Міністерство освіти і науки України

Національний університет «Києво-Могилянська академія»

Факультет інформатики

Кафедра математики

Магістерська робота

освітній ступінь – магістр

на тему: **«ТРЕНУВАННЯ МОДЕЛЕЙ ДЕТЕКЦІЇ ОБ'ЄКТІВ ПРИ НАЯВНОСТІ ШУМУ У РОЗМІТЦІ ДАНИХ»**

Виконала: студентка 2-го року навчання

освітньо-наукової програми «Системний аналіз»,

спеціальності 124 Системний аналіз

Судорженко Анна Володимирівна

Керівник: Швай Н. О.,

кандидат фіз.-мат. наук, доцент

Рецензент_____

(прізвище та ініціали)

Кваліфікаційна робота захищена

з оцінкою _____

Секретар ЕК _____

«____» _____ 20____ р.

Київ – 2022

РЕФЕРАТ

Обсяг роботи: 57 сторінок, 44 рисунки, 4 таблиці, 7 джерел посилань, 3 додатки.

ТРЕНУВАННЯ МОДЕЛЕЙ ДЕТЕКЦІЇ, ШУМ В НАВЧАЛЬНІЙ ВИБІРЦІ, ВПЛИВ ШУМУ НА ТРЕНУВАННЯ МОДЕЛІ, МОДЕЛЬ YOLOV5.

Об’єктом роботи є модель розпізнавання об’єктів на зображеннях, що натренована на навчальній виборці з вмістом шуму. Предметом цієї роботи є вплив різних типів шумів на якість розпізнавання об’єктів моделлю детекції.

Метою роботи є реалізацію алгоритму методу підвищення стійкості моделей детекції об’єктів до шуму у навчальній виборці та порівняння його ефективності на різних типах помилок, що можуть виникати.

Методи розробки: Навчання моделі детекції об’єктів архітектури YOLOv5 на тренувальних виборках, в яких наявний шум, порівняльний аналіз впливу шуму різних типів на якість розпізнаванн об’єктів, дослідження ефективності методів ансамблю моделей та TTA для підвищення стійкості моделі до шуму в навчальній виборці, програмування на мові Python.

Результат роботи: Визначено типи шуму, які виникають в процесі розмітки даних. Наведено графіки зміни метрик якості моделі детекції під впливом шуму різних типів. Показано покращення цих метрик після використання методів TTA та ансамблю моделей.

ЗМІСТ

ВСТУП

Галузь комп’ютерного зору швидко розвивається у сучасному світі, велика кількість спеціалістів займаються удосконаленням алгоритмів детекції об’єктів, щоби досягнути точності, наближеної до здатностей людського зору. Проте комплексність методів глибинного навчання спричиняє чутливість цих систем до шуму в навчальних виборках даних. Процес розмітки великих об’ємів даних для моделей детекції об’єктів є дуже кропітким та потребує задіяння значних ресурсів: зокрема, виконання роботи натренованими спеціалістами та оплати їхньої праці. Вирішенням цієї проблеми стали алгоритми автоматизованої розмітки. Однак у їхньому використанні слід зважати на одне суттєве застереження: сам функціональний характер цих алгоритмів спричиняє шум в даних та у кінцевому підсумку негативно впливає на точність моделі.

Процес аналізу даних, що передує тренуванню моделі як такому, є важливим етапом реалізації будь-якої моделі машинного навчання. Процес перевірки та очищення навчальних вибірок для моделей детекції застосовується тільки для формату зображення, тобто представлення обмежувальних рамок об’єктів, проте не зважає на коректність визначення цих рамок та класів об’єктів. На цьому етапі не існує способів автоматизованої перевірки якості розмітки зображень для моделей детекції, а ручна перевірка великих об’ємів даних є непродуктивною з огляду на співвідношення витрат людського ресурсу до потенційно очікуваної ефективності. Питання впливу шуму у навчальній вибірці досліджувалося лише в аспекті якості самого зображення, на якому необхідно розпізнати певні об’єкти. Тому виникає актуальна наукова проблема природи та ступеню впливу шуму у розмітці об’єктів на моделі детекції.

Об’єктом роботи є модель розпізнавання об’єктів на зображеннях, що натренована на навчальній виборці з вмістом шуму. Предметом цієї роботи є вплив різних типів шумів на якість розпізнавання об’єктів моделлю детекції.

За мету цієї роботи було поставлено реалізацію алгоритму методу підвищення стійкості моделей детекції об’єктів до шуму у навчальній виборці

та порівняння його ефективності на різних типах помилок, що можуть виникати.

Мета роботи зумовила наступне наукове завдання:

1. Дослідити вплив шуму в даних на метрики якості натренованої моделі розпізнавання об’єктів.
2. Зробити огляд відомих методів покращення якості моделей детекції об’єктів.
3. Застосувати розглянуті методи до тренування моделей за умови наявності шуму у навчальній вибірці.
4. Виконати порівняльний аналіз ефективності використаних методів для кожного виду шуму.

РОЗДІЛ 1.

1.1 Задача детекції об’єктів.

Задача детекції об’єктів – це задача машинного навчання, мета якої виявити наявність або відсутність об’єктів з множини заданих класів на зображенні або відео, визначення кордонів цього об’єкту у системі координат пікселів вихідного зображення або відео. Кордони об’єкту характеризуються координатами обмежувальної коробки, ключовими точками або його контурами. В цій роботі ми будемо розглядати алгоритми, для яких об’єкт характеризується координатами обмежувальної коробки (рамки) та знаходиться на зображенні.

Моделі детекції об’єктів включає в себе задачу знаходження координат обмежувальної коробки об’єкту та його класифікацію з множини заданих класів. Кількість об’єктів, які зображені на фото завчасно невідома. Системи виявлення об’єктів конструює модель для класу об’єкта за допомогою навчальної виборки. Об’єм навчальної вибірки залежить від варіацій позиціювання об’єктів з множини класів. Тобто, у випадку фіксованого стабільного об’єкту знадобиться мала кількість фото в датасеті, на яких зображений цей об’єкт, але в більшості випадків для тренування моделі детекції необхідна велика навчальна вибірка, щоб охопити певні аспекти мінливості класу. Однак можуть виникати складності при визначені моделі, яка відображає величезну різноманітність об’єктів, знайдених на зображеннях.

Зображення об'єктів певного класу мають високий рівень мінливості. Одним із джерел варіацій є фактичний процес зображення, тобто зміни в освітленні, положення камери, а також цифрові артефакти — все це призводить до значних змін у вигляді зображення, навіть у статичній сцені. Друге джерело варіацій пов'язано з різними варіантами зовнішнього вигляду об'єктів у межах класу, навіть якщо припустити відсутність змін у процесі зображення. Наприклад, люди мають різну фігуру, колір шкіри, носять різний одяг; рукописні цифри відрізняються за характером їх написання, шириною шрифта.

Завдання полягає в тому, щоб розробити інваріантні алгоритми виявлення відносно цих різновидів, які є обчислювально ефективними.

Підхід грубої сили до навчання моделей детекції передбачає велику кількість навчальних даних і представляє весь діапазон можливих варіацій об’єкта. Інваріантність неявно визначається з даних під час навчання моделей. В останні роки, зі збільшенням розміру анотованого набору даних і прискоренням обчислень за допомогою графічних процесорів, це був підхід вибору багаторівневої згорткової нейронної мережі. У випадках малої виборки даних для навчання важливим кроком є вбудова інваріантності в модель. Існує два методи досягнення цієї мети. Один включає обчислення інваріантних функцій і ознак, інший передбачає пошук прихованих змінних. Більшість алгоритмів використовують комбінацію цих підходів. Наприклад, багато алгоритмів застосувати локальні перетворення до інтенсивності пікселів таким чином, щоб перетворені значення були незмінними до умов освітлення та малих геометричних варіацій. Ці локальні перетворення призводять до ознак, а масив значень ознак є картою ознак. Більш значущі перетворення часто обробляються шляхом явного пошуку прихованих змінних або шляхом вивчення залишкової мінливості з навчальних даних.

Загальна структура для виявлення об'єктів за допомогою генеративних моделей включає моделювання двох розподілів. Розподіл $p(\theta;\ \eta_p)$ визначається за можливим параметром латентної пози $\theta \in \Theta$. Цей розподіл фіксує припущення щодо того, які пози є більш чи менш вірогідними. Модель зовнішнього вигляду визначає розподіл ознак зображення у коробці, що залежить від пози, $p(X_{s+W}|\ object, \theta;\ \eta_p\ )$, де X_{s+W} позначає елементи зображення у коробці (підзображення) з верхнім лівим кутом $s\ \epsilon\ L$ (L сітка місць на зображенні). Модель зовнішнього вигляду може бути означена за шаблоном, який визначає ймовірність спостереження певних ознак у кожному місці коробки за канонічним вибором пози об’єкта, тоді як θ визначає трансформацію зразку.

Тренувальний датасет з зображеннями об’єкта використовується для знаходження параметрів η_p та η_a. Зауважимо, що зображення зазвичай не містять інформацію про змінні прихованої пози θ, якщо не надано анотацію. Таким чином, для оцінювання потрібні методи припущення, які обробляють неспостережувані змінні, наприклад різні варіанти алгоритму максимізації очікування. У деяких випадках модель ймовірності для фонових зображень також оцінюється з використанням великої кількості навчальних прикладів зображень, які не містять об’єкта. Основний алгоритм виявлення сканує кожне вікно-кандидат на зображенні, обчислює найбільш вірогідну позу під моделлю об’єкта та отримує «постеріорні шанси», тобто співвідношення між умовною ймовірністю вікна за гіпотезою об’єкта в оптимальній позі, і умовна ймовірність вікна за гіпотезою фону. Потім цей коефіцієнт порівнюється з порогом τ, щоб визначити, чи містить вікно екземпляр об’єкта:

$$\frac{p\left(X_{s+W}|\ object, \theta;\ \eta_p\ \right) p(\theta; \eta_p)}{p(X_{s+W}|\ background\ )} > \tau$$

Якщо не використовуються явні змінні прихованої пози, основним припущенням є те, що навчальні дані достатньо багаті, щоб надати вибірку всієї варіації зовнішнього вигляду об’єкта. Дискримінаційний підхід навчає стандартний класифікатор розрізняти вікна зображень, що містять цільові класи об’єктів, і широкий фоновий клас. Для цього використовуються великі обсяги даних з класів об’єктів і фону. Через великий розмір популяції зображень фону та її складність, дискримінаційні методи часто організовують у каскади. Початковий класифікатор навчений розрізняти об'єкт і керовану кількість фону. Класифікатор розроблений таким чином, щоб не було помилкових хибних результатів за ціною більшої кількості хибнопозитивних результатів. Потім оцінюється велика кількість прикладів фону, а неправильно класифіковані збираються для формування нового набору даних фону. Після накопичення достатньої кількості таких хибних виявлень новий класифікатор навчається розрізняти вихідні дані об’єкта та нові «міцніші» фонові дані. Знову ж таки, цей

класифікатор розроблений таким чином, щоб не було помилково хибних результатів. Цей процес можна продовжувати кілька разів.

Класифікація в каскаді застосовується послідовно. Після того, як вікно класифікується як фон, його дослідження завершується міткою фону. Якщо вибрано мітку об’єкта, застосовується наступний класифікатор у каскаді. Тільки вікна, які класифікуються як об'єкти всіма класифікаторами в каскаді, позначаються як об'єкти.

Деякі дискримінаційні моделі також можуть бути реалізовані з параметрами латентної пози. Припустимо загальний класифікатор, визначений у термінах простору функцій класифікатора $f(x;u)$, параметризований u. Звичайне навчання дискримінаційної моделі складається з розв'язування рівняння виду

$$\min_u \sum_{i=1}^{n} D\big(y_i; f(x_i;u)\big) + C(u),$$

для деякої умови регуляції $C(u)$, який запобігає перенавчанню, і функції втрат D, що вимірює відстань між вихідним результатами класифікатора $f(x_i;u)$ і основною міткою істинності $y_i = 1$ для об’єкта і $y_i = 0$ для фону.

Мінімізацію, що визначена вище, можна замінити наступним чином

$$\min_u \sum_{i=1}^{n} \min_{\theta \epsilon \Theta} D\big(0; f(\theta(x_i);u)\big) + \min_u \sum_{i=1}^{n} \min_{\theta \epsilon \Theta} D\big(1; f(\theta(x_i);u)\big) + C(u)$$

Тут $\theta(x)$ визначає перетворення прикладу x. Інтуїтивно для вірного прикладу хотілося б, щоб було якесь перетворення, за яким x_i класифікується як об’єкт, тоді як для хибного прикладу хотілося б, щоб не було перетворення, за яким x_i класифікується як об’єкт.

1.2 Типи моделей виявлення об’єктів

Через обмеження низькорівневих локальних функцій і складність ручного вказування функцій вищого рівня, нейронні мережі стають все більш популярними як метод вивчення ефективних представлень функцій, від низькорівневих до високорівневих, з використанням великих анотованих

наборів даних. Ці мережі складаються з ієрархії шарів, індексованих сітками зі спадною роздільною здатністю. Перший шар — це необроблені пікселі вхідного зображення. Кожен шар обчислює векторний вихід у кожній точці сітки, використовуючи список локальних фільтрів, застосованих до даних на попередньому шарі. За цією лінійною операцією зазвичай слідує нелінійна операція, яка застосовується по координатам, а на певних шарах роздільна здатність сітки зменшується шляхом підвибірки після локального максимуму або операції усереднення. Мережа закінчується одним або кількома вихідними рівнями, які роблять прогнози відповідно до дизайну моделі (наприклад, класифікатор категорії об’єкта та оцінювач пози, якщо вихідні дані моделі містять інформацію).

Усі коефіцієнти фільтра та параметри вихідних шарів піддаються навчанню. Навчання здійснюється за допомогою стохастичного градієнтного спуску на функції втрат, визначеної в термінах вихідних шарів і міток у наборі даних. Таким чином мережа спільно вивчає лінійні класифікатори та складну ієрархію нелінійних ознак, які дають остаточне представлення ознак. Існує дві основні конструкції нейронних мереж виявлення об’єктів: однорівневі та дворівневі.

В обох цих конструкціях дві згорткові підмережі застосовуються паралельно. У кожному місці відносно сітки ознак одна з цих підмереж діє як класифікатор, а інша як передбачення знаходження. Метод оцінки знаходження прогнозує відносний зсув і масштабування обмежувальної коробки, тоді як класифікатор прогнозує, чи є це об’єкт або фон. Створюючи кілька пар таких шарів, причому кожна пара спеціалізується на коробці певного розміру та співвідношення сторін, набір попередньо визначених розсувних кордонів може краще наблизити набір усіх можливих обмежувальних коробок зображення. Завдання оцінювача розташування об’єкта - передбачити залишкову похибку між квантованими вікнами та істинними обмежуючими рамками об’єкта. У першій парадигмі проектування, яку часто називають «одноетапним» методом, класифікатор розсувного вікна робить прогнозування кількох класів для набору

всіх категорій об’єктів і фону. Ці прогнози разом із уточненими вікнами містять вихідні дані моделі.

У другій парадигмі проектування класифікатор розсувного вікна виконує класифікацію двох класів між об’єктом: (будь-якої категорії) і фоном. Удосконалені вікна потім використовуються як кандидати для розташування об’єктів, які часто називають регіонами інтересу (RoIs), для подальшої класифікації та етапу удосконалення вікна. Цей другий етап, як це є типовим для каскадної обробки, отримує лише високі результати RoI від класифікатора об’єкта та фону. Вхідні об’єкти для другого етапу, як правило, обчислюються шляхом вилучення об’єктів у кожному RoI з карти об’єктів. Процес виділення ознак RoI може включати квантування координат RoI та використання максимального об’єднання або білінійної інтерполяції карти ознак без виконання квантування. Вихідні значення другого етапу є багатокласове передбачення класифікації та удосконалення вікна (зсув і масштабування) для кожного RoI. Всі параметри цих мереж навчаються спільно за допомогою стохастичного градієнтного спуску на багатозадачній функції втрат, яка включає терміни для класифікації та оцінки пози.

1.3 Архітектура моделі YOLOv5

YOLOv5 – однорівнева модель детекції об’єктів. Її назва розшифровується наступним чином: «You only look once». Модель YOLO є першою моделлю детекції об’єктів, в якій процедура передбачення класу об’єкта з визначенням обмежувальних рамок об’єкта. Вона має три головні частини: хребет моделі, шия моделі та голова моделі.

Хребет моделі – це згорткова нейронна мережа, яка агрегує та формує елементи зображення з різною деталізацією, головним чином використовується для витягування важливих характеристик вхідної картинки. Тут використовуються перехресні часткові мережі (CSPNets) з метою екстракції інформативних ознак з зображення. CSPNets показали значні покращення часу обробки з глибшими мережами.

Шия моделі використовується для створення пірамід ознак, які допомагають узагальнити масштабування об’єкту. Тобто цей метод допомагає ідентифікувати один і той самий об’єкт в різних модифікаціях розміру та масштабу.

Голова моделі використовується для виконання фінальної частини детекції, застосовує прив’язані поля до об’єктів і генерує кінцеві вихідні вектори з ймовірностями класів, оцінками об’єктивності та обмежувальними рамками.

На рис. 1 зображена архітектура моделі YOLOv5

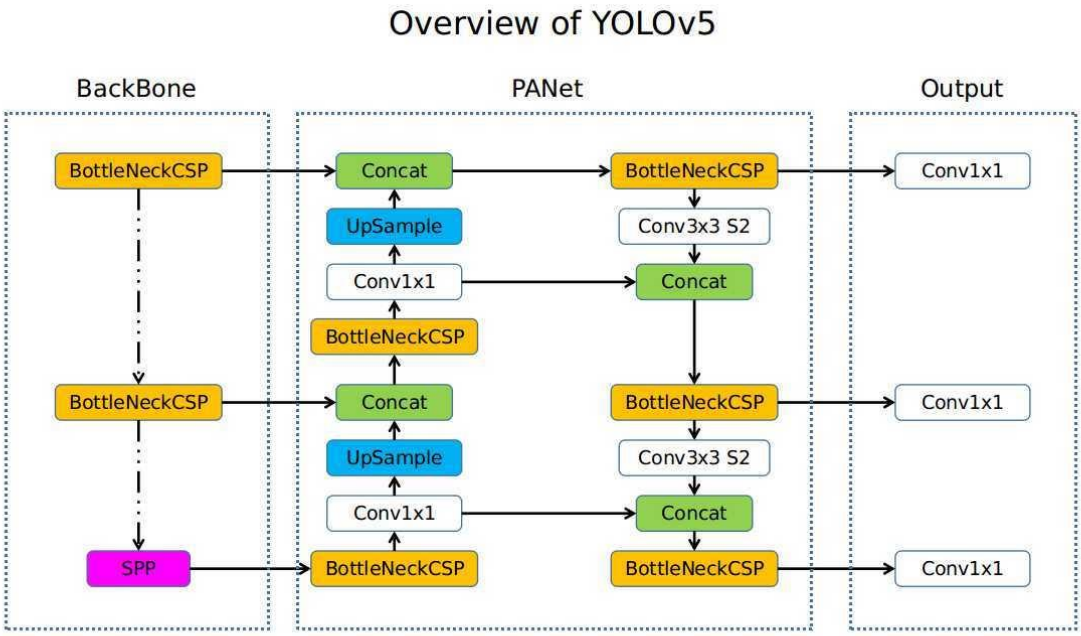


Рис. 1 Архітектура моделі YOLOv5

Особливість моделі YOLO полягає в тому, що мережа використовує ознаки всього зображення, щоб передбачити обмежувальну рамку кожного об’єкта. Це означає, що мережа глобально міркує про повне зображення та всі об’єкти, що на ньому зображені, а також прогнозує всі обмежувальні рамки для всіх об’єктів одночасно.

