

БАЛОВСЯК СЕРГІЙ

Чернівецький національний університет імені Юрія Федьковича

<https://orcid.org/0000-0002-3253-9006>e-mail: s.balovsyak@chnu.edu.ua**СТЕЦЬ СЕРГІЙ**

Чернівецький національний університет імені Юрія Федьковича

<https://orcid.org/0009-0007-0231-9970>e-mail: stets.serhii@chnu.edu.ua

АВТОМАТИЗОВАНЕ СТВОРЕННЯ СПЕЦІАЛІЗОВАНОГО ДАТАСЕТУ ДЛЯ ЗОБРАЖЕНЬ АВТОМОБІЛІВ

Предметом дослідження є алгоритм та програмна для системи автоматизованого створення спеціалізованого датасету (набору даних) зображень автомобілів. Розроблена програма функціонує на базі мікрокомп'ютера Raspberry Pi 5, а початкові зображення зчитуються з відеокамери. Програму розроблено на мові Python із використанням хмарних технологій. Завдання програми полягають у зчитування початкових зображень, селекції унікальних кадрів та формування датасету з таких кадрів. Селекція кадрів складається з 3 етапів. Перший етап полягає у глобальній селекції кадрів із використанням перцептивного хешування (pHash), яка дає змогу відкинути всі зображення, які в цілому є подібними до зображень датасету. На другому етапі виконується селекція кадрів із використанням детекції та аналізу зображень автомобілів, у результаті чого отримуються тільки кадри, які містять нові автомобілі порівняно з кадрами датасету. Детекція автомобілів на зображеннях виконується за допомогою згорткової нейронної мережі YOLO. Третій етап полягає у локальній селекції ділянок автомобілів із застосуванням перцептивного хешування, у результаті чого отримуються лише кадри з новими для датасету об'єктами. Під час експериментальної перевірки розроблена система упродовж 1,5 години роботи зафіксувала 993 зображення, з яких після селекції отримано 121 унікальний кадр (~12 % від початкового набору). Виявлення автомобілів на зображеннях виконується попередньо навченою моделлю YOLOv8m середнього розміру, яка забезпечує високу точність детекції об'єктів. Отриманий спеціалізований датасет зображень застосовано для донавчання більш швидких моделей YOLOv8n. Сформований датасет характеризується високою різноманітністю та придатністю для подальшого використання у задачах навчання штучних нейронних мереж.

Ключові слова: селекція зображень, датасет, виявлення автомобілів, згорткові нейронні мережі, хмарні технології, програмне забезпечення.

BALOVSYAK SERHIY**STETS SERHII**

Yuriy Fedkovych Chernivtsi National University

AUTOMATED CREATION OF A SPECIALIZED DATA SET FOR CAR IMAGES

The subject of this research is the algorithm and software for a system that automatically creates a specialized dataset of car images. The developed program runs on a Raspberry Pi 5 microcomputer, and the initial images are captured from a video camera. The software is written in Python and utilizes cloud technologies. The main tasks of the program are to capture initial images, select unique frames, and form a dataset from these frames. The frame selection process consists of three stages. The first stage involves global frame selection using perceptual hashing (pHash), which allows filtering out images that are generally similar to those already present in the dataset. In the second stage, frame selection is performed using car detection and image analysis, resulting in the extraction of frames containing new cars compared to those already in the dataset. Car detection in the images is performed using a YOLO convolutional neural network. The uniqueness of detected objects is determined using the Hungarian algorithm, taking into account the Intersection over Union (IoU) metric between bounding boxes of new and existing dataset images. The uniqueness of a car image is also determined by comparing the color values of car regions in the new image with those in the dataset. The third stage involves local selection of car regions using perceptual hashing, which results in frames containing only objects that are new to the dataset. During experimental testing, the developed system operated for 1.5 hours and captured 993 images, from which 121 unique frames (~12% of the initial set) were obtained after selection. The first stage provided the most significant data reduction – over 84%. Car detection was performed using a pre-trained YOLOv8m model (medium size), which ensures high detection accuracy. The resulting specialized image dataset was then used for fine-tuning faster YOLOv8n (nano size) model, which requires fewer computational resources. The final dataset is characterized by high diversity and suitability for further use in the fine-tuning of artificial neural networks.

Keywords: image selection, dataset, car detection, convolutional neural networks, cloud technologies, software.

Стаття надійшла до редакції / Received 16.10.2025

Прийнята до друку / Accepted 15.11.2025

Постановка проблеми

Створення якісного датасету (набору даних) є критичним етапом при навчанні та донавчанні штучних нейронних мереж (ШНМ), які застосовуються для вирішення задач комп'ютерного зору, зокрема, для детектування зображень автомобілів. Важливо, щоб зображення у датасеті були унікальними, представляли різні дорожні ситуації, погодні умови, освітлення, трафік та ракурси об'єктів. Проте, створення набору даних значного розміру в ручному режимі є надто трудомістким. Автоматизоване отримання знімків дорожнього руху з відеокамер забезпечує високу швидкість генерації датасету, але при цьому створює ризик накопичення подібних кадрів. Тобто, якщо датасет містить кілька кадрів із однаковими автомобілями, то це може призвести до надлишкового збільшення розміру датасету, зменшення швидкості навчання моделі ШНМ та до зниження її здатності до узагальнення. Зокрема, при отриманні знімків дорожнього руху із відеокамер з інтервалом у 5 секунд у багатьох випадках отримані кадри можуть бути майже ідентичними, оскільки положення

транспортних засобів та інших об'єктів часто залишаються без змін. У результаті датасет насичується «дублями» зображень, які не несуть нової інформації для нейронної мережі. Тому існує потреба у розробці алгоритму та програмного забезпечення для автоматизованого створення спеціалізованого датасету для зображень автомобілів, які б не містили значну кількість подібних зображень. Відбір (селекцію) зображень у новий датасет потрібно виконувати з урахуванням подібності різних кадрів відеопотоку.

