутний вихід.
- Прихована функція активації ReLU, за винятком останнього шару, який використовує лінійну функцію активації.
- Певні додаткові методи, такі як пакетна нормалізація та вилучення, упорядковують модель і запобігають її перенасиченню.

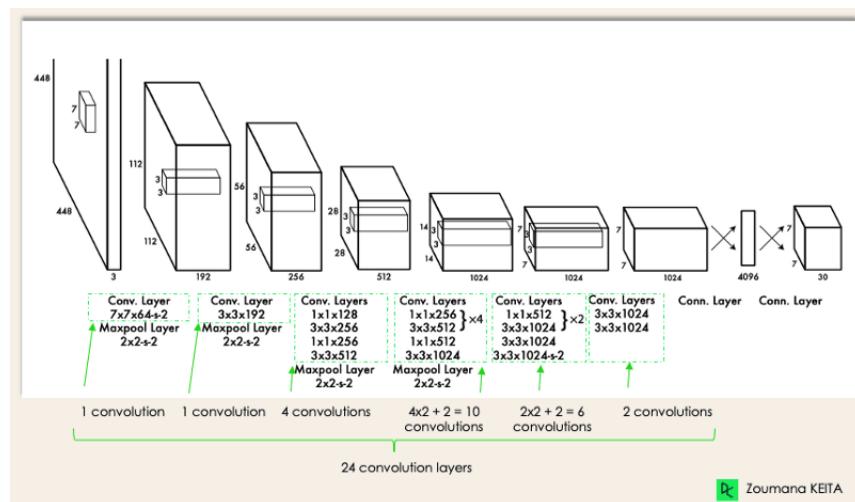


Рис. 3.1: Мережа складається з 24 згорткових рівнів, за якими слідують 2 повністю зв'язні шари.

### 3.3. Розпізнання

- Система YOLO розбиває вхідне зображення на сітку  $S * S$ . Якщо центр об'єкта лежить у цій клітинці сітки, то ця клітинка відповідає за виявлення цього об'єкта.
- Кожна комірка передбачає обмежувальні коробки В і пов'язані з нею оцінки достовірності. Коробки характеризуються: Висота, ширина відносно всього зображення, тобто між 0 і 1. Координати x і y відносно

меж комірки, що знаходяться між 0 і 1. Оцінка впевненості подається як  $\text{Pr}(\text{Object}) * \text{IOU}_{\text{Pre}}$ .

**Pr(Object)** це ймовірність того, що об'єкт присутній у комірці.

Intersection over Union(IoU) : Це метрика оцінки, яка використовується для перевірки точності передбачуваної обмежувальної коробки відносно фактичної правдивості.

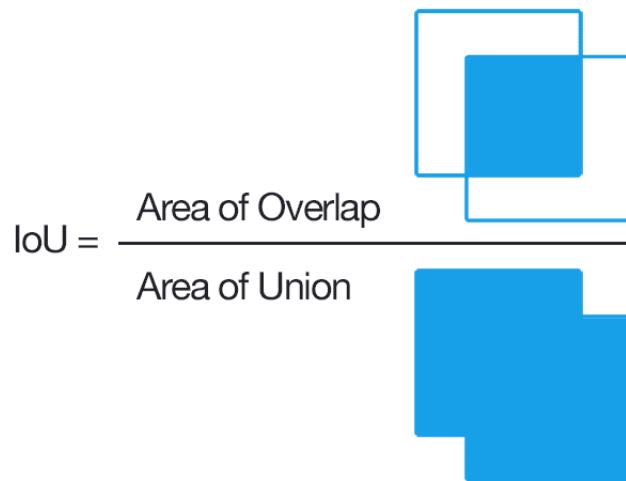


Рис. 3.2:

- Кожна комірка також передбачає лише набір умовних класових ймовірностей, які під час перевірки множаться на оцінки достовірності для створення прогнозів довіри за класами.