Конструкція YOLO забезпечує комплексне навчання та швидкість в реальному часі, зберігаючи високу середню точність. Система ділить вхідне зображення на сітку S × S. Якщо центр об’єкта потрапляє в клітинку сітки, ця

клітинка сітки відповідає за виявлення цього об’єкта. Кожна клітинка сітки передбачає B обмежувальних рамок та показники впевненості для цих квадратів. Ці показники відображають, наскільки модель переконана у тому, що коробка містить об’єкт, а також наскільки точно вона вважає, що прогнозує коробка. Формально впевненість визначається як $\Pr(\text{Object}) * \text{IOUTtruth pred}$. Якщо в цій комірці немає жодного об’єкта, показники впевненості мають дорівнювати нулю. В іншому випадку оцінка довіри дорівнює перетину об’єднання (IOU) між передбаченою коробкою та основною істиною. Кожна обмежувальна рамка складається з 5 передбачень: x, y, w, h і впевненість. Координати (x, y) представляють центр поля відносно меж осередку сітки. Ширина та висота (w, h) прогнозуються відносно всього зображення. Нарешті, прогноз являє собою IOU між прогнозованою коробкою і будь-якою базовою коробкою істини.

Кожна клітинка сітки також прогнозує C умовних ймовірностей класу, $\Pr(\text{Classi}|\text{Object})$. Ці ймовірності обумовлені коміркою сітки, що містить об’єкт. Алгоритм передбачає лише один набір ймовірностей класів на клітинку сітки, незалежно від кількості блоків B.

Під час тестування множаться умовні ймовірності класів і прогнози довіри окремої коробки,

$$\Pr(\text{Class}_i|\text{Object}) * \Pr(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}},$$

що дає специфічні для класу бали впевненості для кожної коробки. Ці оцінки кодують як ймовірність появи цього класу в коробці, так і те, наскільки добре передбачена коробка відповідає об’єкту.

1.4 Метрики, які використовуються для аналізу якості моделі.

Існують різні показники, які використовуються для вимірювання точності та продуктивності моделей детекції об’єктів. У цій частині ми розглянемо

головні з них – precision, recall, IOU та mAP. Розглянемо кожну з цих метрик більш детально.

Precision (влучність) вимірює, наскільки правильні ваші прогнози, тобто частка релевантних зразків серед знайдених.

Recall (повнота) вимірює частку загального числа позитивних зразків, які було знайдено.

Ці метрики мають наступне математичне представлення:

$$Precision = \frac{TP}{TP+FP};\quad Recall = \frac{TP}{TP+FN},$$

де $TP = True\ positive,\ TN = True\ negative, FP = False\ positive, FN = False\ negative$.

Тут TP визначає кількість правильно розпізнаних об'єктів, TN визначає кількість правильно вказує, що область не містить об'єкту, FP – визначає випадки, коли об'єкта немає на картинці, але модель його виявила і FN – визначає випадки, коли об'єкт є на картинці, але модель не виявила його.

Два показника, які ми описали, можна об'єднати в міру $F1\ score$. Оцінка $F1$ визначається як середнє гармонійне значення влучності та повноти. Формула розрахунку $F1\ score$:

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall},$$

Оскільки оцінка F1 є середнім показником влучності та влучності, це означає, що оцінка F1 надає однакову вагу влучності та влучності:

- Модель отримає високе значення оцінки $F1$, якщо і $Precision$, і $Recall$ високі
- Модель отримає низьке значення оцінки $F1$, якщо і $Precision$, і $Recall$ низькі
- Модель отримає середнє значення оцінки $F1$, якщо один із параметрів $Precision$ або $Recall$ низький, а інший високий.

Далі будемо розглядати метрику IoU (Intersection over union). IoU вимірює перекриття між двома межами. Ця метрика використовується для вимірювання площі перекриття передбаченої межі з істинною межею (межею реального

об’єкту). У деяких наборах даних попередньо визначається поріг IoU (наприклад, 0.5), щоб класифікувати чи є прогноз істинно позитивним чи хибно позитивним. IoU приймає значення від 0 до 1. На рис. 2 показано приклад визначення значення метрики IoU.

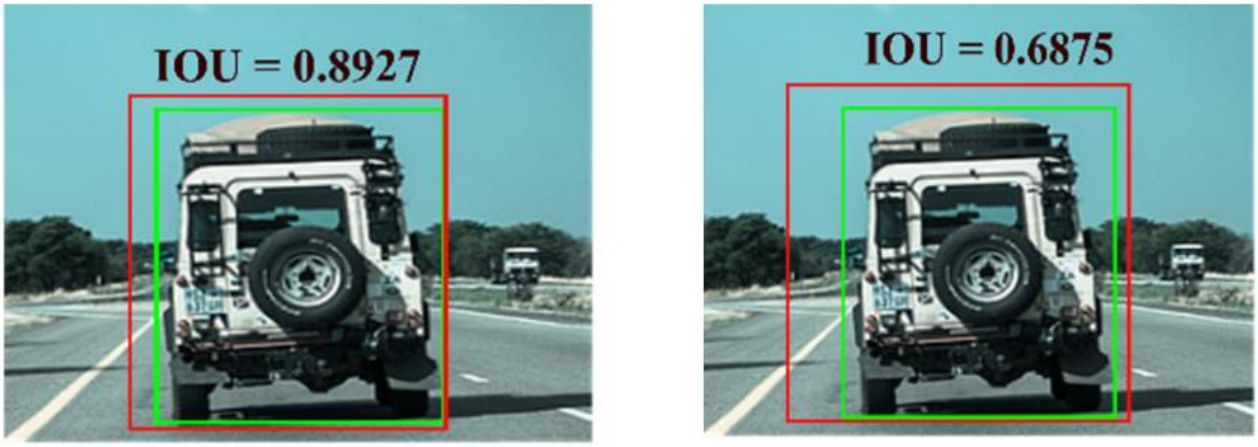


Рис. 2 Приклад знаходження метрики IoU

Після знаходження метрик $Precision$ та $Recall$, а також визначення порогу IoU, будується крива $Precision\ Recall\ curve$ ($PR\ curve$), за якою ми можемо визначити наступну метрику - $Average\ Precision$ (середня влучність).

Загальне визначення середньої точності (AP) — це знаходження площі під кривою $PR\ curve$. Вона визначається наступною формулою:

$$AP@\alpha = \int_0^2 p(r)dr,$$

де α – порогове значення IoU, p – влучність, r – повнота.

При розгляді багатокласового детектора об’єктів метрика mAP дає нам середню AP для всіх класів. Метрика mAP визначається наступною формулою:

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q},$$

де $AP(AP(q)$ – середня точність для класа q, Q – кількість класів.

В цьому розділі ми розглянули основні метрики для оцінки ефективності моделі, але треба зауважити, що жоден з цих показників не охоплює всієї картини, оскільки оцінка ефективності є суб’єктивним завданням.

Також розглянемо функцію втрат у моделях архітектури YOLOv5. Вона складається з трьох компонент: box loss, class loss, object loss.

Box loss – це втрата регресії обмежувальної коробки об’єкту. Обчислюється box loss наступним чином:

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} \left[(x_i - \overline{x}_i)^2 + \left(y_i - \overline{y}_i\right)^2 \right] +$$

$$\lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} \left[\left(\sqrt{w_i} - \sqrt{\overline{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\overline{h}_i}\right)^2 \right],$$

де ℓ_{ij}^{obj} означає, що j передбачена обмежувальна коробка в i-тій комірці відповідає за передбачення (передбачено об’єкт), λ_{coord} – параметр, що відповідає за збільшення втрат за передбачення обмежувальної коробки, x_i, y_i, w_i, h_i – це характеристики заданої обмежувальної коробки, $\overline{x}_i, \overline{y}_i, \overline{w}_i, \overline{h}_i$ – це характеристики обмежувальної коробки, яку передбачила модель.

Class loss – це класифікаційна втрата, яка визначається за допомогою перехресної ентропії та обчислюється наступним чином:

$$\sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{noobj} \left(C_i - \hat{C}_i\right)^2,$$

де λ_{noobj} – параметр, що відповідає за зменшення втрат від прогнозів довіри для блоків, які не містять об’єктів, C_i та $\hat{C}_i$ – умовна ймовірність класу об’єкту (заданого класу та передбачуваного класу), ℓ_{ij}^{noobj} означає, що j передбачена обмежувальна коробка в i-тій комірці відповідає за передбачення (передбачено фон зображення).

Object loss – це впевненість в присутності об’єкту, бо обчислюється за визначається бінарною перехресною ентропією та обчислюється наступним чином:

$$\sum_{i=0}^{S^2} \ell_i^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2,$$

де ℓ_i^{obj} означає, що в комірці i знаходиться об’єкт, $p_i(c)$ та $\hat{p}_i(c)$ – ймовірність того, що в комірці i знаходиться об’єкт класу c (задана та передбачена коробка).

РОЗДІЛ 2. ШУМ В ДАНИХ

2.1 Визначення даних з шумом

Дані з шумом – це дані з великою кількістю додаткової безцінної інформації або пошкодження даних, яка називається шумом. Він також включає будь-які дані, які система користувача не може зрозуміти та правильно інтерпретувати. Дані з шумом можуть негативно вплинути на результати тренування моделі і спотворити висновки.

Для задачі детекції об’єктів дані з шумом можемо інтерпретувати наступним чином: дані з шумом – це дані, в яких шум може знаходитись на зображенні (тобто, фото поганої якості, нечітко видно об’єкт, тощо) та/або в розмітці об’єктів (невірно визначений клас об’єкту, некоректно задана обмежувальна коробка для об’єкту тощо).

Далі ми будемо розглядати шум в даних, який спричинений некоректною розміткою даних. Ця проблема дуже часто зустрічається в реальному житті. Головна причина виникнення шуму в розмітці даних заключається в тому, що для тренування моделей детекції об’єктів необхідно використовувати великі набори даних. Розмітка великих наборів даних може здійснюватися двома методами:

1) Ручна розмітка даних – це процес розмітки зображення, який вручну роблять наймані співробітники (модератори)
2) Автоматична розмітка даних – це процес розмітки зображень, за який відповідають алгоритми машинного навчання.

Розглянемо плюси і мінуси кожного підходу. Ручна розмітка даних забезпечує високу якість розмічених екземплярів, бо розмітка об’єктів на зображенні є суб’єктивною задачею і в цьому випадку кожен модератор є особою, що приймає рішення. Як було визначено раніше, для моделей детекції об’єктів дуже важливо використовувати великий тренувальний датасет, тому важливо зауважити, що ручна розмітка даних для створення такого датасету буде впливати на ціну створюваної моделі. Також варто зазначити, що якість

роботи модераторів може падати через обробку великої кількості монотонних схожих між собою зображень.

Стосовно методу автоматичної розмітки даних, його плюси заключаються в тому, що процес створення датасету не вимагає багато часу та ціна його створення значно нижча від ручної розмітки зображень модераторами. Проте великий мінус цього методу заключається в високому рівню неточності визначення рамок або хибному визначенні класу об’єкту.

На рис 3 зображена схема автоматичної розмітки даних інструментом Amazon SageMaker Ground Truth.

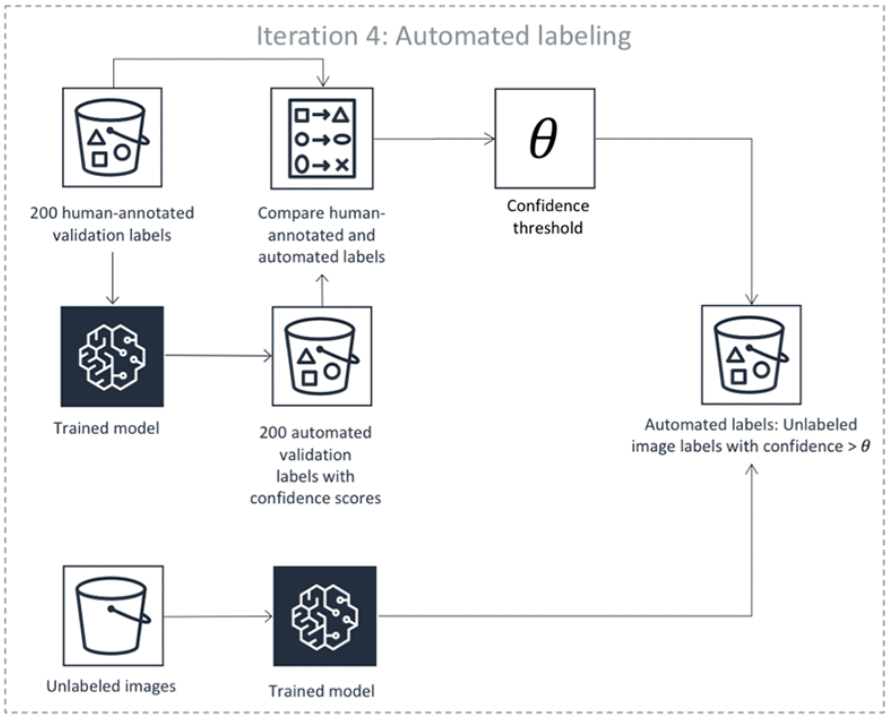


Рис. 3 Схема автоматичної розмітки Amazon SageMaker Ground Truth

2.2 Методи покращення якості моделі детекції, що навчена на тренувальній вибірці з шумом.

Після перших етапів навчання нейронних мереж на розмічених даних часто метрики якості моделі не відповідають очікуваним. У цій роботи для навчання моделей детекції об’єктів використовуються навчальні виборки з шумом, що в будь-якому випадку спричиняють погіршення якості

розпізнавання. В цьому розділі розглянуті два методи покращення основних метрик, що оцінюють якість моделі, без зміни даних в навчальній вибірці.

2.2.1 Метод ансамблю нейронних мереж.

Метод ансамблю моделей – процес комбінації передбачень, які є виробляються кількома моделями, з метою отримання кінцевого результату. Мета цього методу – це підвищення точності моделі передбачення. Далі розглядаємо використання методу ансамбля моделей для задачі виявлення об’єктів на зображеннях.

Для задачі детекції об’єктів можна виділити два види методів ансамблю моделей:

1) ті, які засновані на природі алгоритмів, що використовуються для побудови моделей детекції
2) ті, які працюють з передбаченнями моделей.

Перший вид ансамблю моделей запроваджує декілька видів модифікацій ансамблю моделей: об’єднання функцій з різних джерел перед подачею їх в алгоритм пропозиції області об’єкту, застосування ансамблю на етапі класифікації або застосування ансамблю на обох цих стадіях. Другий вид ансамблю моделей полягає в тому, щоб за допомогою додаткової моделі в ансамблю корегувати прогнози первинної моделі детекції об’єктів. Ще одна можлива модифікація ансамблю моделей, що працює з передбаченнями, полягає в застосуванні методів усунення зайвих коробок. Однак ці методи не враховують класи виявлених об'єктів або кількість моделей, які виявили конкретний об'єкт; і, отже, якщо їх застосовувати наосліп, вони, як правило, дають багато помилкових результатів.

У цій роботі буде розглянуто практичне застосування одної з модифікацій методу ансамблю моделей: буде об’єднано в ансамбль первинну модель, що навчена на вибірці даних train та вторинна модель, що навчена на вибірці test. Архітектури першої та другої моделі будуть однаковими.

2.2.2 Метод «test-time augmentation»

Нарощування даних (data augmentation) — це метод, який полягає у створенні нових навчальних зразків із вихідного набору навчальних даних шляхом застосування перетворень, які не змінюють клас даних (перевернення зображення, наближення, зміна кута, відзеркалення, масштабування, обрізання тощо). Test-time augmentation – вид нарощування даних для тестової вибірки. Цей підхід створює випадкові модифікації зображень з тестової вибірки, на якому модель передбачає характеристики об’єктів (їх клас та координати обмежувальної коробки цього об’єкту), і вихідним результатом стає ансамбль з цих передбачень. Така стратегія збільшення даних широко використовується як метод покращення якості моделей детекції через високі витрати на збір даних для навчання моделей цього типу.

Навпаки, через відсутність загального методу поєднання прогнозів детекторів об’єктів, збільшення часу тестування в основному застосовувалося в контексті класифікації зображень. У моделях детекції моделей метод test-time augmentation застосовується тільки для зміни кольорів зображення. Інші методи data augmentation не застосовуються для моделей розпізнавання об’єктів через те, що вони змінюють позицію об’єкта, що зображено, і це треба враховувати на етапі об’єднання результатів в вихідний.

РОЗДІЛ 3. АНАЛІЗ ВПЛИВУ ДАНИХ З ШУМОМ НА ТРЕНУВАННЯ МОДЕЛІ ДЕТЕКЦІЇ ОБ'ЄКТІВ

3.1 Постановка задачі

У цій роботі було проведено дослідження впливу шуму в даних на тренування моделі детекції об'єктів. Було використано стандартизований датасет PASCAL VOC 2007 року – це стандартизований набір зображень, на яких об'єкти розмічені шляхом виділення обмежувальної коробки (bounding box). У цій задачі ми вважаємо, що розглянутий набір даних немає шуму.

Мета цього аналізу заключається в тому, щоб шляхом редагування набору даних (додавання шуму в дані), виявити вплив шуму на якість моделі розпізнавання об'єктів.

Дослідження проводиться на наступних типах шуму в даних:

- Об'єкт зображений на фото, але обмежувальної коробки немає.
- Об'єкт зображений на фото, координати обмежувальної коробки коректні, але його категорія неправильна
- Об'єкт зображений на фото, його категорія обрана вірно, проте координати обмежувальної коробки не відповідають дійсним розмірам об'єкта (обмежувальна коробка може бути меншою або більшою за об'єкт)

Для кожного типу шуму буде проведено три експерименти:

1) Шум в наборі даних складає 10% від усіх розмічених об'єктів
2) Шум в наборі даних складає 30% від усіх розмічених об'єктів
3) Шум в наборі даних складає 50% від усіх розмічених об'єктів.

В усіх експериментах буде проведено навчання моделей, які мають архітектуру YOLOv5 та попередньо підготовлені на наборі даних COCO.

Для дослідження якості натренованої моделі детекції об'єктів буде розглянуто наступні метрики: train/validation box loss, train/validation class loss, train/validation object loss, precision, recall, mAP 0.5, mAP 0.5:0.95, F1 curve.

3.2 Дослідження набору даних, який буде використаний для тренування моделей детекції об’єктів.

Для дослідження я обрала набір даних PASCAL VOC 2007 року. Цей датасет включає в себе фотографії, на яких розмічено об’єкти 20 категорій:

- Люди: «person»
- Тварини: «horse», «bird», «cow», «dog», «cat», «sheep»
- Транспортні засоби: «aeroplane», «car», «bicycle», «boat», «train», «motorbike», «bus»
- Предмети, що знаходяться в кімнаті: «diningtable», «chair», «tvmonitor», «bottle», «pottedplant», «sofa».

В цьому датасеті знаходиться 5011 фотографій і розмічено 15662 об’єкта. Далі ми детально подивимось на розподіл розмічених об’єктів в наборі даних.

Розбиття даних на вибірки представлений у таблиці 1.

назва набору даних	відсоткова частка від загального датасету	кількість елементів
train	80%	4009
validation	10%	501
test	10%	501

Таблиця 1. Розбиття набору даних на вибірки

На малюнках 4, 5 та 6 зображені діаграми розподілу розмічених об’єктах на фото, що знаходяться у вибірках train, validation та test відповідно. Як можна побачити з цих діаграм, кількість об’єктів категорій «person», «car» та «chair» сильно відрізняється від середнього значення кількості об’єктів категорій інших класів. Проте об’єкти категорій «person», «car» та «chair» мають набагато вищий рівень варіативності в порівнянні з іншими класами, тому

незбалансованість кількості об’єктів цих категорій в порівнянні з іншими не призведе до зміщення детекції об’єктів до цих класів.