Аналіз досліджень та публікацій

На даний час у сферах комп'ютерного зору та машинного навчання питання автоматичного формування високоякісних, різноманітних та некорельованих (відносно дублювання) датасетів зображень набуває все більшої уваги. Автори нижчеописаних статей аналізують сучасні методи дедуплікації та пропонують алгоритми для вирішення проблеми селекції унікальних зображень, які застосовуються для якісного навчання штучних нейронних мереж.

У роботі [1] автори розглянули вплив наявності подібних (майже дубльованих) зображень на ефективність навчання ШНМ. Запропоновано виконувати хешування зображень та адаптувати поріг дедуплікації так, щоб зменшити об'єм датасету без суттєвої втрати точності навчання ШНМ. Зокрема, показано, що можна скоротити датасет до $\approx 23\%$ без падіння точності, або навіть до 75% з допустимим зниженням точності на 5% . Це підтверджує, що навіть на етапі побудови датасету важливо ретельно балансувати між видаленням надлишкових зображень і збереженням інформаційного насичення.

Автори дослідження [2] аналізували великомасштабні набори зображень (на прикладі завдання картографування) і виявили значну кількість точних дублікатів між навчальною та тестовою вибірками. Застосовано техніки перцептивного хешування для ефективною дедуплікації та виявлення витоків даних з одного піднабору до іншого. Проведений аналіз показав, що навіть у великих репозиторіях часто більше ніж 90% зображень можуть бути дубльованими або повторюватися у різних піднаборах. Це підкреслює практичну важливість вбудованих перевірок на унікальність зображень ще на етапі формування датасету.

Збірник [3] присвячений огляду сучасних методів дедуплікації зображень, серед яких виділено хешування, кластеризацію, комбіновані підходи, порогове порівняння хешів та інші. Роботи збірника обґрунтовують важливість балансу між швидкістю та точністю формування датасету, розглядають обчислювальні витрати та вплив порогів селекції на втрату інформації. Описані дослідження є корисними як теоретична основа для вибору методів селекції зображень (наприклад, pHash, aHash, dHash чи гібридні підходи).

У статті [4] показано, як дублікати (зокрема, подібні копії або "near-duplicates") у навчальному датасеті зображень впливають на час та точність навчання згорткових нейронних мереж. Показано, що агресивна дедуплікація може зменшити дані до 75% при збереженні прийнятної точності, але надмірне видалення може призвести до втрати представлення для рідкісних випадків.

Формування нових датасетів та селекцію зображень можливо виконувати за допомогою інструменту Roboflow [5], який також призначений для керування, анотації та обробки датасетів для задач комп'ютерного зору. Функція створення версій датасету (Dataset Versioning) дозволяє завантажувати лише ті зображення, які пройшли внутрішні кроки селекції. У результаті цього створюються миттєві (snapshot) версії датасету для подальшого навчання ШНМ. Отримані датасети перевіряються на наявність дублікатів. За допомогою препроцесингу і аугментації відібраних даних додатково збільшується їх варіативність без порушення унікальності. Однак при використанні Roboflow виникають деякі складнощі. Зокрема, генерація нової версії датасету іноді може тривати доволі значний час, особливо якщо обробляється багато зображень або вже існує багато версій у проєкті. Тому для практичного застосування важливою є швидкість селекції зображень при створенні унікального датасету. Трапляються помилки при створенні версій чи експорту датасету, особливо коли зображення чи анотації не додані до датасету коректно. У безкоштовних планах у Roboflow можуть бути обмеження щодо розміщення приватних проєктів або обсягу даних. У загальному Roboflow є ефективним інструментом для селекції унікальних зображень, особливо якщо відсутні жорсткі обмеження на час та точність формування датасету. Крім цього, Roboflow є універсальним інструментом для вирішення задач комп'ютерного зору, який не враховує специфіки обробки зображень автомобілів.

Дослідження [6-8] показують перспективність застосування згорткових нейронних мереж з архітектурою YOLO для детекції зображень різноманітних об'єктів, зокрема, автомобілів. У цією метою застосовуються моделі різних версій, наприклад модель YOLOv8. До переваг ШНМ YOLO належить висока точність детекції та висока швидкість, яка є достатньою для обробки кадрів відеопотоку в реальному часі.

Аналіз проведених досліджень показує потребу у розробці власного програмного забезпечення для формування спеціалізованого набору зображень автомобілів. Створений датасет може ефективно застосовуватися для навчання ШНМ, наприклад нейронних мереж з архітектурою YOLO [9]. Такі ШНМ призначені для вирішення прикладних завдань детекції автомобілів у різноманітних дорожніх умовах [10]. У мету уникнення дублювання зображень у датасеті доцільно застосовувати методи відбору зображень, засновані на їх подібності, зокрема, методи перцептивного хешування.