Confidence:  $\text{Pr}(\text{Object}) * \text{IOU}_{\text{Pre}} * \text{Pr}(\text{Class}_i/\text{Object}) = \text{Pr}(\text{Class}_i) * \text{IOU}_{\text{Pre}}$

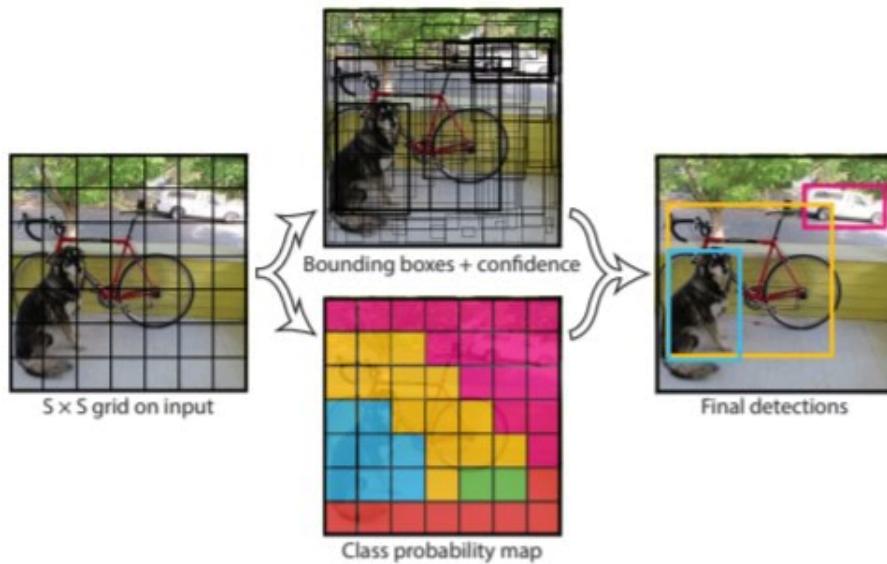


Рис. 3.3:

Вихідний тензор має розмір:  $S^*S^*(B^*5 + C)$ , де  $B$  — кількість обмежувальних коробок, 5 прогнозів на одну клітинку:  $x, y, h, w$ , показник достовірності та ймовірності класу  $C$  на клітинку.

У цій роботі використано Yolov8 - останню версію від Ultralytics. Будучи передовою, сучасною моделлю, YOLOV8 заснована на успіху попередніх версій, впроваджуючи нові функції та вдосконалення для підвищення продуктивності, гнучкості та ефективності. YOLOv8 підтримує повний спектр завдань у сфері комп’ютерного зору, включаючи виявлення, сегментацію, оцінку пози, відстеження та класифікацію.

## Розділ 4

### Дані

Для тренування моделі та демонстрації результатів використано датасет з роботи Амір Рошан Замір і Мубарак Ша [8].

Набір даних Google Street View містить 62 058 високоякісних зображень Google Street View. Зображення охоплюють центр міста та сусідні райони Піттсбурга, Пенсильванія; Орландо, Флорида та частково Манхеттен, Нью-Йорк. Також надаються точні GPS-координати зображень і їх напрямок за компасом. Для кожної позначки місця Перегляду вулиць (тобтоожної точки на одній вулиці) сферичний огляд  $360^\circ$  розбивається на 4 види збоку та 1 вид угору. На кожну позначку місця є одне додаткове зображення, яке показує деякі накладені маркери, наприклад адресу, назви вулиць тощо.



Рис. 4.1: Caption

На малюнку вище показано зразки зображень перегляду вулиць, що належать до восьми позначок місця в наборі даних ліворуч. Праворуч показано шістнадцять зразків завантажених користувачами зображень, які використовувалися як зображення запитів у відповідній статті.

Тренувальні/валідаційні та тестові зразки було випадково відібрано з датасету, згаданого вище. Наразі він складається з 600-а зображень і меншої кількості файлів анотації. Його було створено і упорядковано за допомогою сервісу Roboflow [9], а самі дані доступні для завантаження у будь-який момент.