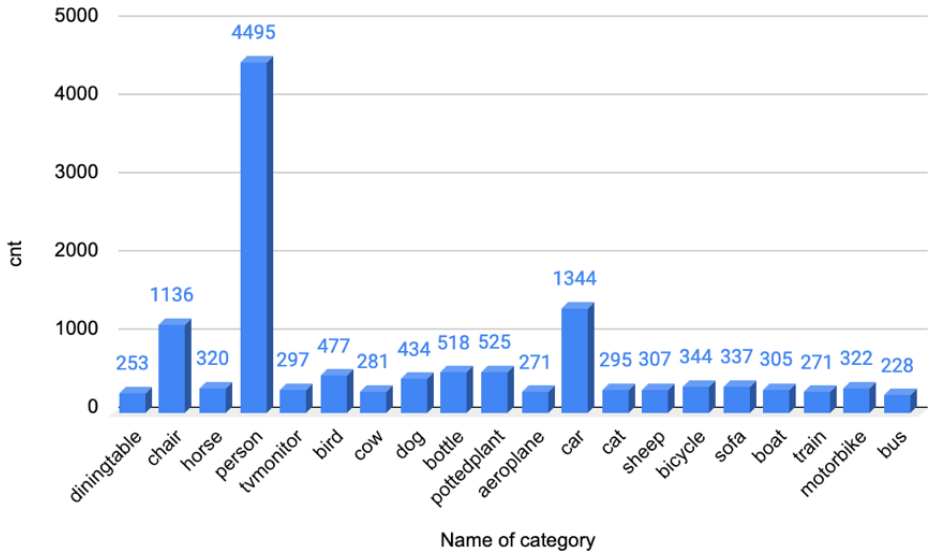


Рис. 4. Діаграма співвідношення кількості розмічених об’єктів до їх класів у вибірці train

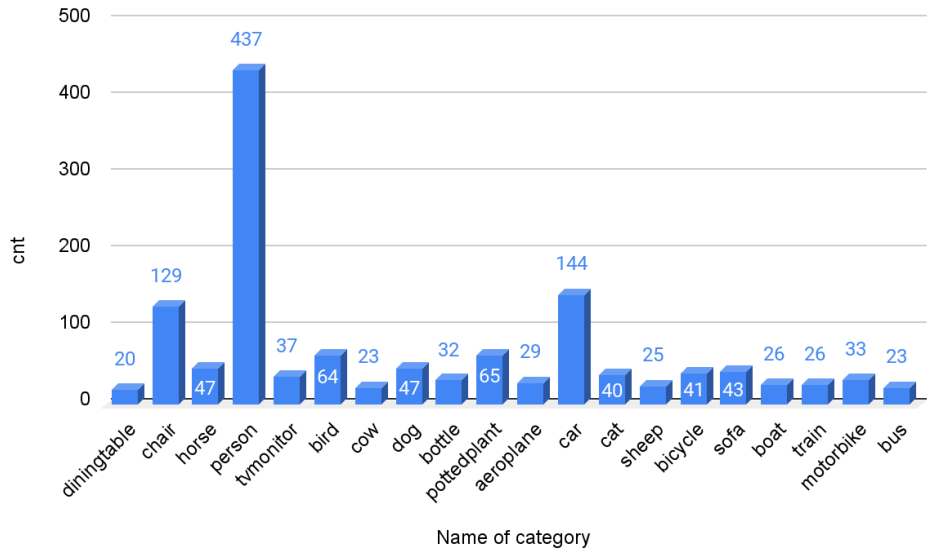


Рис. 5. Діаграма співвідношення кількості розмічених об’єктів до їх класів у вибірці validation

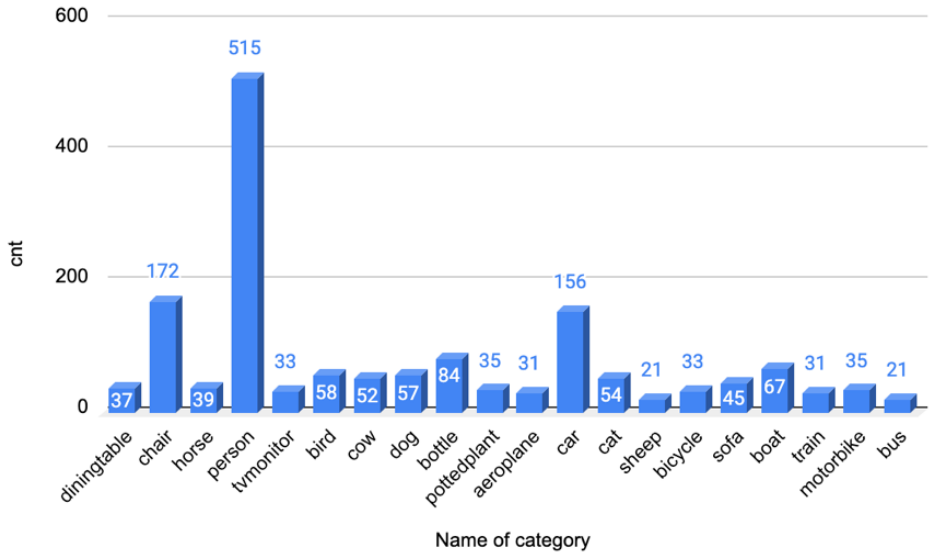


Рис. 5. Діаграма співвідношення кількості розмічених об’єктів до їх класів у вибірці test

У датасеті PASCAL VOC обмежувальні коробки об’єктів представлені у вигляді $(x_{min},\ y_{min},\ x_{max},\ y_{max})$, де x_{min} та y_{min} – координати верхньої лівої вершини коробки, в x_{max} та y_{max} – координати нижньої правої вершини коробки. Обрана у цьому дослідженні модель YOLOv5 приймає вхідні значення обмежувальної коробки об’єкту наступного вигляду: (x, y, w, h), де x та y – координати центру обмежувальної коробки, а w та h - ширина та висота обмежувальної коробки відповідно. У процесі підготовки даних до навчання файли, в яких вказано перелік об’єктів та координати їх обмежувальних коробок, були змінені відповідним чином.

Редагування даних та додавання в них шуму буде здійснюватись тільки для вибірки train з метою дослідження впливу помилок в розмітці саме на процес тренування моделі. Вибірки validation та test залишаються незмінними.

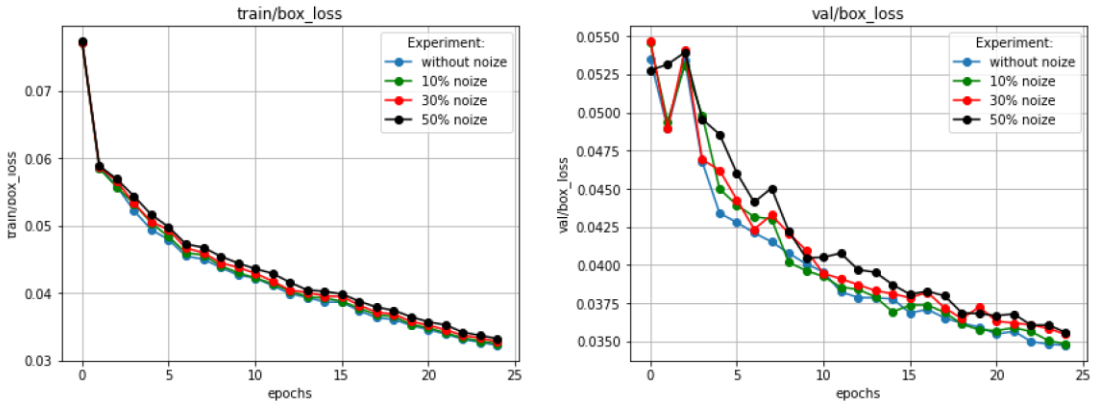
3.3 Дослідження впливу відсутності обмежувальної коробки на об’єкті, який зображений на фото

Як було зазначено раніше, для дослідження впливу шуму кожного типу було проведено три експеримента - навчання моделі детекції об’єктів на тренувальній вибірці, в якій вміст шуму становить: а) 10%; б) 30%; в) 50%.

Об’єкти, що піддавалися відповідним змінам, було обрано випадковим чином, кількість яких дорівнює $n\%$ від кількості всіх розмічених об’єктів у тренувальній виборці, де $n = \{10, 30, 50\}$. У цих експериментах додавання шуму відбувалось шляхом видалення обмежувальних коробок вибраних об’єктів.

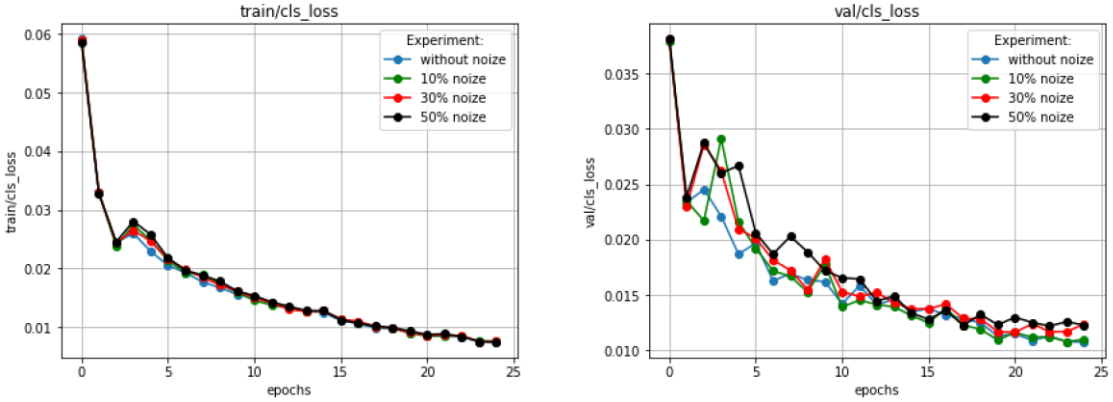
На малюнках 6-10 відображені графіки зміни значень метрик впродовж тренування моделі з шумом для експериментів: 1) вміст шуму – 0%; 2) вміст шуму – 10%; 3) вміст шуму – 30%; 4) вміст шуму – 50%.

На малюнку 11 зображено F1-криву для експериментів: 1) вміст шуму – 0%; 2) вміст шуму – 10%; 3) вміст шуму – 30%; 4) вміст шуму – 50%.



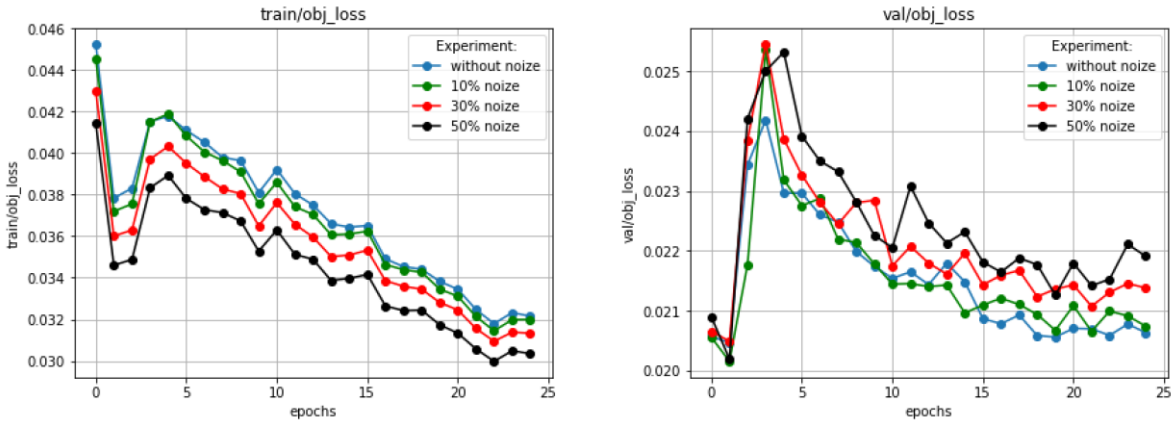
а) б)

Рис. 6. Графік зміни значення box loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



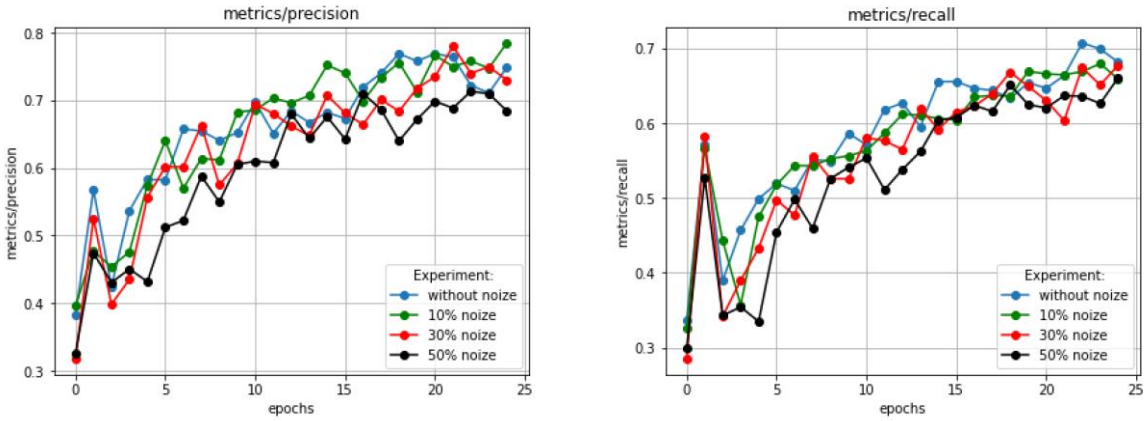
а) б)

Рис. 7. Графік зміни значення class loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



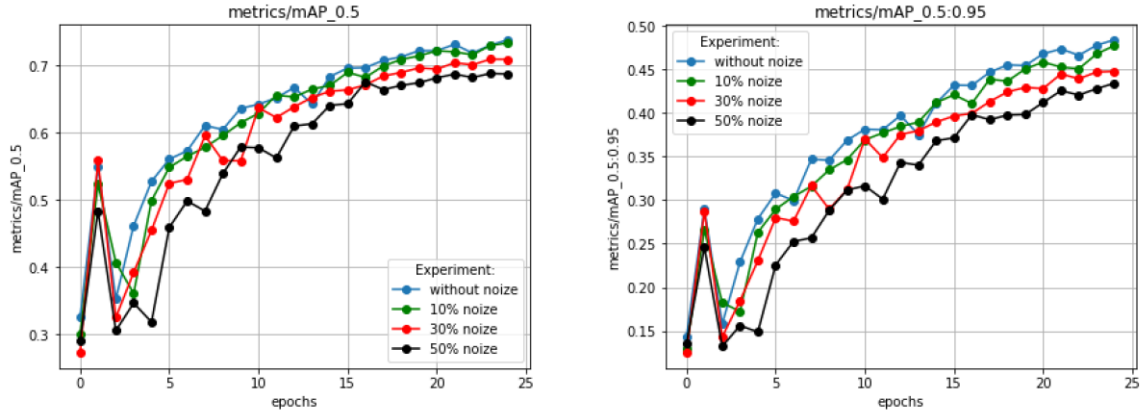
а) б)

Рис. 8. Графік зміни значення object loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



а) б)

Рис. 9. Графік зміни значення метрик для кожної епохи навчання моделі: а) влучність (precision); б) повнота (recall).



а) б)

Рис. 10. Графік зміни значення метрики mAP для кожної епохи навчання моделі: а) поріг $IoU = 0{,}5$; б) поріг $IoU \in (0{,}5;\ 0{,}95]$.

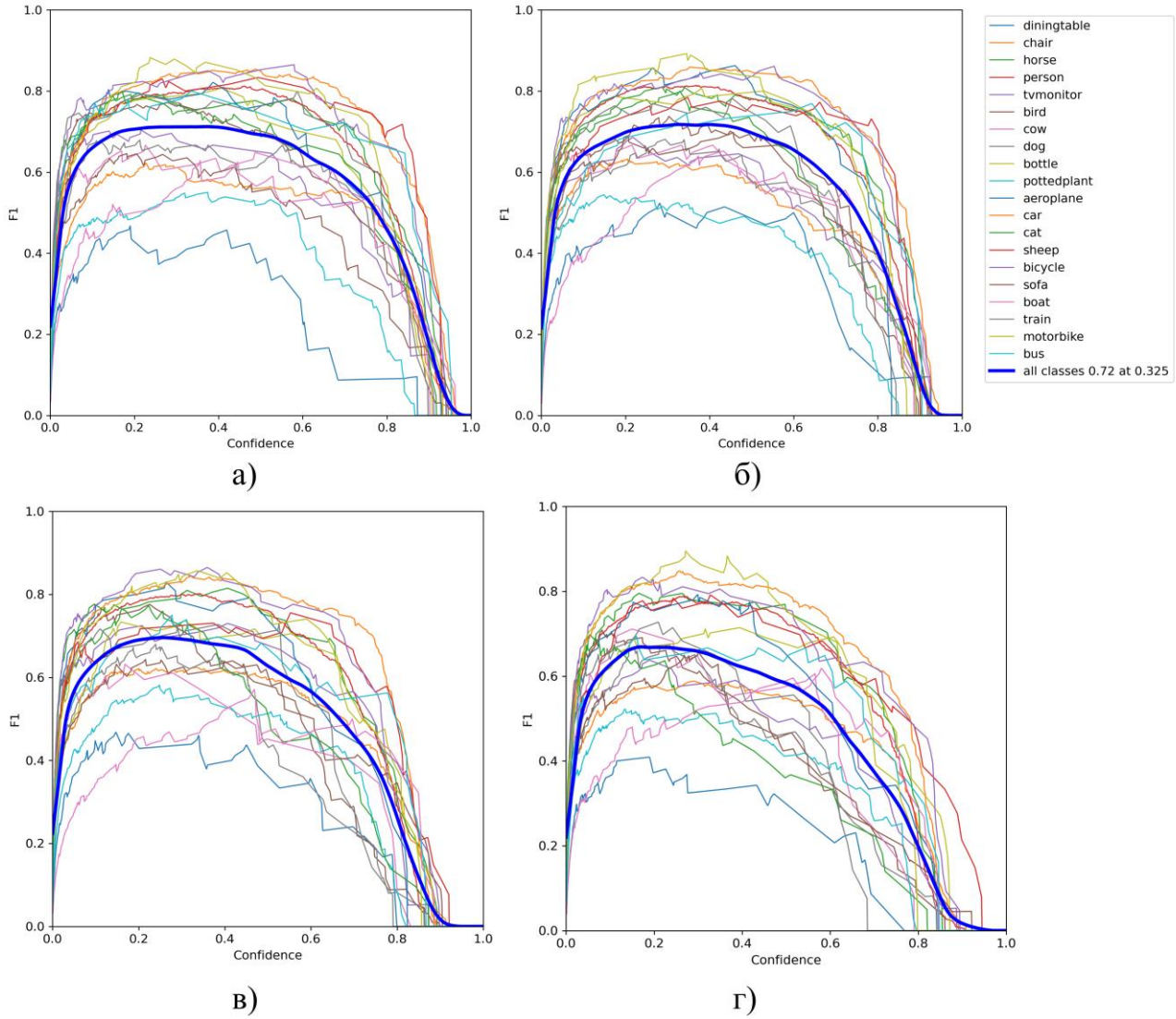


Рис. 11. Крива F1, коли навчання відбувалось на тренувальній вибірці з вмістом шуму: а) 0%; б) 10%; в) 30%; г) 50%.

З малюнків 6-8 можемо побачити, що компоненти функції втрат box loss та class loss змінюються незначною мірою на кожному експерименті, проте object loss і на етапі тренування, і на етапі валідації моделі змінюється на кожному експерименті. Величина object loss на етапі тренування пропорційно знижується, а на етапі валідації моделі на останньому етапі втрати пропорційні вмісту шуму в даних.

Метрика recall майже незначною мірою зазнала змін з підвищенням вмісту шуму. На графіку зміни метрики precision можна помітити, що значення precision при наявності шуму на 50% об’єктів на кожному етапі тренування приймає нижчі результати, ніж в інших експериментах.

Метрика mAP за умови що поріг $IoU = 0{,}5$ та $IoU \in (0{,}5;\ 0{,}95]$ знижується при збільшенні вмісту шуму в наборі даних.

Якщо проаналізувати криву F1, то для кожного експерименту можемо отримати значення, що вказані в таблиці 2.

	Значення F1	Confidence
0% шуму	0,71	0,372
10% шуму	0,72	0,325
30% шуму	0,70	0,254
50% шуму	0,67	0,170

Табл. 2 Значення F1 та confidence для розглянутих експериментів

За отриманими значеннями F1 та confidence ми можемо помітити, що тільки 50% вмісту шуму в даних сильно впливає на детекцію об’єктів моделлю. Ці висновки ми можемо підтвердити і за матрицею помилок (Додаток А).