Метою роботи є розробка алгоритму та програми для автоматизованого створення спеціалізованого датасету для зображень автомобілів, який формується шляхом селекції початкових зображень з використанням перцептивного хешування і призначений для навчання та донавчання штучних нейронних мереж.

1. Алгоритм створення унікального датасету для зображень автомобілів

Розроблено алгоритм для селекції початкових зображень та створення унікального датасету зображень автомобілів, який призначений для навчання штучної нейронної мережі (рис. 1). Початкові кольорові

зображення I_{RGB} дорожнього руху отримувалися з веб-камери із заданим інтервалом t_R (наприклад, $t_R = 5$ с) як кадри відеопотоку. У відеопотоці багато кадрів є подібними, тому початкові зображення додаються до датасету тільки після проходження багатоетапної перевірки на унікальність.

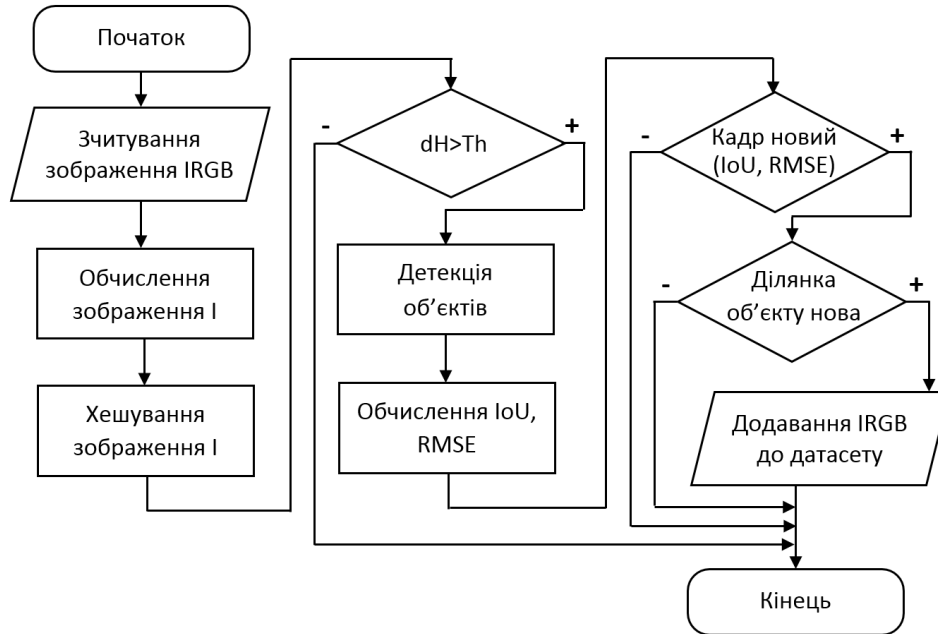


Рис. 1. Схема алгоритму створення спеціалізованого датасету для зображень автомобілів

Згідно з алгоритмом кольорове зображення I_{RGB} (розміром $M_0 \times N_0$ пікселів) перетворюється у зображення у відтінках сірого I , яке масштабується до розміру $N \times N$ пікселів. Після цього виконується перцептивне хешування зображення I і його перевірка на унікальність. На першому етапі зображення вважається новим, якщо різниця d_H його хешу та хешу зображення з датасету перевищує поріг T_h . На другому етапі перевірки виконується детекція автомобілів на зображенні за допомогою ШНМ YOLO. Для рамок детектованих автомобілів на початковому зображенні на зображенні датасету обчислюється метрика IoU їх перекривання та корінь середньої квадратичної похибки (RMSE). Кадр вважається новим, якщо значення IoU менше заданого порогу або значення RMSE перевищує поріг. На третьому етапі зображення вважається новим, якщо його ділянки автомобілів є новими з урахуванням їх хешів. У випадку успішного проходження всіх перевірок зображення I_{RGB} додається до датасету.

Результатом роботи алгоритму є унікальні зображення, на яких положення детектованих автомобілів вказується прямокутними рамками. Виявлення об'єктів на зображеннях виконується попередньо навченою моделлю №1 ШНМ yolov8m (середнього розміру), оскільки така модель забезпечує високу точність детекції об'єктів навіть у складних сценах. Окремі зображення датасету, за потреби, коректуються в ручному режимі. Отриманий спеціалізований датасет використовується для донавчання (fine-tuning) моделі №2 ШНМ yolov8n (розміру нано). У результаті після донавчання модель ШНМ малого розміру повинна не тільки забезпечувати точність на рівні моделі великого розміру (при детектуванні автомобілів), але й споживати менше обчислювальних ресурсів і демонструвати при цьому вищу швидкість. Селекція зображень зменшує розмір датасету, що сприяє зменшенню часу навчання ШНМ малого розміру. Такий датасет зображень з детектованими автомобілями може використовуватися для навчання моделей ШНМ YOLO різних версій та розмірів за допомогою бібліотеки Ultralytics. Таким чином, в значній мірі автоматизується не тільки процес зчитування зображень з відеокамери та формування датасету, але й донавчання ШНМ.