## Розділ 5

# Тренування та приклади

### 5.1. Технічна база

Навчання моделі YoloV8 проводилось з використання наступних девайсів та технологій:

- Nvidia Geforce RTX 3060;
- AMD Ryzen 5 3600 6-Core Processor;
- Nvidia Cuda Toolkit;
- Nvidia cudnn;
- Anaconda Navigator;
- Python=3.9;
- Tensorflow-gpu==2.10.0.

## 5.2. Приклади

Нижче приведені приклади розпізнавання моделлю Yolov8 класу "building". Було проведено 50 епох тренування з кількістю батчів(batch(кількість зображень опрацьованих одночасно)) 2. Приклади обрані випадково з датасету.

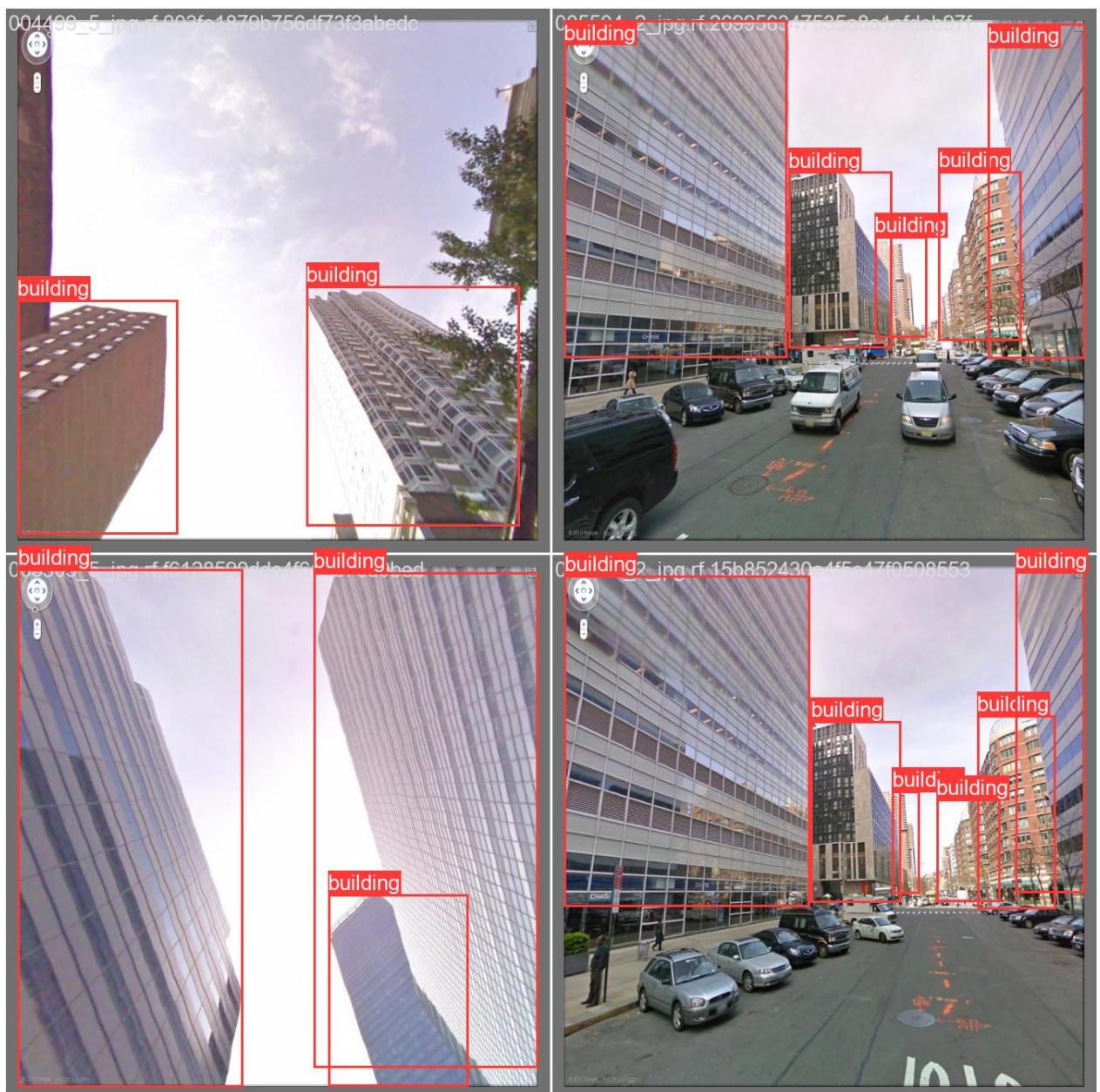


Рис. 5.1: Приклад 1: val1\_labels.jpg

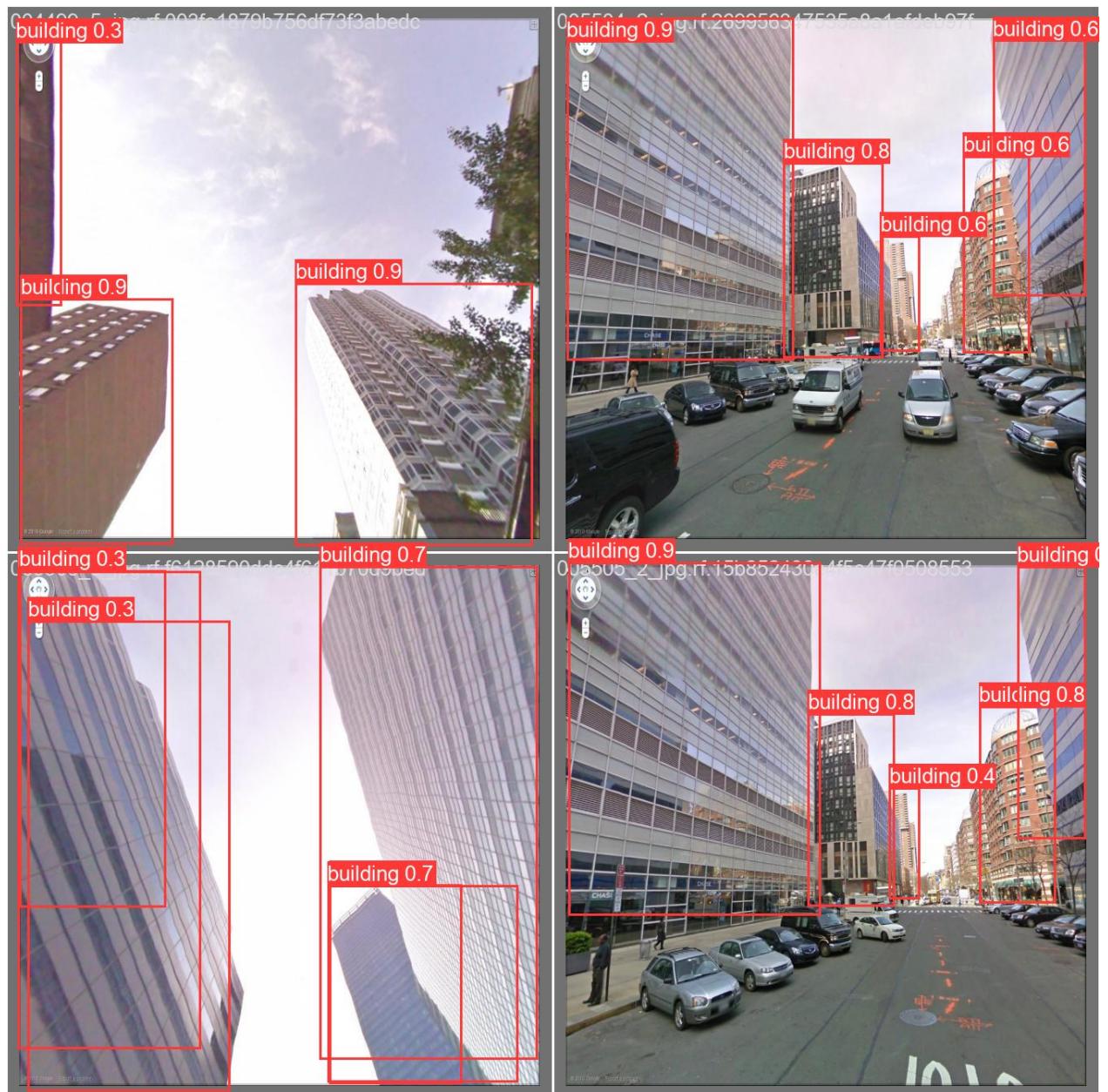


Рис. 5.2: Приклад 1: val1\_pred.jpg

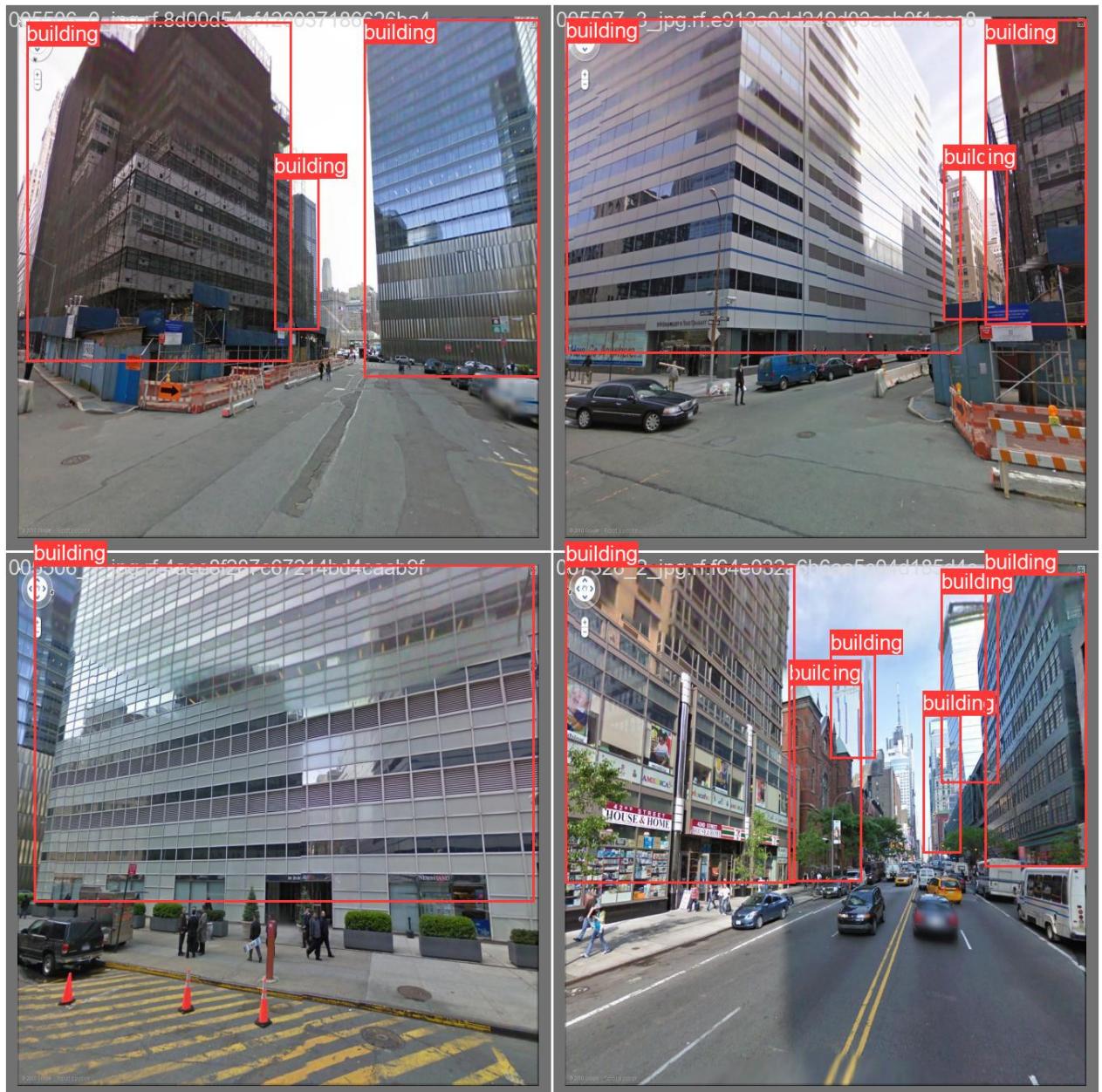