Тобто, з отриманих результатів можна припустити, що оглянутий тип шуму помітно впливає на якість моделі розпізнавання об’єктів у випадку наявності 50% шуму в тренувальному датасеті.

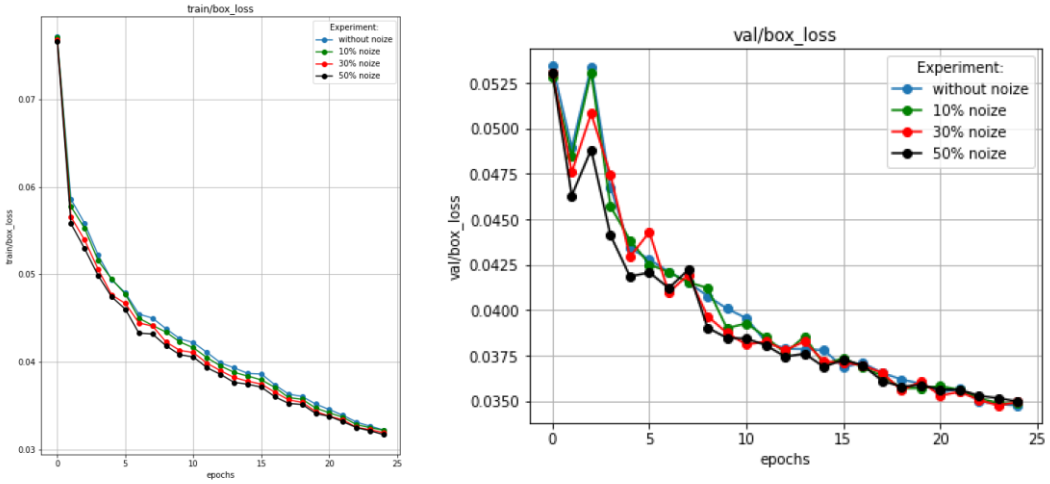
3.4 Дослідження впливу присвоєння неправильного класу об’єкта, який зображений на фото

Проведено три експерименти, в яких було додано: 1) 10% шуму в тренувальну вибірку; 2) 30% шуму в тренувальну вибірку; 3) 50% шуму в тренувальну вибірку.

Алгоритм додавання шуму зберігається таким самим, як і в попередньому дослідженні. У цих експериментах додавання шуму відбувалось шляхом зміни класу об’єкту.

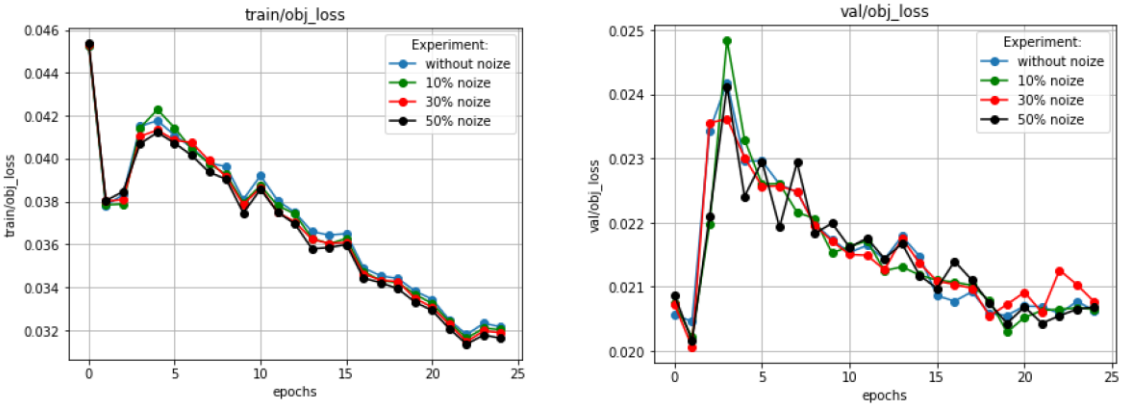
На малюнках 12-16 відображені графіки зміни значень метрик впродовж тренування моделі з шумом для експериментів: 1) вміст шуму – 0%; 2) вміст шуму – 10%; 3) вміст шуму – 30%; 4) вміст шуму – 50%.

На малюнку 17 зображено F1-криву для експериментів: 1) вміст шуму – 0%; 2) вміст шуму – 10%; 3) вміст шуму – 30%; 4) вміст шуму – 50%.



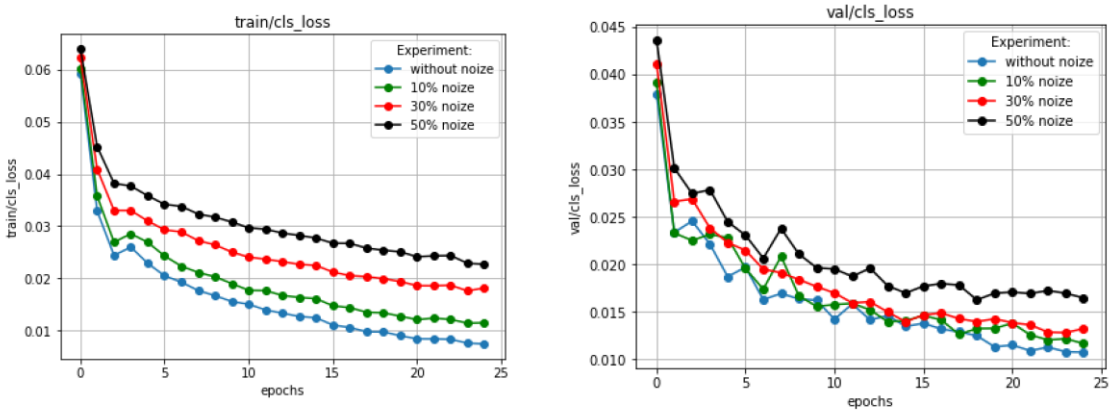
а) б)

Рис. 12. Графік зміни значення box loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



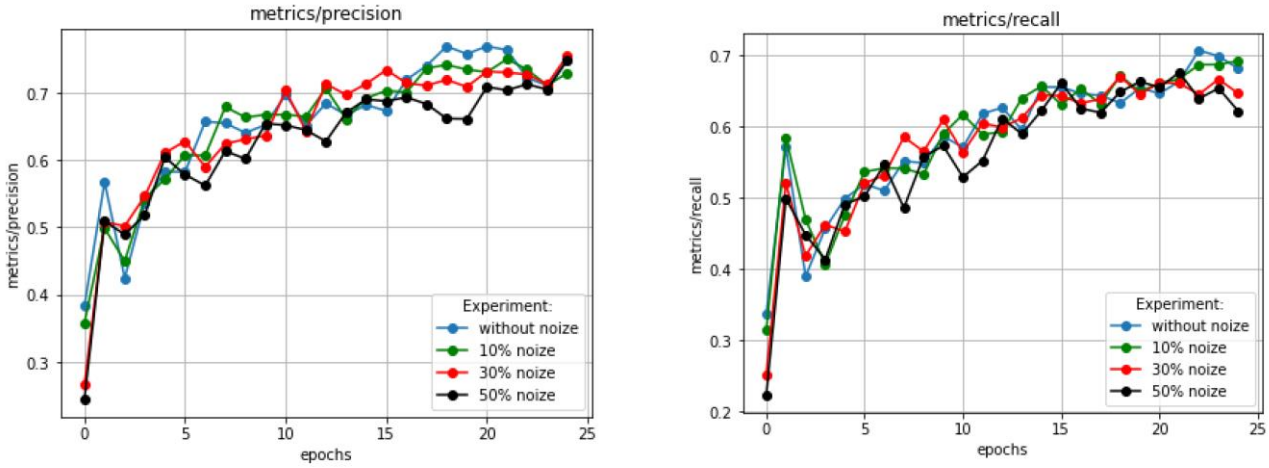
а) б)

Рис. 13. Графік зміни значення obj loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



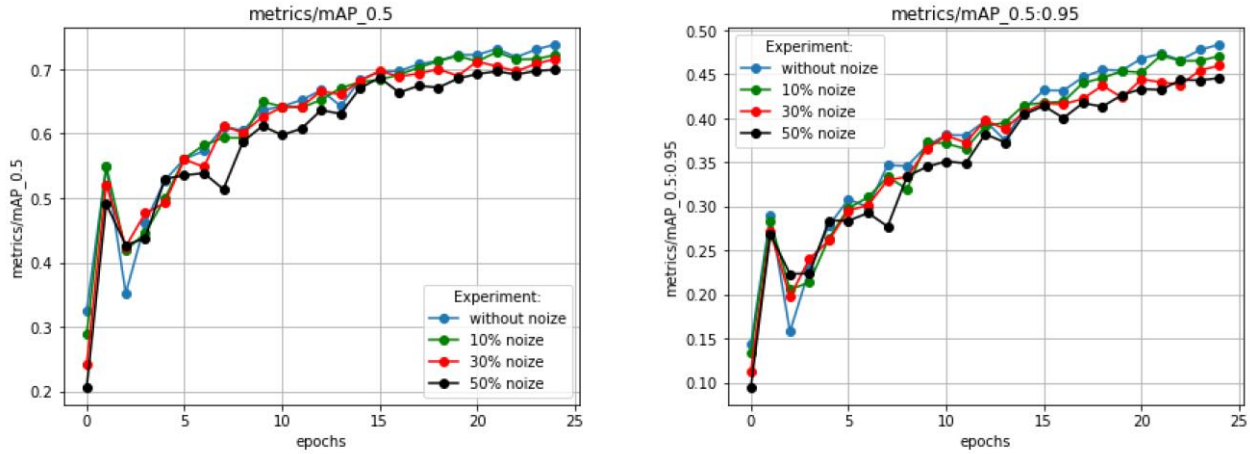
а) б)

Рис. 14. Графік зміни значення class loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



а) б)

Рис. 15. Графік зміни значення метрик для кожної епохи навчання моделі: а) влучність (precision); б) повнота (recall).



а) б)

Рис. 16. Графік зміни значення метрики mAP для кожної епохи навчання моделі: а) поріг $IoU = 0{,}5$; б) поріг $IoU \in (0{,}5;\ 0{,}95]$.

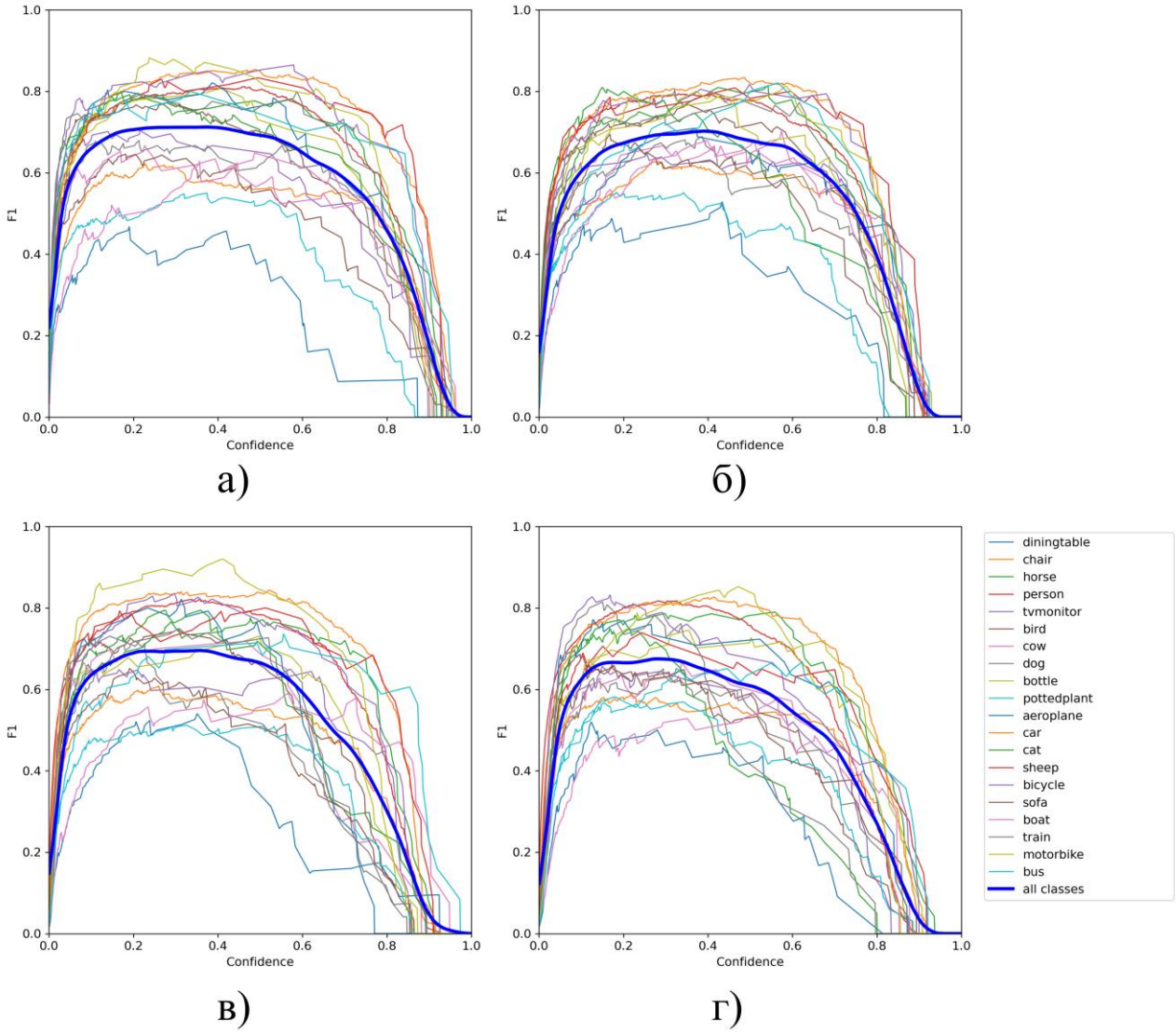


Рис. 17. Крива F1, коли навчання відбувалось на тренувальній вибірці з вмістом шуму: а) 0%; б) 10%; в) 30%; г) 50%.

З малюнків 12-14 можемо побачити, що компоненти функції втрат box loss та object loss змінюються незначною мірою на кожному експерименті, проте class loss і на етапі тренування, і на етапі валідації моделі видимо змінюється на кожному експерименті. Величина class loss на етапі тренування пропорційно знижується, а протягом всього етапу валідації моделі значення втрат пропорційні вмісту шуму в даних.

З графіку на малюнку 25 видно, що метрика precision на останньому етапі у всіх експериментах майже не відрізняється, проте можна побачити, що найкраще значення precision у експерименті, де немає шуму вище, ніж в усіх експериментах в яких шум наявний. На графіку зміни метрики recall можна

помітити, що значення recall на кінцевому етапі навчання пропорційно вмісту шуму в тренувальному наборі даних.

Метрика mAP за умови що поріг $IoU = 0{,}5$ та $IoU \in (0{,}5;\ 0{,}95]$ знижується при збільшенні вмісту шуму в наборі даних.

Якщо проаналізувати криву F1, то для кожного експерименту можемо отримати значення, що вказані в таблиці 3.

	Значення F1	Confidence
0% шуму	0,71	0,372
10% шуму	0,70	0,391
30% шуму	0,70	0,354
50% шуму	0,67	0,279

Табл. 3 Значення F1 та confidence для розглянутих експериментів

За отриманими значеннями F1 та confidence ми можемо помітити, що при збільшенні вмісту даних з шумом ці значення майже не змінюються. Ці висновки ми можемо підтвердити і за матрицею помилок (Додаток Б).

Тобто, з отриманих результатів можна припустити, що оглянутий тип шуму помітно впливає на значення метрики recall при будь-якому вмісті шуму, проте якість розпізнавання об’єктів майже не змінюється.

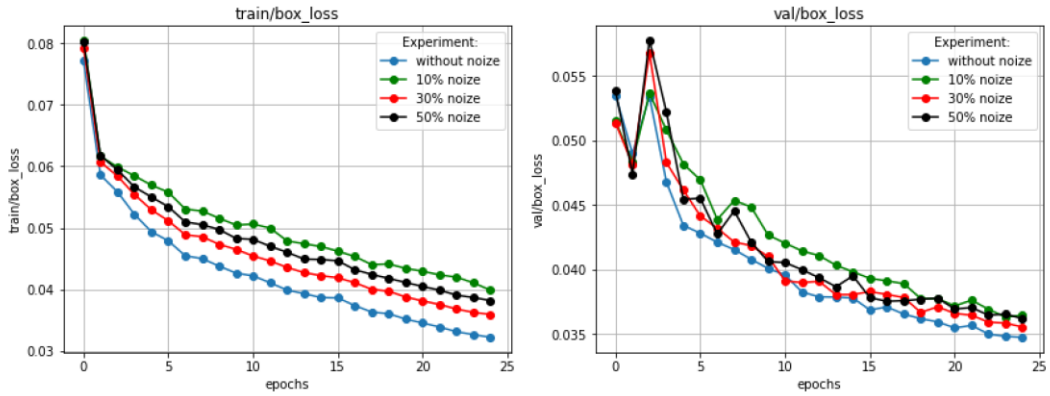
3.5 Дослідження впливу неправильного визначення обмежувальної коробки об’єкту

Проведено три експерименти, в яких було додано: 1) 10% шуму в тренувальну вибірку; 2) 30% шуму в тренувальну вибірку; 3) 50% шуму в тренувальну вибірку.

Алгоритм додавання шуму зберігається таким самим, як і в попередньому дослідженні. У цих експериментах додавання шуму відбувалось шляхом зміни координат обмежувальної коробки – зменшення її розмірів або збільшення.

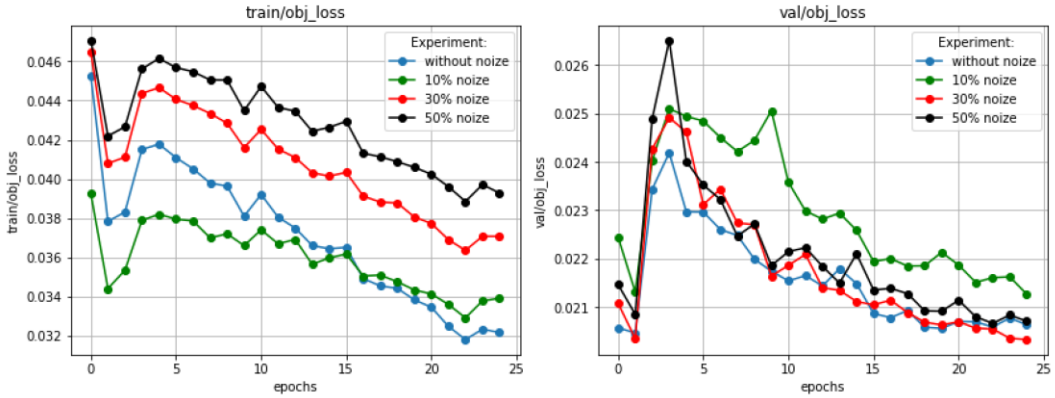
На малюнках 18-22 відображені графіки зміни значень метрик впродовж тренування моделі з шумом для експериментів: 1) вміст шуму – 0%; 2) вміст шуму – 10%; 3) вміст шуму – 30%; 4) вміст шуму – 50%.

На малюнку 23 зображено F1-криву для експериментів: 1) вміст шуму – 0%; 2) вміст шуму – 10%; 3) вміст шуму – 30%; 4) вміст шуму – 50%.