2. Глобальна селекція кадрів з використанням перцептивного хешування

Розглянемо детальніше етап №1 відбору кадрів. На реальних зображеннях дорожнього руху, отриманих з відеокамери, багато кадрів є подібними і містять зображення тих самих автомобілів. Тому з метою усунення надлишковості даних виконується глобальна селекція зображень з використанням перцептивного хешування (perceptual hashing – pHash) [11-12]. На відміну від криптографічних хешів (SHA-256, MD5), перцептивні хеші нечутливі до незначних змін освітлення, масштабу, шумів або компресії зображень. Перцептивні хеші мають суттєву перевагу при виявленні унікальних зображень у порівнянні з безпосереднім порівнянням зображень, оскільки зміни освітлення та шуми призводять до певної різниці яскравості для різних зображень навіть тих самих об'єктів [13, 14]. У той же час перцептивні хеші будуть змінюватися при суттєвих змінах на зображеннях, зокрема, при переміщенні автомобілів. Крім цього, розмір хешу значно менший за розмір зображення, тому використання хешів підвищує швидкість селекції зображень.

Глобальна селекція зображень за допомогою перцептивного хешування полягає в наступному. Для кожного нового зображення обчислюється його хеш (глобальний для всього зображення) і порівнюється з хешами зображень, які вже містяться в датасеті. До датасету додаються тільки зображення з унікальними хешами, тобто зі значними змінами відносно зображень датасету (наприклад, зображення з новим автомобілем).

Перцептивне хешування pHash [11, 12] базується на двовимірному дискретному косинусному

перетворенні (Discrete Cosine Transform – DCT). Перед таким перетворенням початкове кольорове зображення перетворюється у відтінки сірого і масштабується до розміру $N \times N$ пікселів, у результаті чого отримується зображення $I(x, y)$, де x, y – координати пікселів зображення за шириною та висотою відповідно. Двовимірне DCT виконується над зображенням $I(x, y)$ за формулою:

$$C(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I(x, y) \cos\left(\frac{\pi(2x+1)u}{2N}\right) \cos\left(\frac{\pi(2y+1)v}{2N}\right), \quad (1)$$

де $C(u, v)$ – частотна матриця; u, v – значення частот за шириною та висотою відповідно;

$$\alpha(u) = \sqrt{\frac{1}{N}} \text{ для } u = 0; \alpha(u) = \sqrt{\frac{2}{N}} \text{ для } u > 0.$$

В матриці $C(u, v)$ аналізується низькочастотний блок розміром 8×8 елементів, який відображає загальну структуру сцени. Кожен коефіцієнт низьких частот $C(u, v)$ порівнюється з медіанним значенням блоку, у результаті чого формується бінарний хеш $h \in \{0, 1\}^{64}$. Таким чином, хеш зберігає загальну структуру зображення, а не абсолютні значення яскравості пікселів. Для порівняння двох зображень використовується різниця між 64-бітними хешами h_1 та h_2 зображень, яка обчислюється як відстань Хеммінга (Hamming distance):

$$d_H(h_1, h_2) = \sum_{i=1}^{64} 1[h_1(i) \neq h_2(i)]. \quad (2)$$

де i – номер елемента хеша; $d_H = 0$ означає ідентичні кадри;

$d_H = 64$ означає повністю різні зображення.

Ключовим параметром селекції зображень є поріг відстані Хеммінга T_h , який визначає подібність двох зображень. Нове зображення з хешом h_{new} вважається відмінним від зображення в датасеті з хешем h_{stored} , якщо

$$d_H(h_{new}, h_{stored}) > T_h. \quad (3)$$

Якщо умова (3) виконується для всіх зображень датасету, то нове зображення успішно проходить глобальну селекцію з використанням перцептивного хешування. У практичній реалізації поріг T_h обрано експериментально шляхом перебору його значень. Для кожного значення T_h визначалася кількість кадрів, які відкидалися як дублікати або оцінювалися як унікальні. Найкращий результат отримано для порогу $T_h = 2$, при якому кадри з однаковою сценою та автомобілями відкидаються, а унікальні кадри з новими або зміщеними автомобілями оцінюються як унікальні.

У процесі глобальної селекції кожне нове зображення перетворювалося в хеш та порівнювалося з уже збереженими хешами зображень у базі даних Redis. Використання Redis забезпечує швидкий доступ до даних та масштабованість, оскільки в такому випадку дані зберігаються в оперативній пам'яті. Також, якщо датасет вже має велику кількість збережених унікальних хешів, то такі кеші можна аналізувати по частинам у межах ковзного вікна порівняння (sliding window), наприклад, розміром 200 кешів. Після швидкодіючої глобальної селекції з використанням перцептивного хешування, під час якої відкидається більшість кадрів-дублікатів, відібрані зображення проходять додаткові перевірки на унікальність. Інші етапи селекції зображень потребують більше обчислювальних ресурсів (порівняно з етапом №1), тому застосування глобальної селекції забезпечує високу швидкодію всього процесу відбору унікальних зображень у датасет.