Рис. 5.3: Приклад 2: val2\_labels.jpg

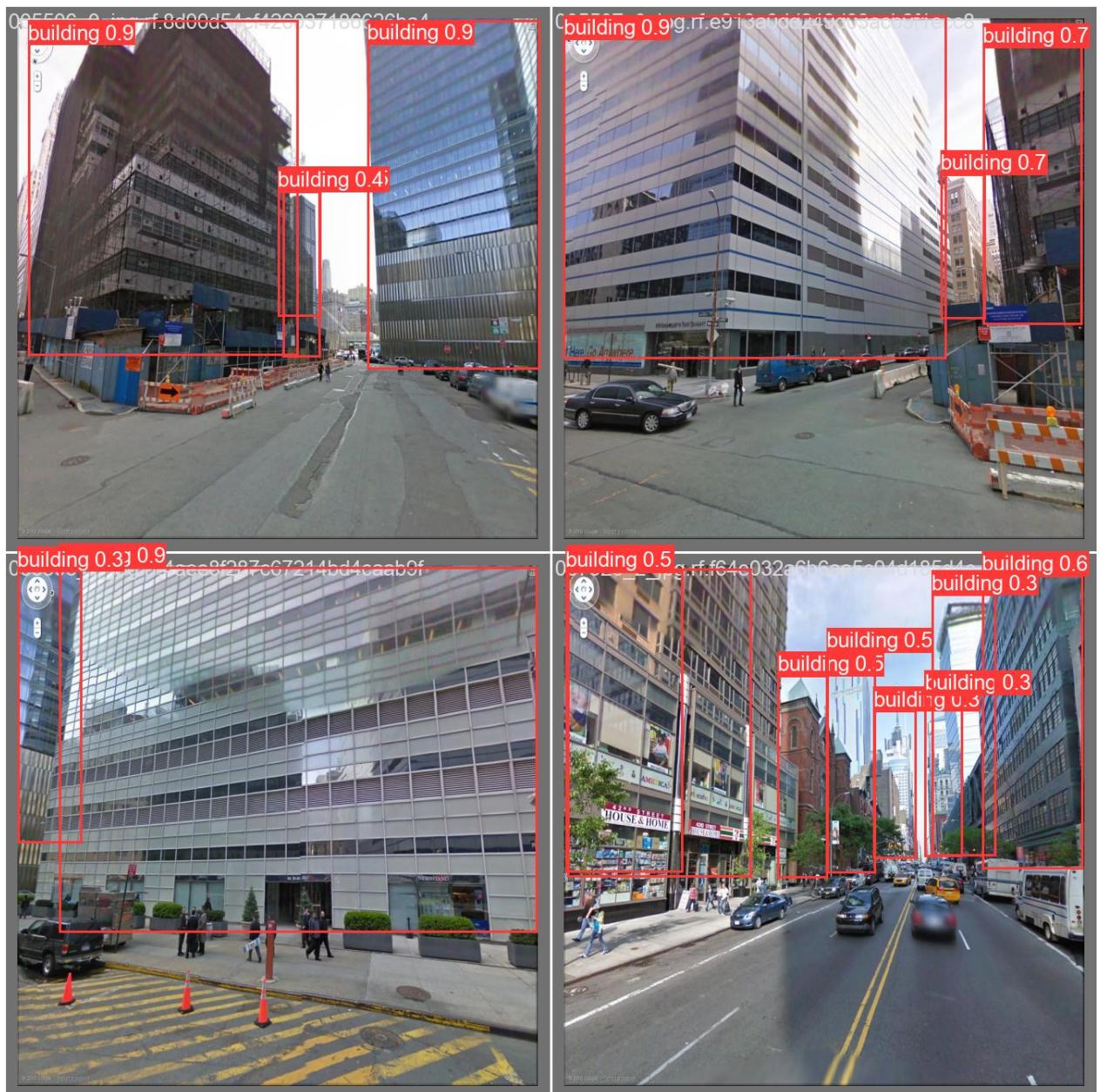


Рис. 5.4: Приклад 2: val2\_pred.jpg

Наступні приклади випадково вибрані з Інтернету, щоб продемонструвати ефективність моделі на сторонніх даних.