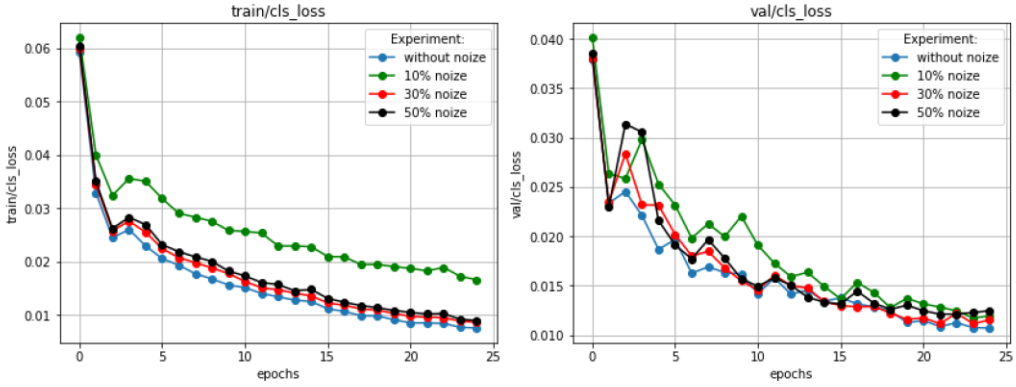
а) б)

Рис. 18. Графік зміни значення box loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



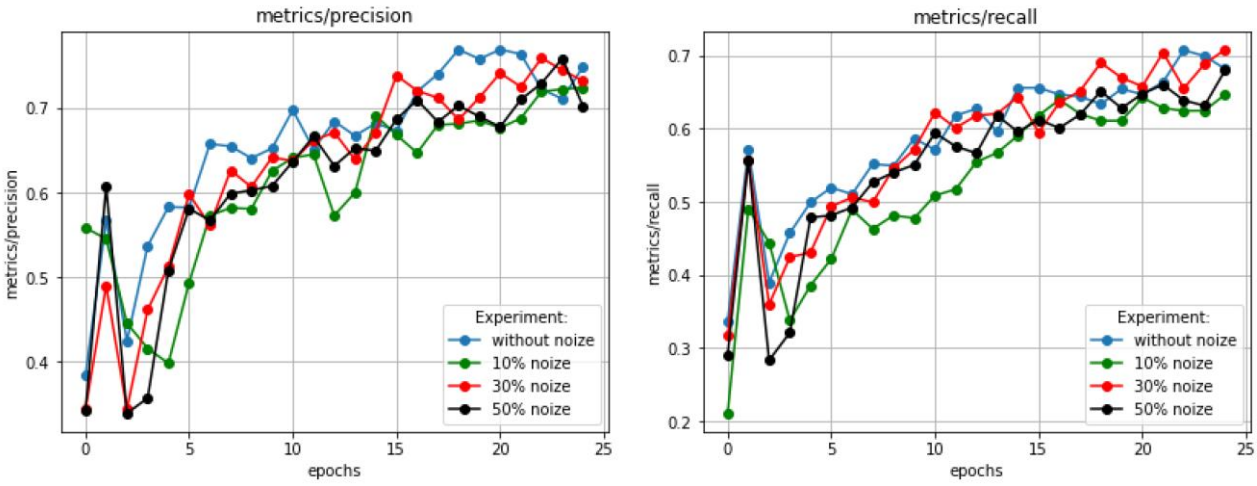
а) б)

Рис. 19. Графік зміни значення obj loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



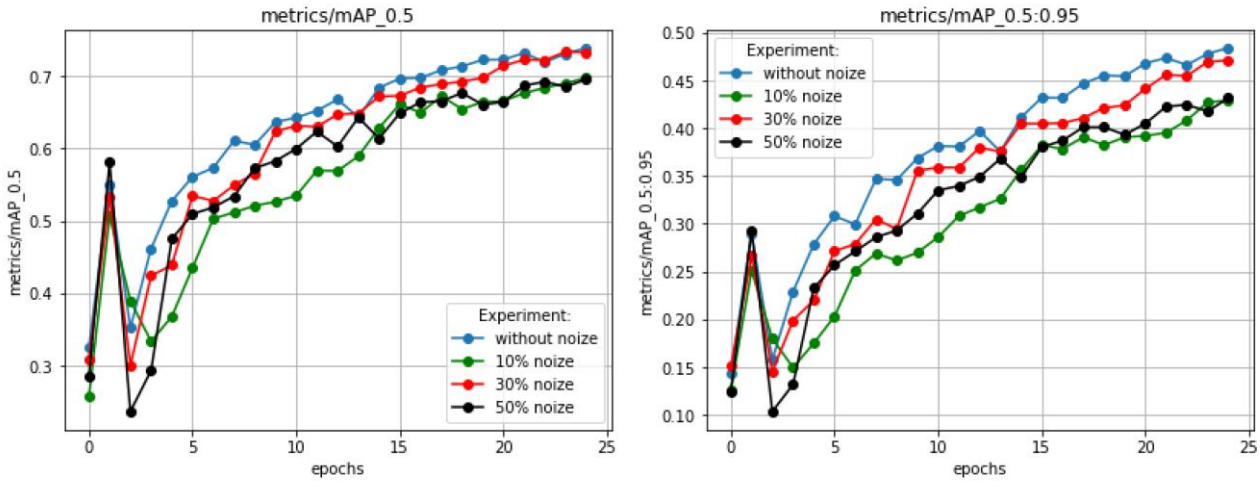
а) б)

Рис. 20. Графік зміни значення class loss для кожної епохи навчання моделі: а) етап тренування моделі; б) етап валідації моделі.



а) б)

Рис. 21. Графік зміни значення метрик для кожної епохи навчання моделі: а) влучність (precision); б) повнота (recall).



а) б)

Рис. 22. Графік зміни значення метрики mAP для кожної епохи навчання моделі: а) поріг $IoU = 0{,}5$; б) поріг $IoU \in (0{,}5;\ 0{,}95]$.

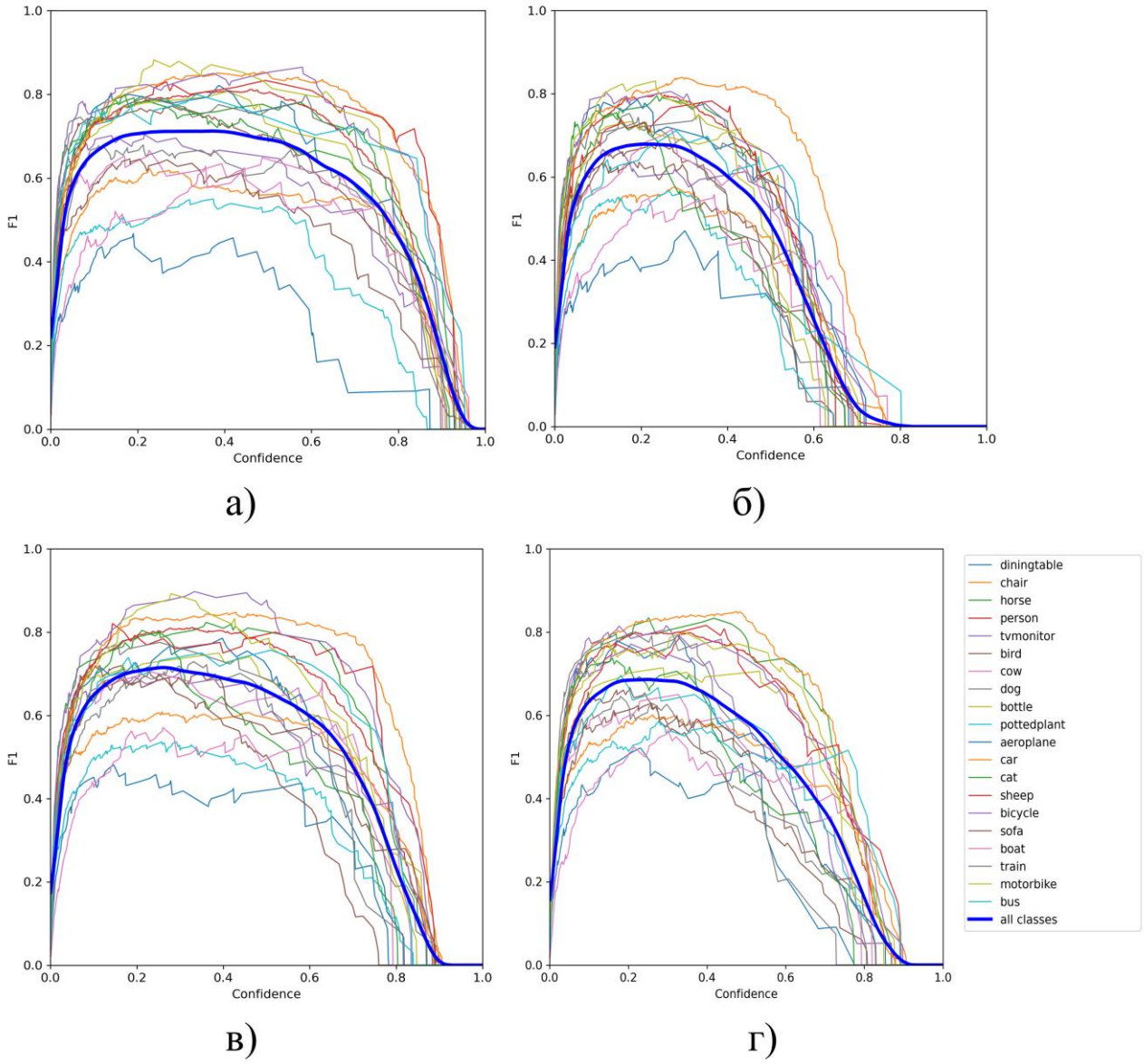


Рис. 23. Крива F1, коли навчання відбувалось на тренувальній вибірці з вмістом шуму: а) 0%; б) 10%; в) 30%; г) 50%.

З малюнків 18-20 можемо побачити, що всі компоненти функції втрат (box loss, object loss та class loss) змінюються значною мірою на кожному експерименті, при чому значення втрат на етапі валідації моделі видимо змінюється на кожному експерименті пропорційно вмісту шуму в даних.

Також з графіка на малюнку 21 можемо побачити, що досліджуваний тип шуму впливає і на значення метрики precision, і на значення метрики recall. Значення метрики precision і на найкращому етапі навчання, і на останньому етапі змінюється пропорційно вмісту шуму в навчальному датасеті.

Метрика mAP за умови що поріг $IoU = 0{,}5$ та $IoU \in (0{,}5;\ 0{,}95]$ знижується при збільшенні вмісту шуму в наборі даних, проте знижується не значущим чином.

Якщо проаналізувати криву F1, то для кожного експерименту можемо отримати значення, що вказані в таблиці 4.

	Значення F1	Confidence
0% шуму	0,71	0,372
10% шуму	0,68	0,203
30% шуму	0,71	0,261
50% шуму	0,69	0,252

Табл. 4 Значення F1 та confidence для розглянутих експериментів

За отриманими значеннями F1 та confidence ми можемо помітити, що при збільшенні вмісту даних з шумом ці значення F1 майже не змінюється, проте значення confidence різко знижується. Проглянувши матриці помилок (Додаток В) можна помітити, що цей тип шуму майже не впливає на класифікатор моделі детекції об’єктів.

Тобто, з отриманих результатів можна припустити, що оглянутий тип шуму помітно впливає на значення метрики precision, recall та F1 при будь-якому вмісті шуму, проте на етап класифікації об’єктів майже не впливає.

3.6 Висновки з аналізу впливу даних з шумом на тренування моделі детекції об’єктів.

Як можна побачити з результатів проведених експериментів будь-який шум в даних впливає на якість розпізнавання моделі детекції об’єктів.

Найменшим чином на якість тренування моделі впливає шум, який спричинений некоректно обраним класом об’єкту. Цей тип шуму видимо впливає тільки на значення повноти, проте навіть цей вплив можна вважати незначним.

Шум, що спричинений невідміченими об’єктами, що зображені на фото, більш видимо впливає на якість моделі розпізнавання об’єктів. Вплив цього

шуму на якість моделі можна оцінити зі значень метрик precision та зміни кривої F1 на кожному експерименті. Особливо значний вплив зазнається при 50% вмісту даних з шумом.

Помилки у координатах обмежувальної коробки впливають на якість розпізнавання об’єктів найбільше. Шум заданого типу не змінює якість класифікації об’єктів, проте на передбачення обмежувальної рамки сильно впливає. З даних, що в минулому розділі описують експерименти зі зміною обмежувальної коробки об’єктів, видно, що видимий вплив здійснюється на кожну метрику, яка була розглянута.

З цих результатів експериментів можна зробити висновок, що шум, який обумовлений зміною обмежувальної коробки, найбільше вражає якість моделей детекції об’єктів і його варто уникати.

РОЗДІЛ 4. МЕТОДИ ПОКРАЩЕННЯ ЯКОСТІ МОДЕЛІ ДЕТЕКЦІЇ ОБ'ЄКТІВ ЗА УМОВИ НАВЧАННЯ ЇЇ НА ДАНИХ З ШУМОМ.

У цьому розділі буде розглянуто навчальні вибірки, в яких дані з шумом кожного типу складають 50%.

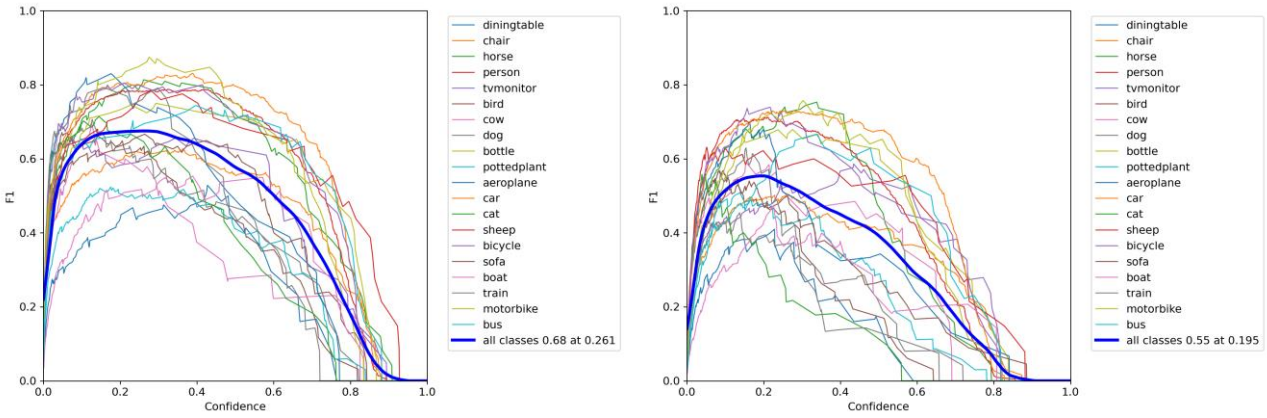
4.1 Використання ансамблю моделей для покращення якості моделі детекції.

4.1.1 Тренування ансамбля моделей на навчальній виборці, де в частині даних немає обмежувальної коробки на наявних на фото об'єктах.

Для тренування ансамбля моделей було використано модель, що навчена на вибірці, в якій 50% фото, на яких зображені об'єкти не мають необхідної обмежувальної коробки на ці об'єкти.

Другим етапом створення ансамблю було тренування нової моделі з такою самою архітектурою, проте як навчальна вибірка була використана тестувальна вибірка (test).

На малюнку 24 відображено криву F1 для першої та другої моделі.



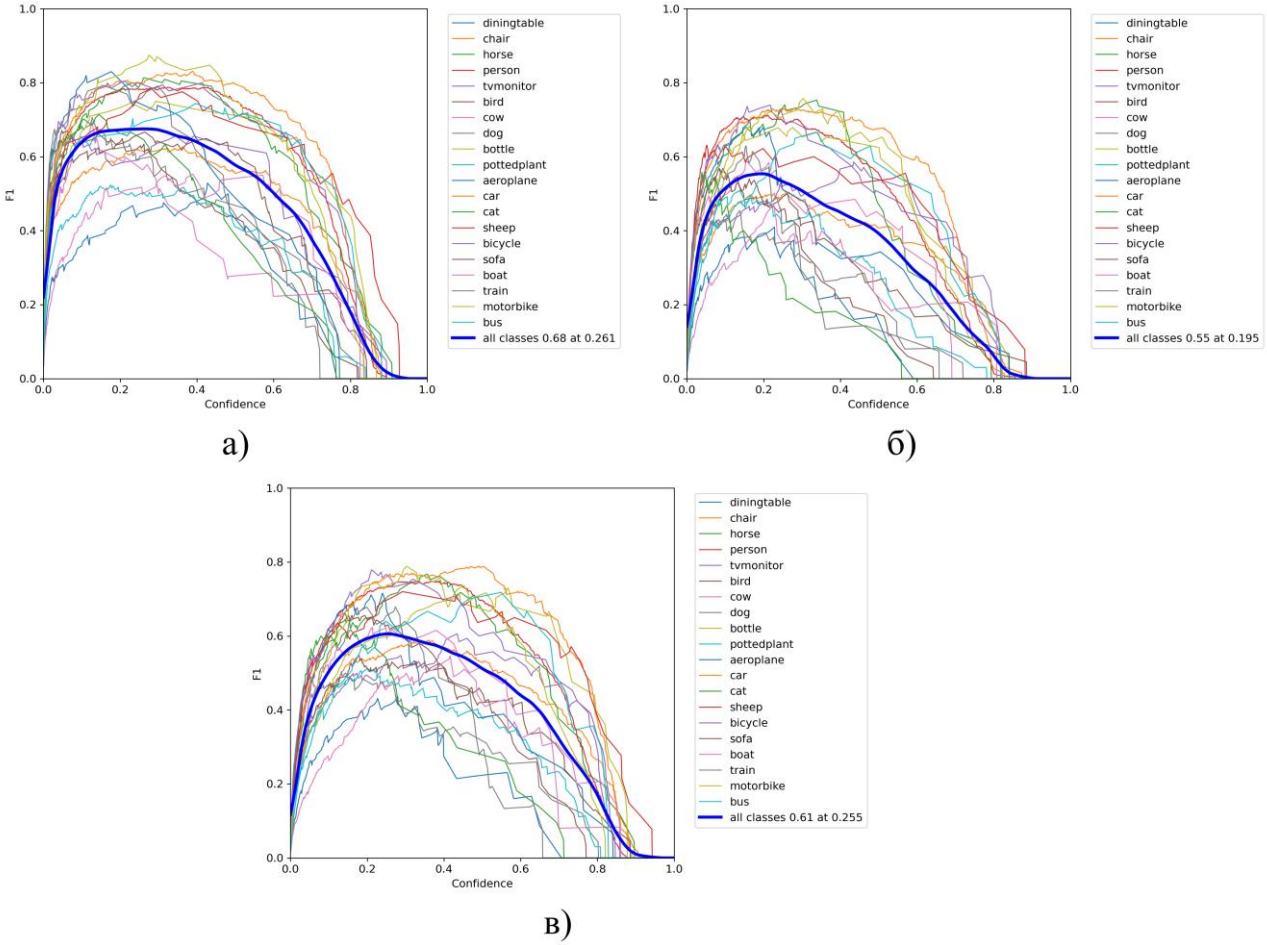
а) б)

Рис. 24. Крива F1 для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті.

Як ми можемо побачити, що друга модель має нижчі значення F1, проте зважаючи на різницю між об'ємом навчальних датасетів для першої та другої моделі (для першої моделі об'єм вибірки складає 4001 зображення, а для другої – 501 зображення).

Було свідомо обрано таку маленьку вибірку для навчання другої моделі в ансамблі, з метою показати покращення результатів моделі не генеруючи додатково великий набір розмічених даних.

Далі валідуємо ансамбль з цих двох моделей та отримуємо наступні показники якості ансамблю цих моделей.



а) б)

в)

Рис. 25. Крива F1 для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті; в) ансамбль двох моделей.