3. Селекція кадрів із використанням детектування та аналізу зображень автомобілів

Етап №2 відбору зображень, який полягає в детектуванні та аналізі автомобілів на зображеннях, дає змогу залишити тільки ті зображення, на яких спостерігаються зміни досліджуваних об'єктів (автомобілів), а не фону. Така перевірка потрібна, оскільки частина зображень (які пройшли глобальну селекцію), відрізняються від зображень у датасеті тільки фоном, що може бути зумовлено значними змінами освітлення, погодними умовами, рухом сторонніх об'єктів (не автомобілів). Етап №2 відбору зображень містить такі кроки:

1. Детекція автомобілів на зображенні за допомогою моделі ШНМ YOLO, а результати чого обчислюються координати прямокутних обмежувальних рамок (bounding box) автомобілів.

2. Порівняння координат рамок автомобілів на новому зображенні з координатами рамок автомобілів для зображень, які містяться в датасеті, за допомогою угорського алгоритму та з урахуванням метрики IoU (Intersection over Union – Перетин над Об'єднанням) перекривання рамок.

3. Порівняння значень кольору ділянок автомобілів на новому зображенні зі значеннями кольору ділянок автомобілів для зображень, які містяться в датасеті; зображення автомобіля в межах рамки вважається новим, корінь середньої квадратичної похибки (Root Mean Square Error – RMSE) значень кольору двох ділянок перевищує заданий поріг T_{RMSE} . Така перевірка дозволяє виявити на кадрах нові автомобілі, навіть якщо їхні рамки просторово збігаються з рамками автомобілів для зображень датасету.

Кадр вважається унікальним і переходить на наступний етап селекції, якщо більш ніж 50% рамок автомобілів на зображенні є унікальними (за їх координатами або за зображеннями автомобілів у межах рамок).

Угорський алгоритм, також відомий як алгоритм Куна-Мункреса, є ефективним для вирішення задачі оптимального співставлення в двочастковому графі [15]. У контексті порівняння зображень, цей алгоритм використовується для асоціації виявлених об'єктів між двома кадрами, мінімізуючи або максимізуючи певну метрику подібності, таку як відстань або коефіцієнт перетину. Метрика IoU використовується для оцінки точності просторового детектування об'єктів і обчислюється як відношення площі перетину двох обмежувальних рамок "Area of Overlap" до площі їх об'єднання "Area of Union" [16]:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}. \quad (4)$$

Значення IoU змінюється від 0 до 1, де 1 означає повне перекривання рамок.

У задачі порівняння нового зображення з попереднім кадром значення IoU використовується для оцінки подібності між просторовими рамками об'єктів. Угорський алгоритм застосовується для оптимального

співставлення рамок об'єктів між двома кадрами, мінімізуючи їх сумарне значення IoU. Якщо сумарне значення IoU двох кадрів перевищує певний поріг T_{IoU} , то об'єкти на ньому вважаються ідентичними, а такий кадр не вважається новим. Експериментальним методом встановлено значення порогу $T_{IoU} = 0.5$. Унікальними вважаються кадри, для яких сумарне значення IoU не перевищує поріг T_{IoU} .

В іншому випадку виконується перевірка співпадіння значень кольору ділянок зображень об'єктів (розміром $M_b \times N_b$ пікселів) за критерієм RMSE, який обчислюється за формулою:

$$RMSE = \sqrt{\frac{1}{M_b \cdot N_b \cdot Q_C} \sum_{i=0}^{M_b-1} \sum_{k=0}^{N_b-1} \sum_{c=0}^{Q_C-1} (I(i, k, c) - I_D(i, k, c))^2}, \quad (5)$$

де $I(i, k, c)$ – значення каналу кольору c для пікселя нового зображення з координатами (i, k) за висотою та шириною відповідно;

$I_D(i, k, c)$ – значення каналу кольору c для пікселя зображення датасету з координатами (i, k) ;

Q_C – кількість каналів кольору (RGB – червоний, зелений, синій).

Якщо RMSE між рамками двох автомобілів перевищує визначений поріг T_{RMSE} , та зображення автомобіля в рамці вважається новий (не дублікатом). Запропонований етап №2 відбору зображень дозволяє точно виявити кадри з новими досліджуваними об'єктами (автомобілями), що продемонстровано на Рисунку 2.