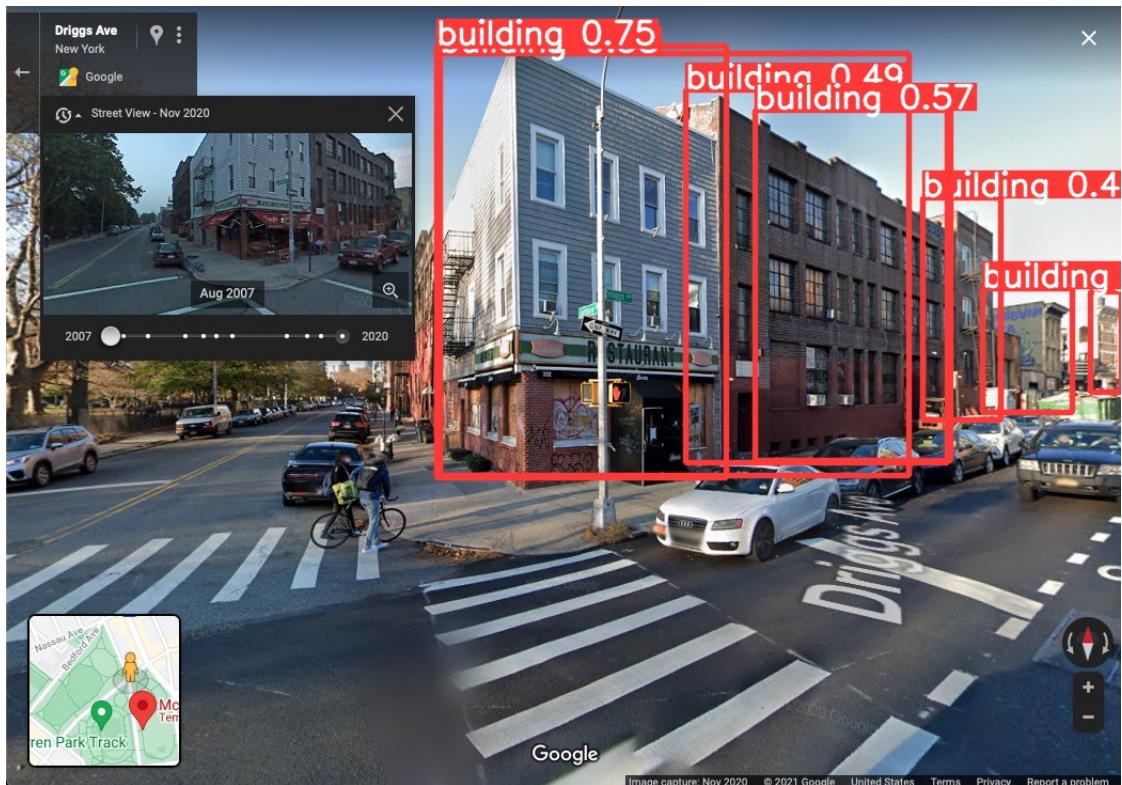


Рис. 5.5: Приклад 3



Рис. 5.6: Приклад 4



Рис. 5.7: Приклад 5

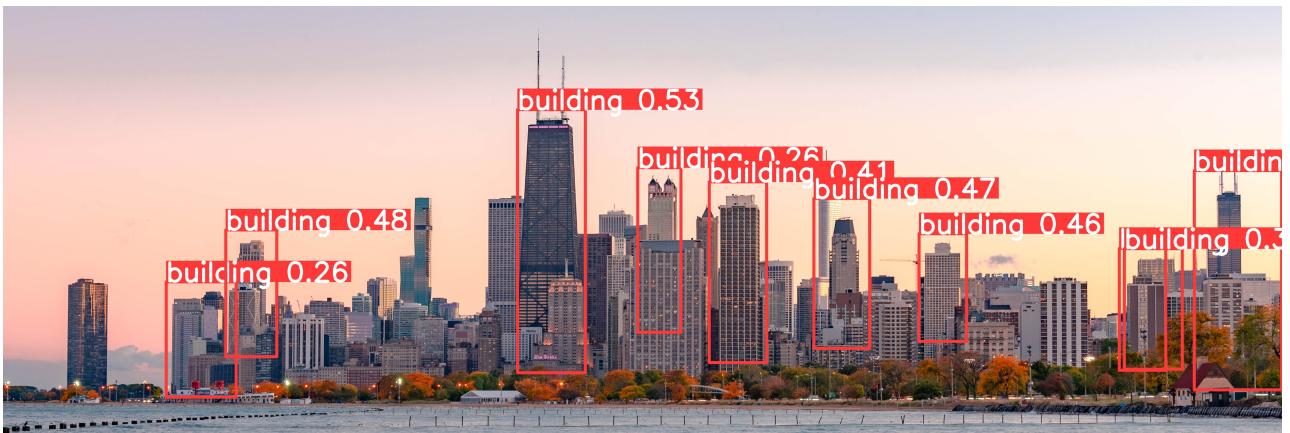


Рис. 5.8: Приклад 6

# Висновки

На основі отриманих результатів можна зазначити, що модель в середньому розпізнає наш об'єкт класу з ймовірністю близько 30-40% і це незважаючи на якість фото, більша частина з яких зроблена під гострими кутами, висвітлена чи ,навпаки, в тіні, або чимось загороджена.

В ході проробленої роботи, можна побачити що модель YoLov8 доволі легко навчити на адекватно підібраному датасеті. На прикладах видно, як, наприклад дерева заважають розпізнати будівлі за ними або авто чи світлофор помилково розпізнається як споруда. Вирішення цією проблему, я бачу, в масштабуванні навчання, тобто більше та якиснх даних на вхід.

В майбутньому планується покращення результатів в контексті даних через розширення датасету, збагачення його новими зображеннями, більш якісне лейбування. А також донавчання моделі на розпізнавання розмірів об'єктів, як відстані від об'єктиву зйомки та реконструкція у 3D з отриманих даних.

# Список використаних джерел

- [1] <https://www.sama.com/blog/computer-vision-history-how-it-works/>
- [2] Development of the computer vision system based on machine learning for educational purposes 14стр
- [3] Fundamental Concepts of Convolutional Neural Network 4-6
- [4] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex
- [5] Receptive fields of single neurons in cat's striate cortex
- [6] Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position
- [7] Joseph Redmon , Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection.
- [8] author=Zamir, A.R. and Shah, M., journal=Pattern Analysis and Machine Intelligence, IEEE Transactions on, title=Image Geo-localization Based on Multiple Nearest Neighbor Feature Matching using Generalized Graphs, year=2014, volume=PP, number=99, pages=1-1, keywords=Generalized Minimum Clique Problem (GMCP);Generalized Minimum Spanning Tree

(GMST);Geo-location;feature correspondence;feature matching;generalized graphs;image localization;multiple nearest neighbor feature matching,  
doi=10.1109/TPAMI.2014.2299799, ISSN=0162-8828,

- [9] <https://universe.roboflow.com/lviv-national-university-of-ivan-franko/buildings2>