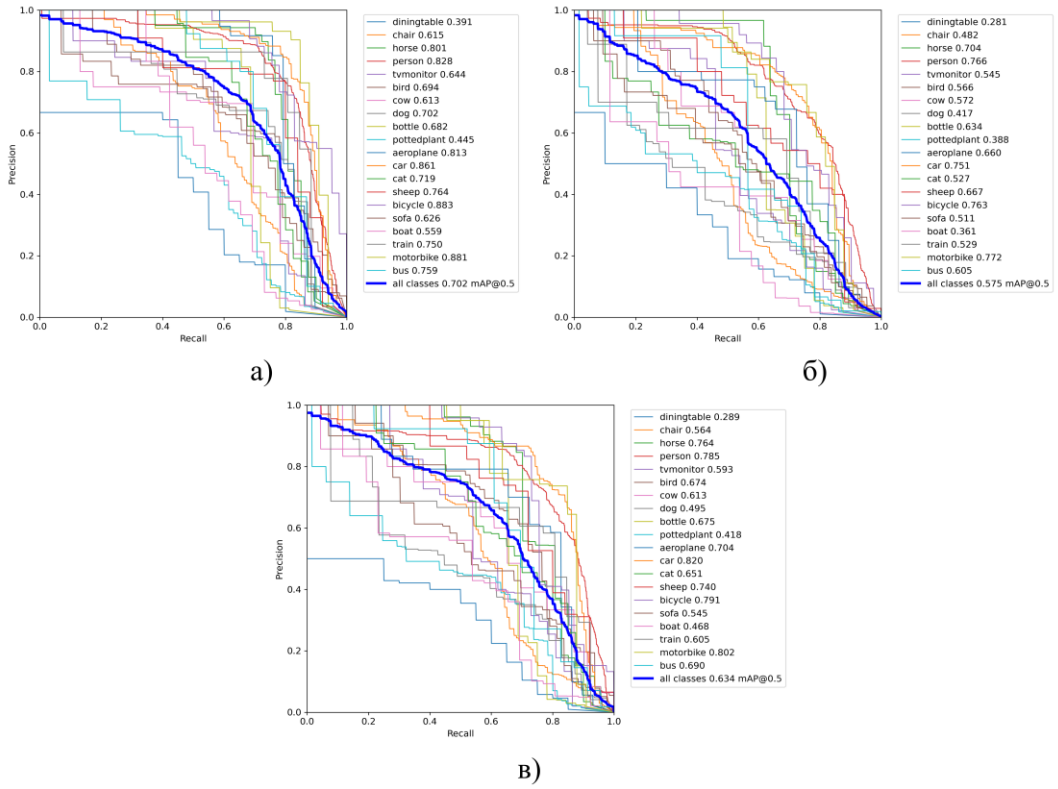


Рис. 26. Крива Precision-recall для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті; в) ансамбль двох моделей.

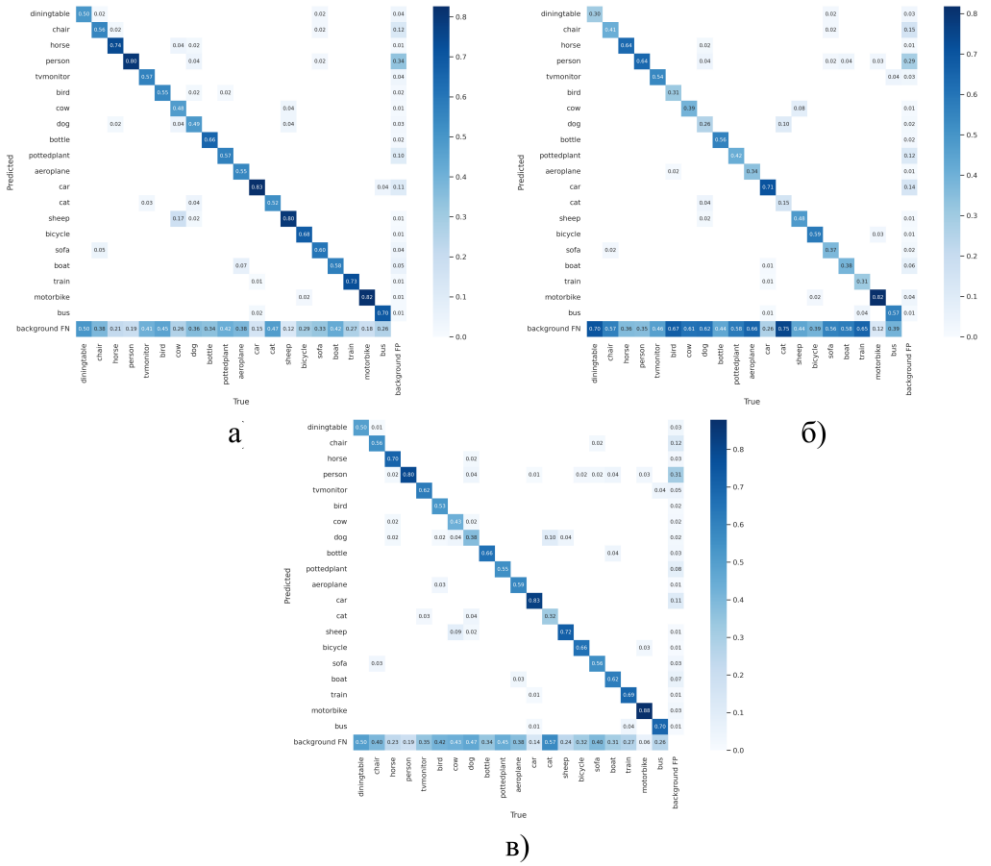


Рис. 27. Матриця помилок для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті; в) ансамбль двох моделей.

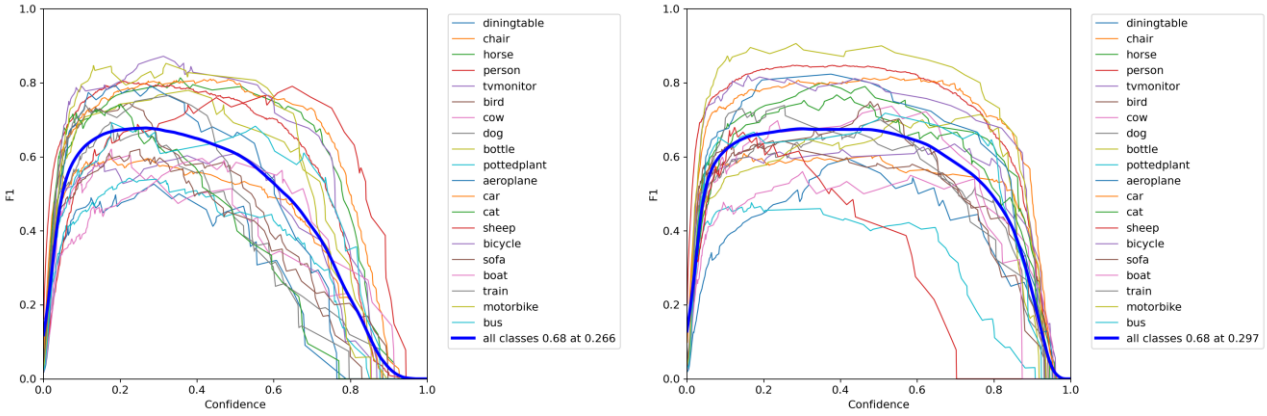
Як ми бачимо, що цей метод не покращує якість моделі, що навчена на вибірці з шумом. Є висока ймовірність, що метод ансамблю моделей покращить якість першої моделі лише за умови збільшення об’єму датасету, на якому навчається друга модель з ансамлю. Проте це призведе до збільшення часу, що необхідне для розробки моделі та збільшить його ціну.

4.1.2 Тренування ансамбля моделей на навчальній виборці, де в частині даних клас об’єкту вказано некоректно.

У цьому експерименті для тренування ансамбля моделей було використано модель, що навчена на вибірці, в якій 50% фото, на яких зображені об’єкти мають некоректний клас.

Друга модель для ансамблю була навчена на тестовому наборі даних, проте в цьому експерименті використана складніша архітектура самої моделі.

На малюнку 28 можемо порівняти криву F1 для кожної моделі, що будемо використовувати в ансамблі.



а) б)

Рис. 28. Крива F1 для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті.

Далі валідуємо ансамбль з цих двох моделей та отримуємо наступні показники якості ансамблю цих моделей.

Як ми можемо побачити з графіків, що зображені на малюнках 28-31, що у цьому експерименті відбулось тільки покращення класифікатора, на що нам

вказує матриця помилок. А значення метрик F1 та precision-recall, тільки зменшились.

В цьому експерименті ансамбль моделей теж не призводить до покращення якості моделі розпізнавання об’єктів, що навчена на датасеті з шумом.

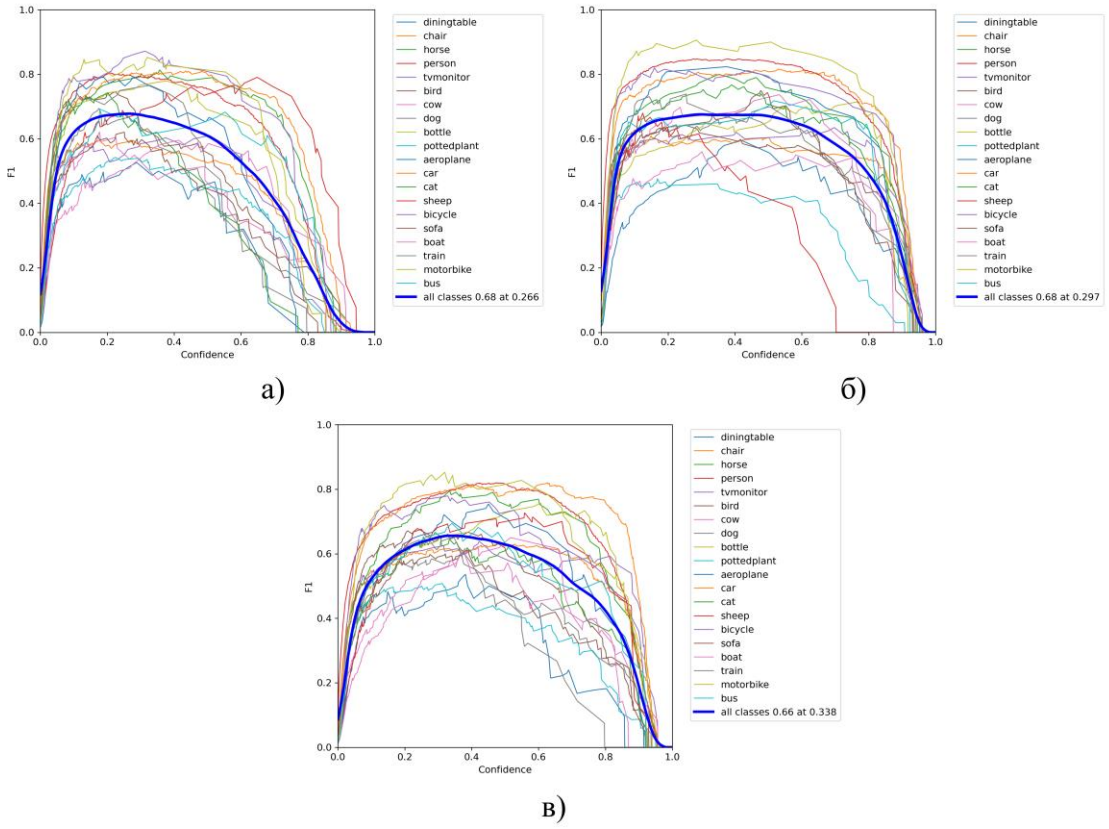


Рис. 29. Крива F1 для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті; в) ансамбль моделей.

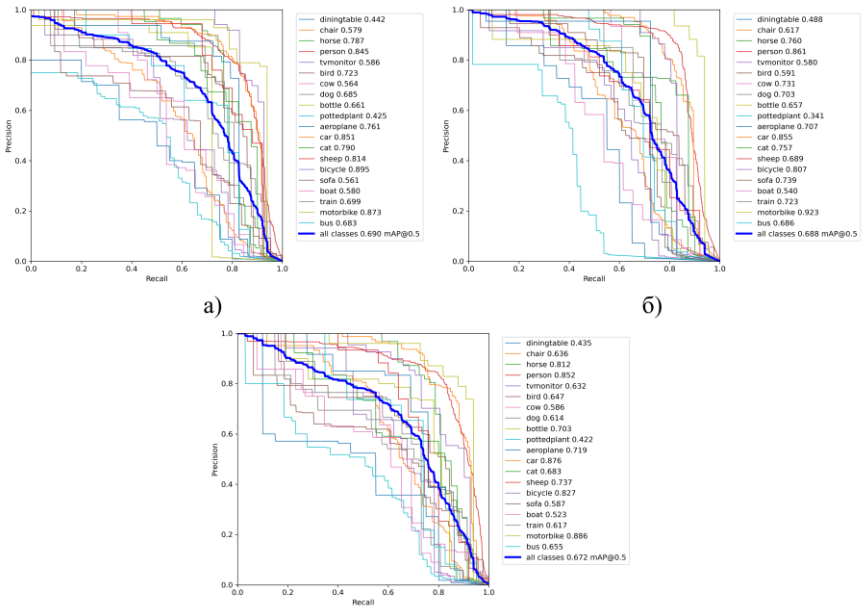


Рис. 30. Крива Precision-recall для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті; в) ансамбль моделей.

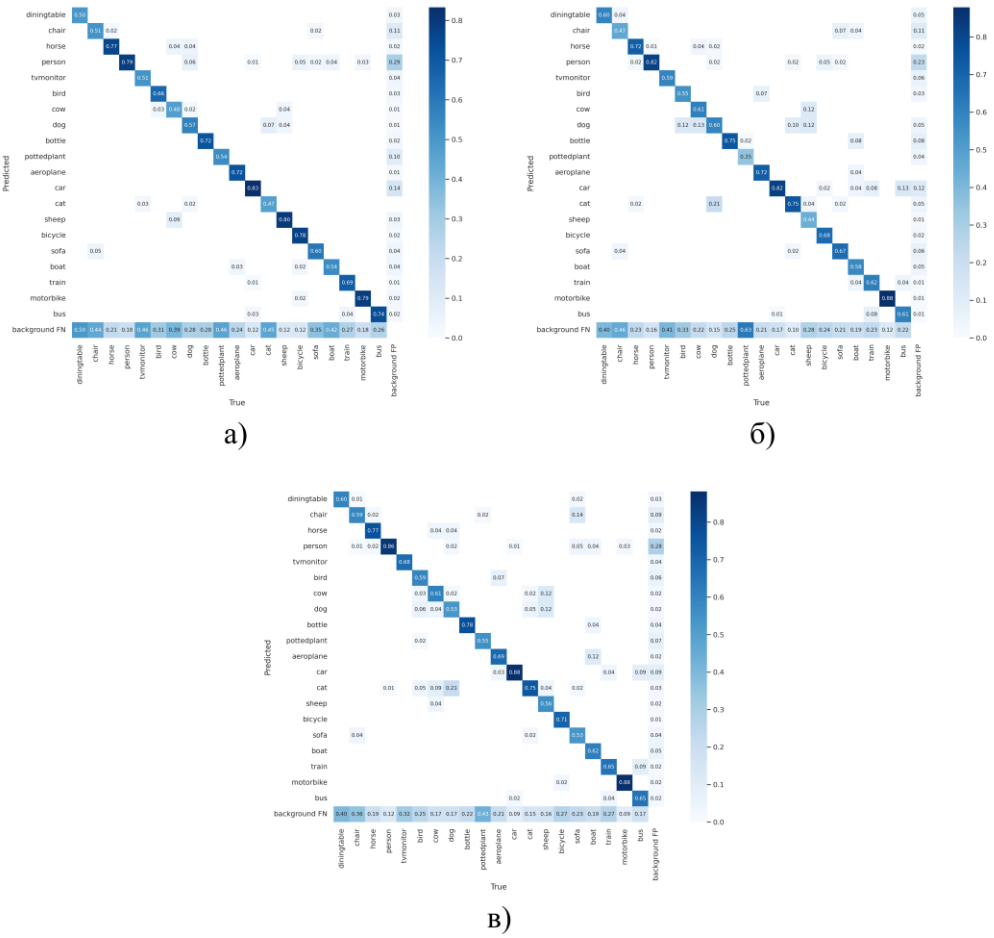


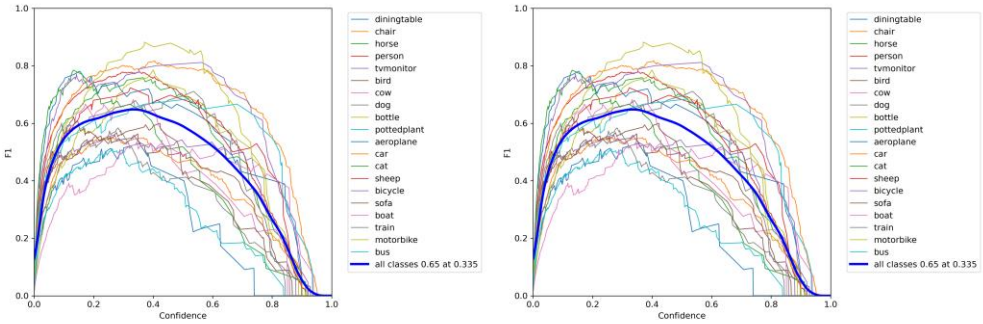
Рис. 31. Матриці помилок для: а) модель, що навчена на датасеті з шумом; б) модель, що навчена на тренувальному датасеті; в) ансамбль моделей.

4.1.2 Тренування ансамбля моделей на навчальній виборці, де в частині даних обмежувальна коробка об’єкту змінена.

У цьому експерименті для тренування ансамбля моделей було використано модель, що навчена на вибірці, в якій 50% фото, на яких зображені об’єкти мають некоректні координати обмежувальної коробки.

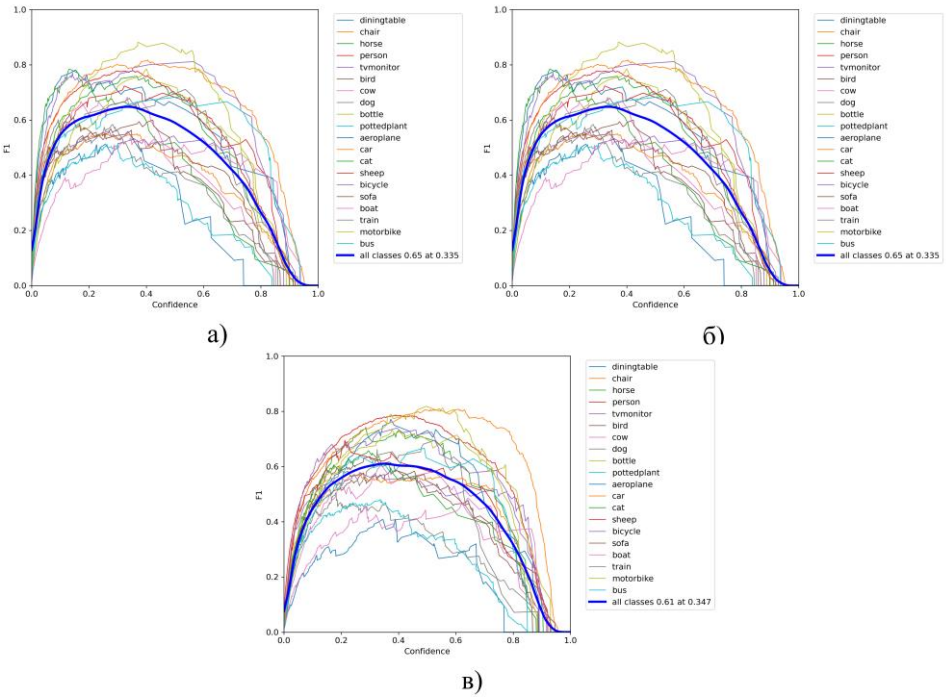
Друга модель для ансамблю була навчена на тестовому наборі даних.

На малюнку 32 можемо порівняти криву F1 для кожної моделі, що будемо використовувати в ансамблі.



а) б)

Рис. 32. Крива F1 для: а) модель, що навчена на даних з шумом; б) модель, що навчена на тестувальному датасеті.



а) б)

в)

Рис. 33. Крива F1 для: а) модель, що навчена на даних з шумом; б) модель, що навчена на тестувальному датасеті; в) ансамбль моделей.