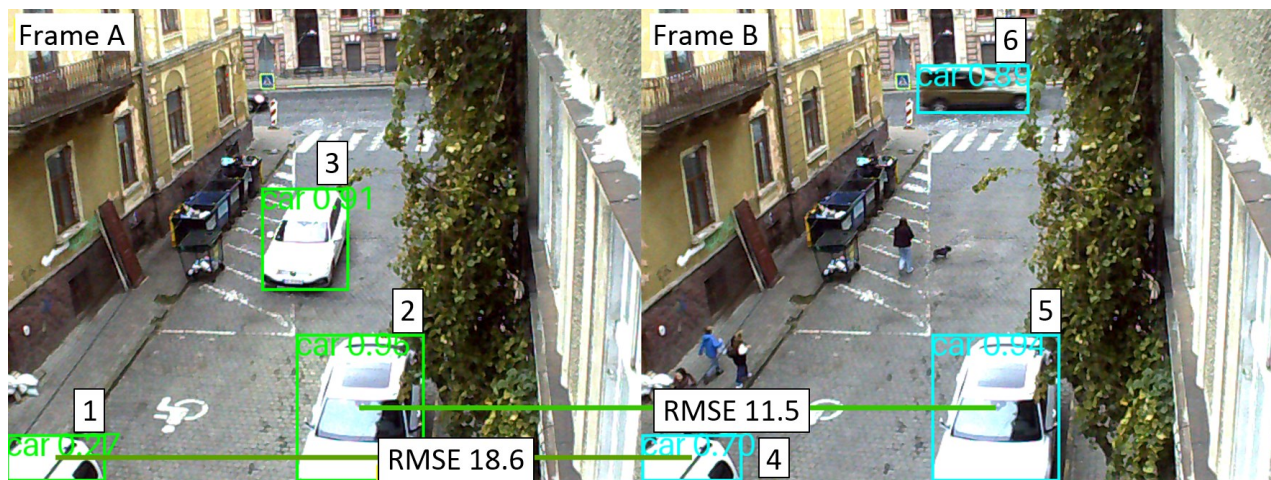


Рис. 2. Приклад селекції кадрів із використанням детекції та аналізу зображень автомобілів

На рис. 2 лівий кадр А належить до датасету, а правий кадр В є новим (проходить перевірку на унікальність). На лівому кадрі модель YOLO детектувала 3 автомобілі (№ 1-№ 3), а на правому кадрі також детектовано 3 автомобілі (№ 4- № 6). У межах обмежувальної рамки кожного автомобіля вказано ймовірність його розпізнавання. За угорським алгоритмом перевірено перекривання рамок лівого та правого зображень, у результаті чого отримано значне просторове перекривання рамок: № 1 з № 4, та № 2 з № 5. Для таких пар рамок обчислено їх RMSE (5). Експериментальним шляхом обрано значення порогу $T_{RMSE} = 30$. Для двох нерухомих автомобілів (рамки № 4 та № 5) отримано значення RMSE 18.6 та 11.5 відповідно (менше за поріг T_{RMSE}), тому такі рамки не вважаються новими. Проте, на правому кадрі є унікальний автомобіль (рамка № 6). Новий кадр проходить перевірку на унікальність на етапі № 2 у тому випадку, якщо відсоток унікальних рамок (з урахуванням їх перекривання та кольору ділянок об'єктів) перевищує поріг T_S (наприклад, $T_S = 10\%$). Розглянутий правий кадр (рис. 2) містить 1/3 унікальних рамок, тому такий кадр успішно проходить перевірку і надсилається на наступний крок селекції.

4. Локальна селекція ділянок автомобілів із використанням перцептивного хешування

На етапі № 3 локальної селекції ділянок автомобілів (обмежених рамками) виконується порівняння детектованих ділянок автомобілів на новому зображенні з ділянками автомобілів для зображень датасету (збережених у базі даних). Така перевірка виконується шляхом порівняння хешів ділянок зображень аналогічно як для етапу №1, але на етапі №3 порівнюються хеші не всього зображення, а тільки локальних ділянок. Новий кадр вважається унікальним, якщо відсоток його унікальних рамок перевищує поріг T_h . Кадри, які пройшли всі етапи селекції №1-3, вважаються повністю унікальними і додаються до датасету.

Перевірка на рівні ділянок автомобілів дозволяє відсіювати повторювані сцени, але водночас не втрачати цінні дані у випадках, коли змінюється склад транспортних засобів на зображенні. Це суттєво підвищує точність відбору унікальних кадрів у порівнянні з підходами, що спираються лише на міжкадрову подібність. Унікальні зображення та координати рамок їх автомобілів зберігаються в датасеті у форматі, який підтримує навчання моделей YOLO. Таким чином формується спеціалізований датасет зображень автомобілів.

Завдяки збереженню кешів у базі даних Redis забезпечується висока швидкість роботи системи, оскільки порівняння навіть тисячі хешів виконується менш ніж за секунду. Повна перевірка унікальності зображення займає менше ніж 5 секунд, при цьому значна частина цього часу припадає на етап селекції №2 (під час якого виконується детекція зображень об'єктів моделями YOLO).

5. Апаратно-програмне забезпечення системи для автоматичного створення датасету

Розроблена система для автоматичного створення датасету зображень автомобілів функціонує на апаратній платформі одноплатного мікрокомп'ютера Raspberry Pi 5, який виконує роль локального обчислювального вузла. Програмне забезпечення вузла виконувало зчитування кадрів з веб-камери, початкову обробку кадрів, перевірку їх на унікальність, збереження унікальних кадрів та їх хешів у датасет. Вибір Raspberry Pi 5 обумовлено комбінацією ряду факторів: його енергоефективністю, низькою вартістю, компактністю та можливістю запускати повноцінне Linux-середовище з підтримкою Docker. На практиці Raspberry Pi використовувався як самостійний вузол, розташований безпосередньо біля відеокамери на фасаді будинку. Камера під'єднувалася до Raspberry Pi по шині USB і налаштовувалася на зчитування кадрів певної роздільної здатності (наприклад, 640×480 пікселів).

Модель веб-камери обрано з урахуванням стабільності потокового захоплення, фіксації фокусу та можливості відключити автоматичні режими (автоекспозиція, автобаланс білого). Отримання відеопотоку без використання автоматичних режимів відеокамери значно зменшує флуктуації яскравості та кольору кадрів, які негативно впливають на точність селекції з використанням перцептивних хешів.