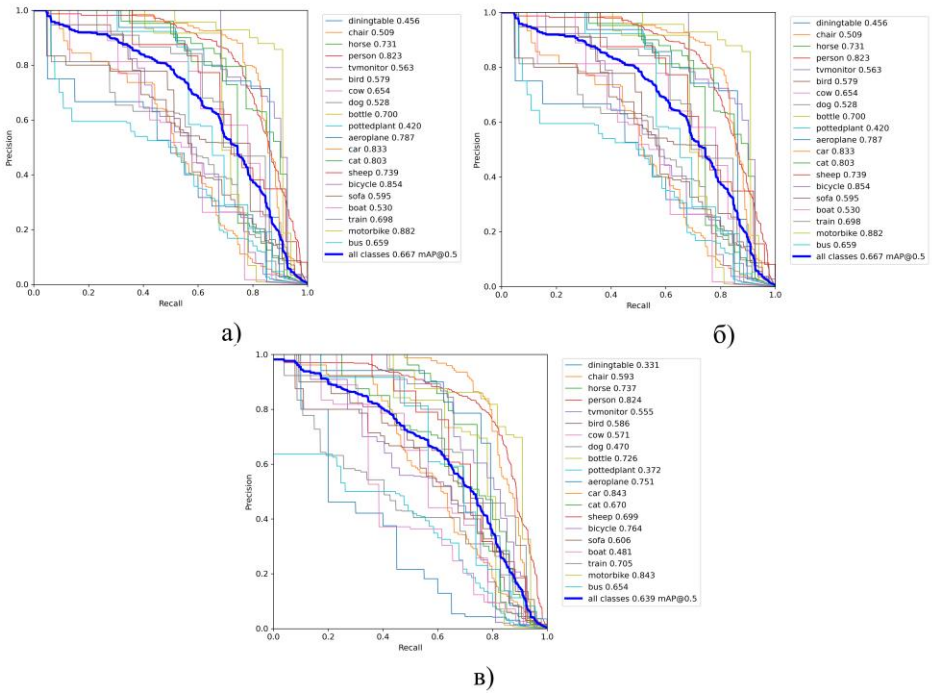


Рис. 34. Крива Precision-recall для: а) модель, що навчена на даних з шумом; б) модель, що навчена на тестувальному датасеті; в) ансамбль моделей.

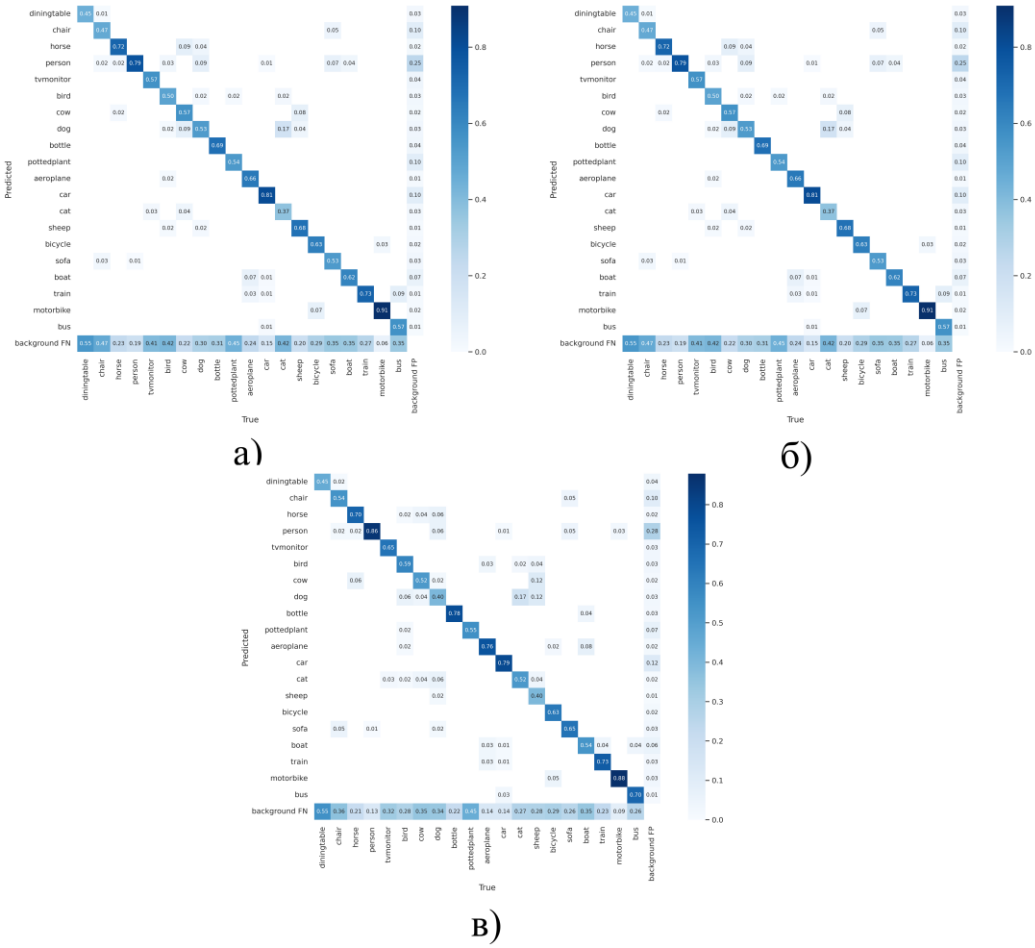


Рис. 35. Матриці помилок для: а) модель, що навчена на даних з шумом; б) модель, що навчена на тестувальному датасеті; в) ансамбль моделей.

Як ми можемо побачити з графіків, що зображені на малюнках 32-35, що у цьому експерименті не відбулось покращення. А значення метрик F1 та precision-recall, тільки зменшились.

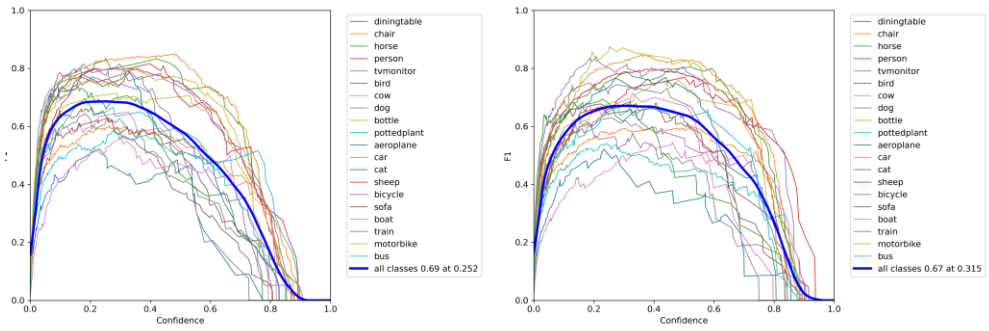
В цьому експерименті ансамбль моделей теж не призводить до покращення якості моделі розпізнавання об’єктів, що навчена на датасеті з шумом.

4.2 Використання методу TTA для покращення якості моделі детекції.

4.2.1 Використання методу TTA при тренуванні моделі детекції об’єктів на навчальній виборці, де в частині даних немає обмежувальної коробки на наявних на фото об’єкта

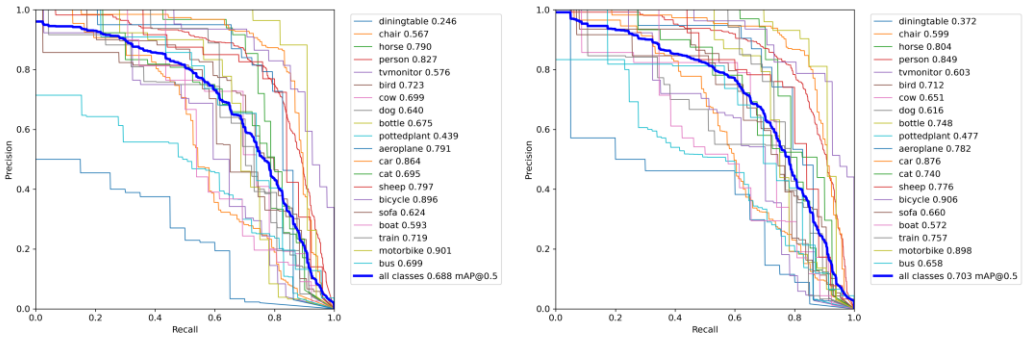
У цьому методі буде використано натреновану модель на навчальній виборці, де в частині даних немає обмежувальної коробки на наявних на фото об’єктах. Такий шум в даних складає 50% від усіх розмічених об’єктів.

На малюнках 36-38 розглянемо метрики якості моделі детекції об’єктів.



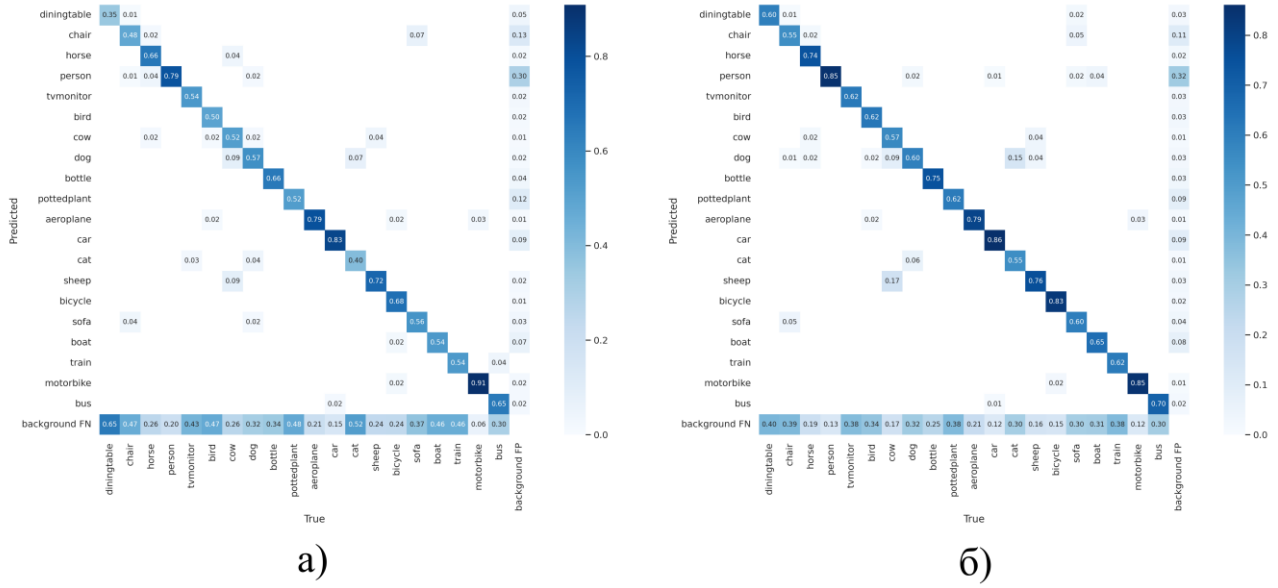
а) б)

Рис. 36. Крива F1 для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.



а) б)

Рис. 37. Крива Precision-recall для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.



а) б)

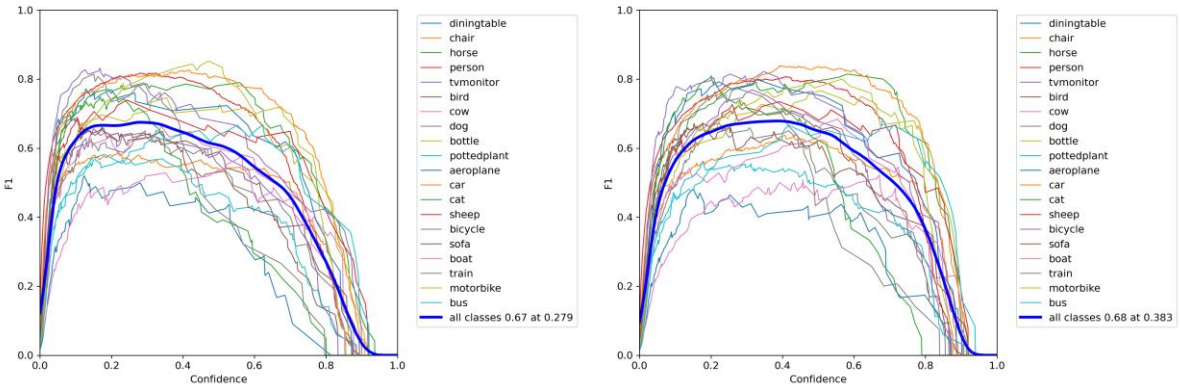
Рис. 38. Матриці помилок для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.

Як можна побачити цей метод покращив всі метрики якості моделі: і передбачення обмежувальних рамок об’єкту, і класифікацію об’єкту. Можна зробити висновок, що метод TTA успішно покращує якість моделі розпізнавання об’єктів, навчальна вибірка якої містить 50% об’єктів з шумом заданого типу.

4.2.2 Використання методу TTA при тренуванні моделі детекції об’єктів на навчальній виборці, де в частині даних клас об’єкту некоректно визначений.

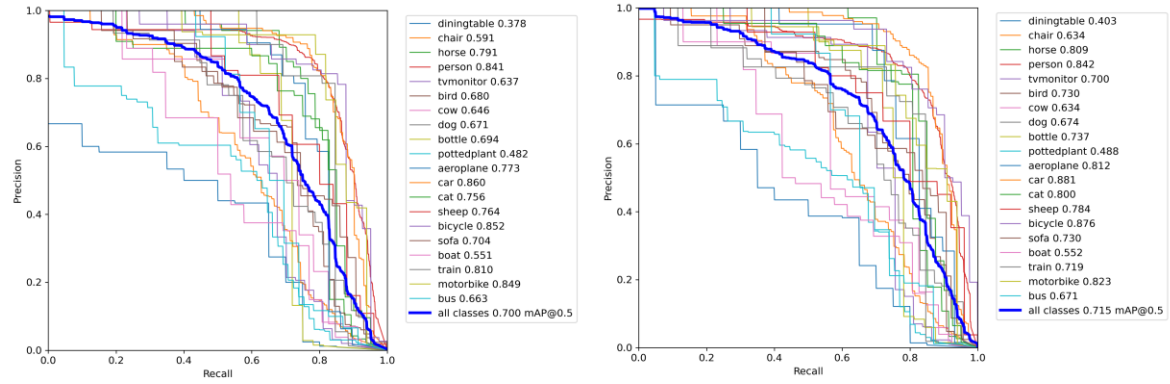
У цьому методі буде використано натреновану модель на навчальній виборці, де в частині даних клас об’єкту визначений некоректно. Такий шум в даних складає 50% від усіх розмічених об’єктів.

На малюнках 39-41 представлено метрики вимірювання якості моделей детекцій об’єктів, що навчена на вибірці з вмістом шуму, та моделі, що завалідована з використанням методу TTA. В цьому експерименті відбулось незначне покращення кожної з метрик. Можна зробити висновок, що метод TTA незначним чином покращую якість моделі розпізнавання об’єктів, навчальна вибірка якої містить 50% об’єктів з шумом заданого типу.



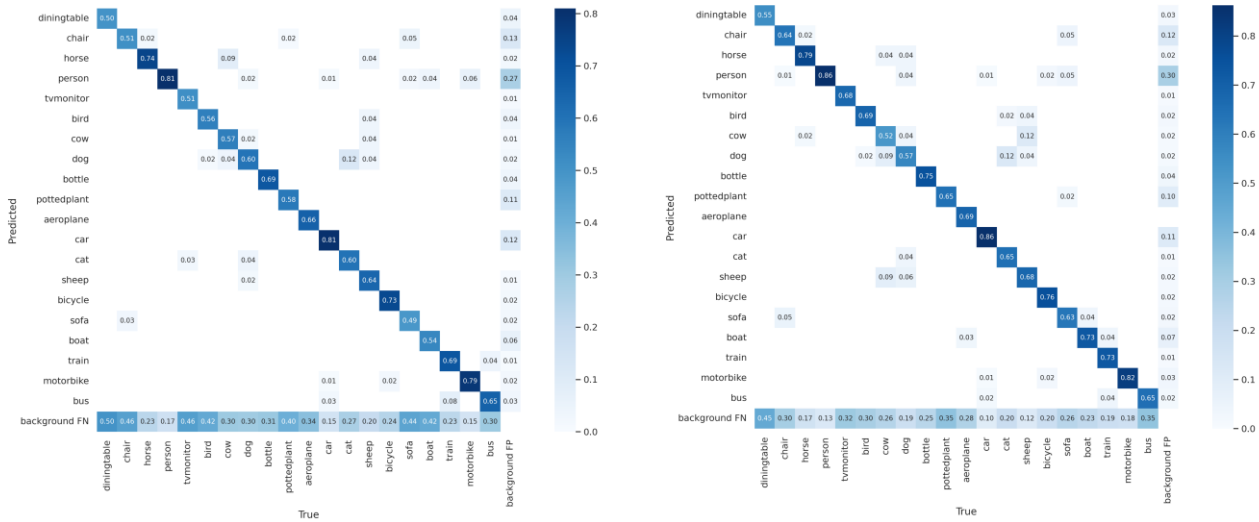
а) б)

Рис. 39. Крива F1 для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.



а) б)

Рис. 40. Крива Precision-recall для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.



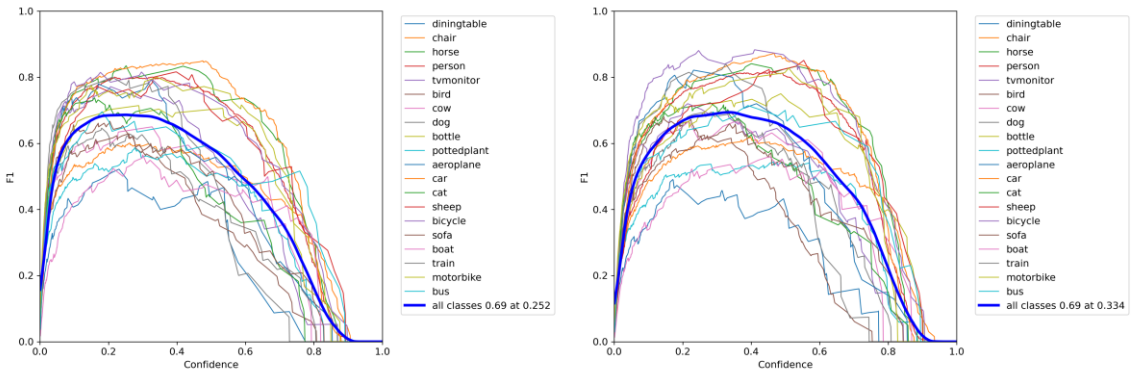
а) б)

Рис. 41. Матриця помилок для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.

4.2.3 Використання методу TTA при тренуванні моделі детекції об’єктів на навчальній виборці, де в частині даних обмежувальна рамка не відповідає розмірам заданого об’єкту.

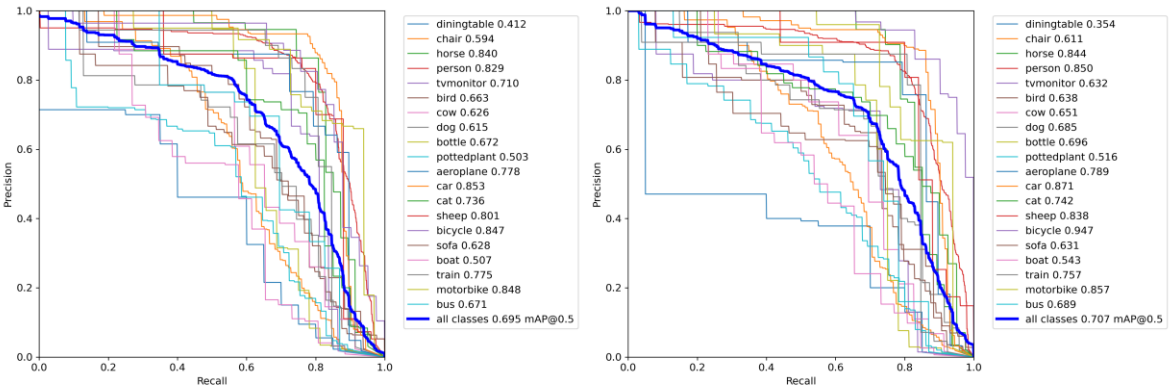
У цьому методі буде використано натреновану модель на навчальній виборці, де в частині даних клас об’єкту визначений некоректно. Такий шум в даних складає 50% від усіх розмічених об’єктів.

На малюнках 42-44 представлено метрики вимірювання якості моделей детекцій об’єктів, що навчена на вибірці з вмістом шуму, та моделі, що завалідована з використанням методу TTA. В цьому експерименті відбулось незначне покращення метрик F1 та precision-recall. Варто зауважити, що значно підвищився рівень класифікації об’єктів. Можна зробити висновок, що метод TTA покращує якість класифікації об’єктів у випадку навчання на вибірці з розглянутим типом шуму.