Мікрокомп'ютер Raspberry Pi 5 забезпечує значну обчислювальну потужність, проте при роботі з моделями YOLO великого розміру демонструє меншу швидкодію у порівнянні зі потужними десктопними комп'ютерами. Саме тому в розробленій програмі використовується етап №1 глобальної селекції зображень, на якому відкидається значна кількість зображень-дублів. Завдяки цьому зменшується кількість перевірок на етапі №2, а відповідно мінімізується кількість запитів до моделі YOLO. Застосування мікрокомп'ютера Raspberry Pi5 як апаратної платформи системи створення датасету зображень також є вигідним економічно з урахуванням низької вартості та енергоспоживання мікрокомп'ютера.

Програму для зчитування зображень з відеокамери, селекції унікальних зображень, детекції зображень об'єктів за допомогою ШНМ YOLO та створення датасету розроблено на мові Python із використанням хмарних технологій. В програмі використано ряд бібліотек, зокрема opencv, ultralytics, numpy, matplotlib та ін. Реалізацію моделей YOLO виконано з використанням бібліотеки Ultralytics [17], яка забезпечує широкий вибір параметрів та гіперпараметрів навчання нейромережі. У програмі для детектування зображень автомобілів використано модель YOLO середнього розміру (medium), а саме модель YOLOv8m. Для перцептивного хешування (pHash) зображень та їх фрагментів (ділянок автомобілів) використано бібліотеку imagehash. Для використання серверу Redis з базою даних використано бібліотеку redis. Сервер Redis розміщено в Docker-контейнері на мікрокомп'ютері Raspberry Pi. Використання контейнера спростило розгортання сервісу, гарантувало ізоляцію середовища та дозволило зручно керувати режимом збереження даних.

Кожен із перелічених компонентів виконує чітко визначену роль у конвеєрі обробки даних; сумісні версії забезпечували відтворюваність і зменшували ризик несумісностей під час експерименту. Перелічені програмні компоненти знаходяться у відкритому доступі і є безкоштовними для власного та наукового використання.

6. Тестування розробленої системи для автоматичного створення датасету

Проведено експериментальну перевірку розробленої системи для автоматичного створення датасету при обробці тестових зображень. Точність функціонування системи визначалася правильністю селекції зображень, які розділялися на кадри-дублікати та унікальні кадри, які містять нові транспортні засоби. Селекція зображень складалася з трьох послідовних етапів, на кожному з поступово зменшується обсяг даних. У ході експериментальної перевірки тривалістю 1,5 години було зчитано 993 зображення з інтервалом у 5 секунд, при цьому для кожного етапу отримано такі результати (табл. 1).

Таблиця 1

Результати багатоетапної селекції кадрів

№ етапу	Назва етапу перевірки	Кількість зображень після етапу	Частка від початкової вибірки	Результати обробки
0	Загальний обсяг даних (усі зчитані кадри)	993	100 %	Усі зображення, зчитані з відеокамери
1	Глобальна селекція кадрів з використанням перцептивного хешування	150	15,1 %	Відкинута кадри, які на рівні всього зображення незначно відрізняються від зображень датасету
2	Селекція кадрів із використанням детектування та аналізу зображень автомобілів	131	13,2 %	Вилучені кадри, які містять нових автомобілів порівняно з кадрами датасету
3	Локальна селекція ділянок автомобілів із використанням перцептивного хешування	121	12,2 %	Відібрано лише ті кадри, в яких містять нові для датасету транспортні засоби

Як видно з Таблиці 1, найсуттєвіше скорочення відбулося вже на етапі №1 (глобальна селекція з

хешуванням), на якому відкинуто понад 84% кадрів як надмірно схожих. Селекція кадрів із використанням детектування та аналізу зображень автомобілів (етап №2) скорочує набір зображень ще приблизно на 12,7 %, що дозволяє уникнути повторів через варіації фону зображень. Етап №3 забезпечує додаткову точність, зменшивши набір до 121 унікального кадру (~12 % від початкових даних). Це гарантує, що в датасет потрапляють лише справді нові приклади транспортних засобів. Таким чином, система ефективно зменшує надлишковість у даних, зберігаючи при цьому найбільш інформативні приклади для формування унікального датасету. Згідно з отриманими даними, найбільший ефект у зменшенні обсягу даних забезпечує перцептивне хешування (етап №1), яке зберігає лише потенційно нові сцени. Це пояснюється тим, що за час між зчитуванням фотографій (5 с) у кадрі не відбувається значної зміни автомобілів.

Багатоетапний підхід до селекції зображень продемонстрував високу ефективність: із майже тисячі зчитаних кадрів сформовано компактний набір із приблизно однієї восьмої від початкового обсягу, який зберігає при цьому максимальне різноманіття автомобілів. Це підтверджує доцільність використання комбінованої перевірки (з використанням перцептивного хешування, детектування та аналізу зображень автомобілів) для створення якісних унікальних датасетів у задачах комп'ютерного зору. Це підтверджує вибірка 10 випадкових зображень автомобілів з датасету, створеного за допомогою розробленої програми (рис. 3). На всіх отриманих кадрах спостерігаються різні автомобілі, тобто відсутні дублікати. Отриманий датасет може ефективно використовуватися для навчання ШНМ. На основі розробленого датасету виконано трансферне навчання нейромережі YOLO малого розміру, тобто попередньо навчені ваги моделей були уточнені на розробленому датасеті. Це дозволило значно прискорити конвергенцію моделей та досягти високої точності детекції об'єктів.