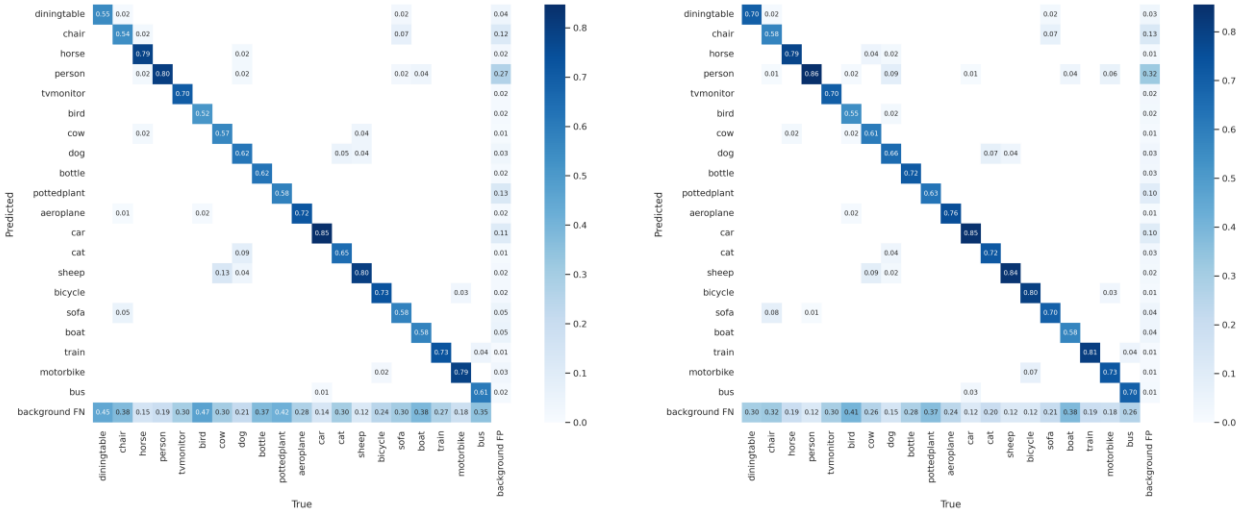
а) б)

Рис. 42. Крива F1 для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.



а) б)

Рис. 43. Крива Precision-recall для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.



а) б)

Рис. 44. Матриця помилок для: а) модель, що натренована на даних з шумом; б) модель з використанням методу TTA.

ВИСНОВОК

У цій роботі було дослідження вплив трьох найпоширеніших типів шуму у навчальній вибірці, які виникають при розмітці даних, на якість натренованої моделі. Було розглянуто наступні типи шуму: неправильно визначений клас виділеного об’єкту, відсутність виділення об’єкту, що присутній на зображенні, та некоректно визначення координат обмежувальної коробки об’єкту. Дослідження відбувалось шляхом порівняльного аналізу метрик якості моделі розпізнавання, що навчена на даних без шуму, та метрик моделі, що в тренувальній вибірці містить певний об’єм шуму визначеного типу.

Для оцінки якості моделі детекції об’єктів було використано наступні метрики: mAP, precision, recall та крива F1. Також для визначення стійкості моделі до обраного типу шуму розглядалися зміни значення трьох складових функцій втрат: box loss, class loss та obj loss.

Результат проведеного аналізу показує, що всі тип шуму найбільшою мірою впливають на визначення обмежувальної коробки об’єкту та майже не впливають на класифікацію виділеного об’єкту. Найнебезпечнішим для якості розпізнавання типом шуму з тих, що було розглянуто, є некоректне визначення координат обмежувальної рамки об’єкту. Неправильне визначення класу виділеного об’єкту майже не впливає на якість моделі розпізнавання.

Наступним етапом дослідження було реалізація двох методів підвищення стійкості моделі детекції об’єктів – метод ансамбля моделей та метод test-time augmentation. З цих експериментів видно, що обидва методи підвищують якість моделі розпізнавання, проте не значущим чином. Для моделі архітектури YOLO більш ефективним виявився метод TTA для всіх типів шуму в навчальній вибірці.

В результаті виконання цього дослідження можна виділити наступні тези: процес розмітки даних є важливим етапом створення моделі детекції об’єктів, а автоматизація цього процесу зменшує витрати на нього проте збільшує ризики отримання в процесі навчання модель детекції низької якості; існуючі методи

збільшення стійкості моделі до шуму, створеного в процесі розмітки, не завжди сильно підвищують якість виділення об’єктів на зображеннях.

Розроблений програмний застосунок можна використовувати для дослідження впливу інших типів шуму, що виникають в процесі розмітки зображень, та аналіз стійкості моделей детекції інших архітектур до помилок у розмітці об’єктів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Joseph Redmon, Ali Farhadi “YOLOv3: An Incremental Improvement”.

[2] Arthur Vilhelm, Matthieu Limbert, Clement Audebert, Tugdual Ceillier “Ensemble Learning techniques for object detection in high-resolution satellite images”. Earthcube, 2019.

[3] Á Casado-García, J Heras “Ensemble Methods for Object Detection”. ECAI 2020, 2020.

[4] Mohammad Eshaghian “Investigation of the Effect of Noise on Tracking Objects using Deep Learning”. International Jornal of Nonlinear Analysis and Applications, 2020.

[5] Yali Amit, Pedro Felzenszwalb, Ross Girshick “Object Detection”. 2020

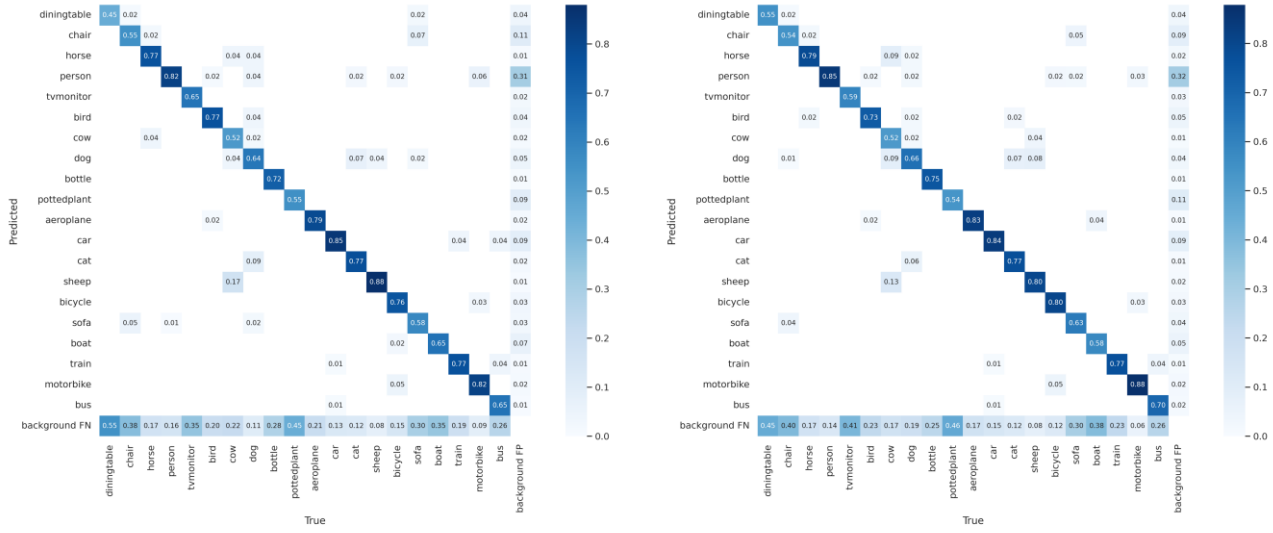
[6] Mark Everingham, John Winn “The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Development Kit”. 2007

[7] Glen Jocher and other YOLOv5 https://github.com/ultralytics/yolov5

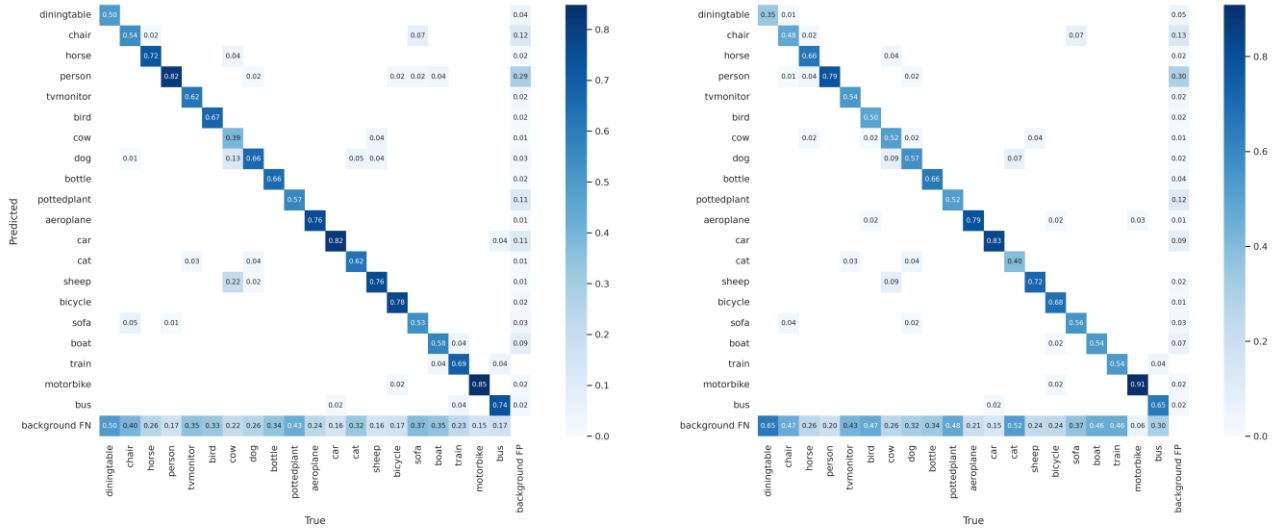
Додаток А

Матриці помилок моделей детекцій об’єктів, при наяності в тренувальному датасеті шуму.

Тип шуму, який додано до набору даних: об’єкт зображено на фото, проте його обмежувальної рамки немає.



а) б)



в) г)

а) Матриця помилок моделі, що навчена на даних без вмісту шуму

б) Матриця помилок моделі, що навчена на даних при вмісту 10% шуму в тренувальному датасеті.

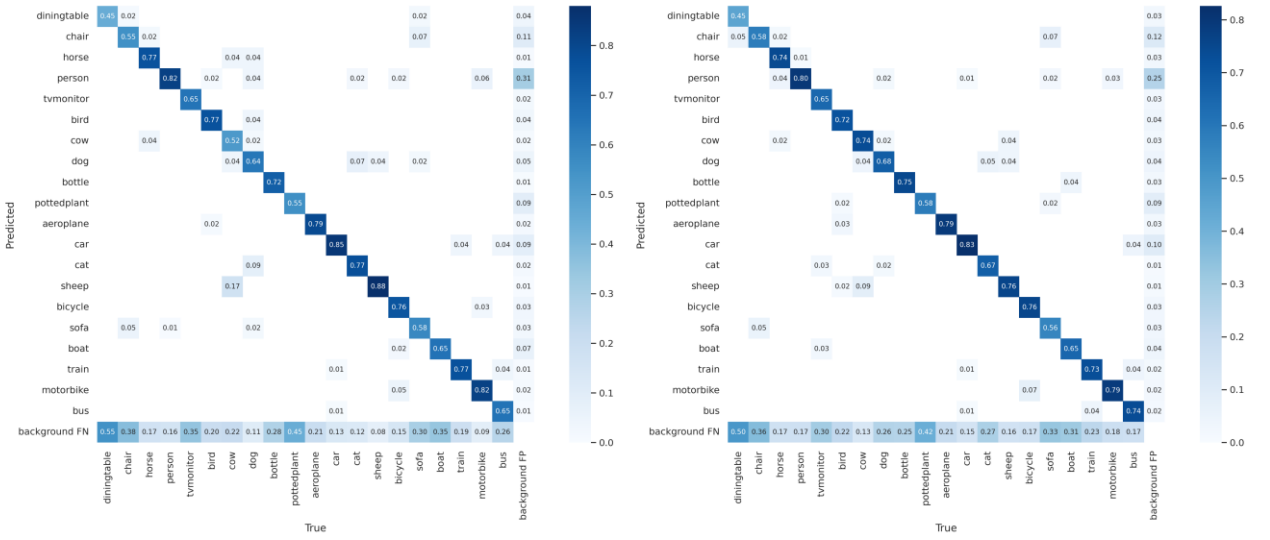
в) Матриця помилок моделі, що навчена на даних при вмісту 30% шуму в тренувальному датасеті.

г) Матриця помилок моделі, що навчена на даних при вмісту 50% шуму в тренувальному датасеті.

Додаток Б

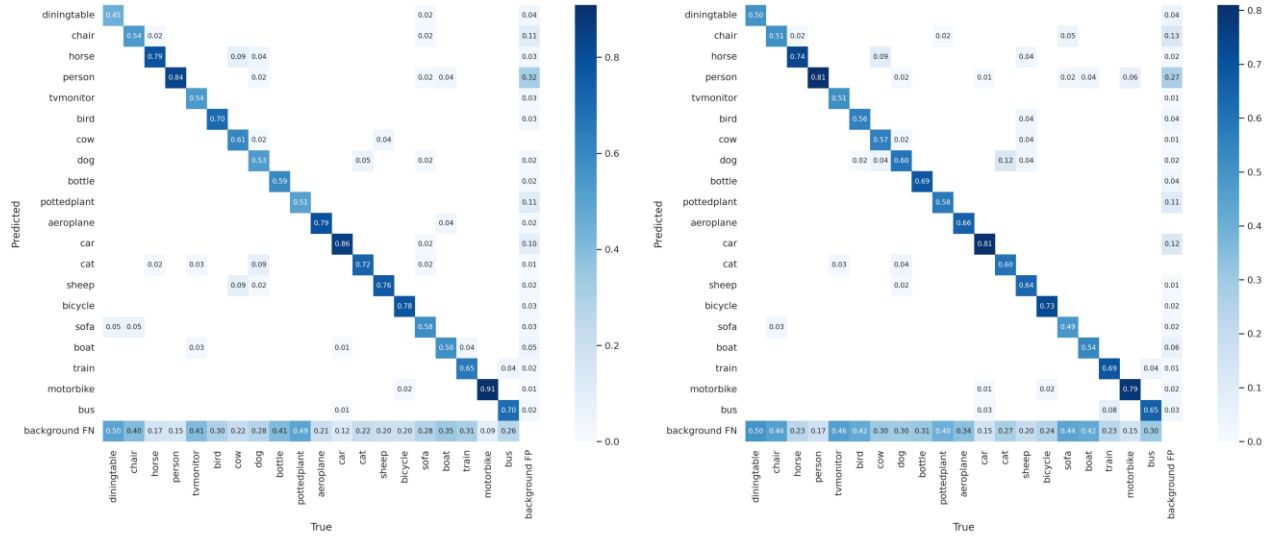
Матриці помилок моделей детекцій об’єктів, при наяності в тренувальному датасеті шуму.

Тип шуму, який додано до набору даних: клас об’єкту містить помилку.



а)

б)



в)

г)

а) Матриця помилок моделі, що навчена на даних без вмісту шуму

б) Матриця помилок моделі, що навчена на даних при вмісту 10% шуму в тренувальному датасеті.

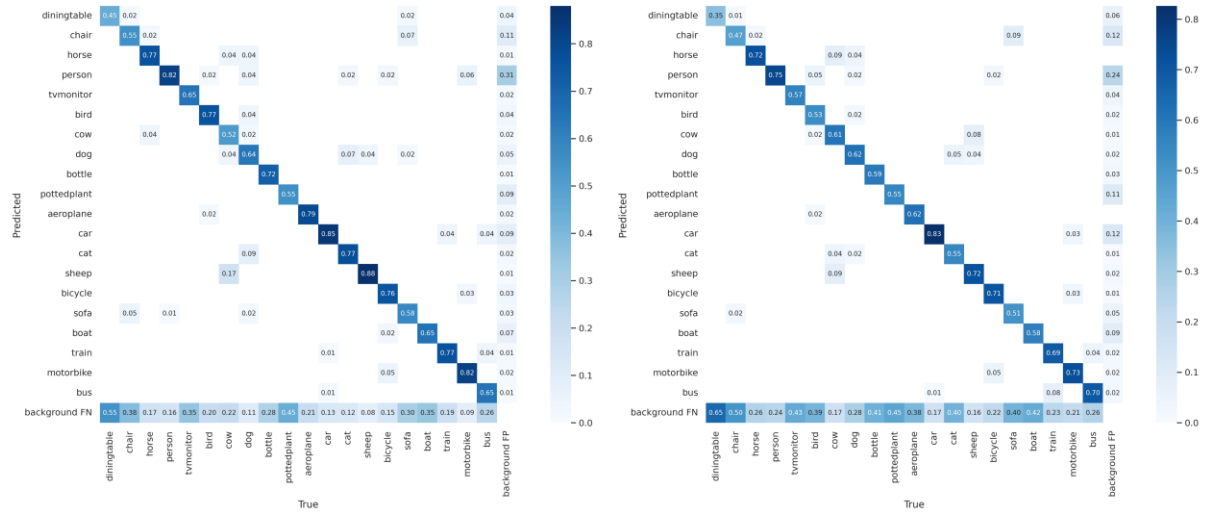
в) Матриця помилок моделі, що навчена на даних при вмісту 30% шуму в тренувальному датасеті.

г) Матриця помилок моделі, що навчена на даних при вмісту 50% шуму в тренувальному датасеті.

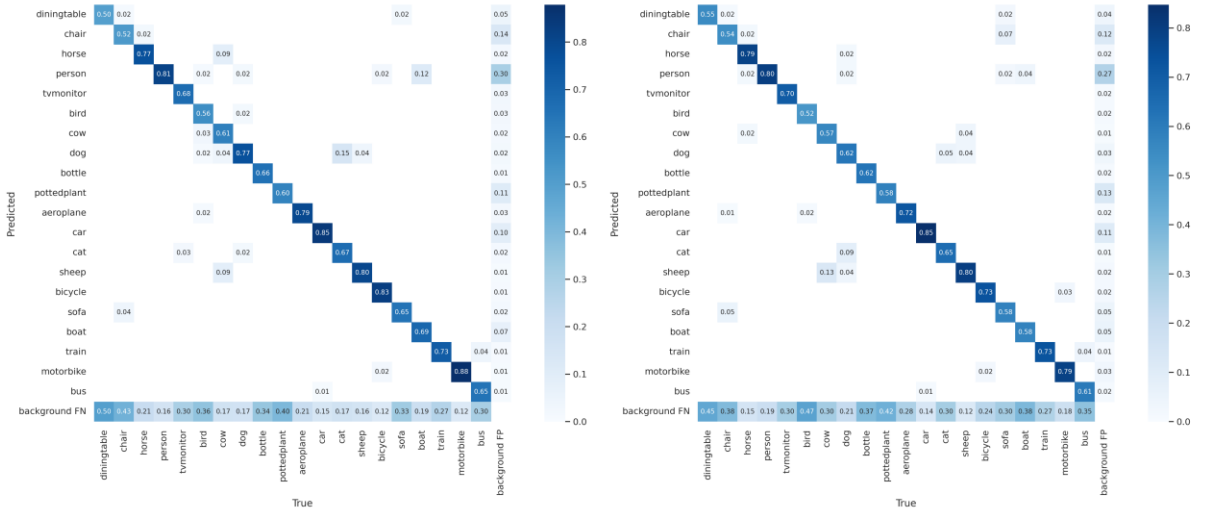
Додаток В

Матриці помилок моделей детекцій об’єктів, при наяності в тренувальному датасеті шуму.

Тип шуму, який додано до набору даних: змінено координати обмежувальної коробки.



а) б)



в) г)

а) Матриця помилок моделі, що навчена на даних без вмісту шуму

б) Матриця помилок моделі, що навчена на даних при вмісту 10% шуму в тренувальному датасеті.

в) Матриця помилок моделі, що навчена на даних при вмісту 30% шуму в тренувальному датасеті.

г) Матриця помилок моделі, що навчена на даних при вмісту 50% шуму в тренувальному датасеті.