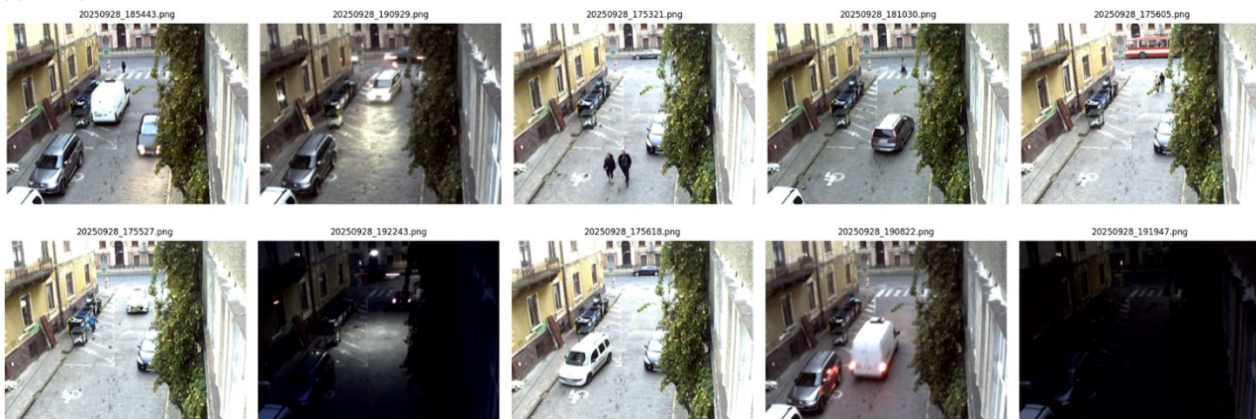


Рис. 3. Приклад кадрів зі створеного датасету

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

Розроблено алгоритм та програмне забезпечення комп'ютерної системи для автоматизованого створення спеціалізованого датасету зображень автомобілів. Систему реалізовано на базі мікрокомп'ютера Raspberry Pi 5, а початкові зображення зчитувалися з відеокамери. Програмне забезпечення розроблено на мові Python із використанням хмарних технологій. З метою видалення кадрів-дублікатів виконано селекцію початкових кадрів, яка складається з 3 етапів. На першому етапі виконується глобальна селекція кадрів з використанням перцептивного хешування (pHash), у результаті чого відкидаються всі зображення, які в цілому незначно відрізняються від зображень датасету. Другий етап полягає у селекції кадрів із використанням детекції та аналізу зображень автомобілів, а на виході етапу отримуються тільки кадри, які містять нові автомобілі порівняно з кадрами датасету. Для цього виконується детекція автомобілів на зображеннях за допомогою моделі ШНМ YOLO, а детектовані об'єкти виділяються прямокутними обмежувальними рамками. За допомогою угорського алгоритму та з урахуванням метрики перекривання рамок IoU виконується порівняння координат рамок автомобілів на новому зображенні з координатами рамок автомобілів датасету. Унікальність зображення також визначається шляхом порівняння значень кольору ділянок автомобілів на новому зображенні зі значеннями кольору ділянок для зображень датасету, а різниця ділянок зображень автомобілів описується параметром RMSE. Кадр вважається унікальним і переходить на наступний етап селекції, якщо 10% рамок автомобілів на зображенні є унікальними за їх положенням або за зображеннями ділянок автомобілів. На третьому етапі виконується локальна селекція ділянок автомобілів із використанням перцептивного хешування, у результаті чого отримуються лише ті кадри, в яких містяться нові для датасету транспортні засоби.

Проведене дослідження засвідчило ефективність багатоетапного відбору зображень для формування унікального датасету автомобілів. Такий підхід дозволяє досягти значного скорочення надлишкових даних: із 993 кадрів, отриманих протягом 1,5 години експерименту, було відібрано лише 121 унікальний кадр (~12 % від початкового обсягу). Найбільший внесок у зменшення даних забезпечило використання перцептивного хешування на першому етапі, яке дозволило видалити понад 84% майже ідентичних кадрів. Подальші етапи перевірки дозволили уникнути повторів, пов'язаних із незначними змінами складу транспортних засобів чи змінами фону. Сформований набір кадрів характеризується високою різноманітністю та придатністю для подальшого використання у задачах донавчання ШНМ. У розробленій системі детектування автомобілів на зображеннях виконується попередньо навченою моделлю YOLOv8m (середнього розміру), яка забезпечує

високу точність детекції об'єктів. Отриманий спеціалізований датасет зображень застосовано для донавчання (fine-tuning) більш швидких моделей YOLOv8n, функціонування яких потребує менше обчислювальних ресурсів. Такі результати демонструють, що запропонована система є ефективним інструментом для автоматизованого створення унікальних датасетів і може бути використана як основа для більш масштабних досліджень у сфері інтелектуальних транспортних систем, моніторингу дорожнього руху та автономного водіння.

Література

1. Li X. CE-Dedup: Cost-Effective Convolutional Neural Nets Training based on Image Deduplication / X. Li, L.Chang, X. Liu // arXiv preprint. – 2021. – <https://arxiv.org/abs/2109.00899>.
2. Adimoolam Y. K. Efficient Deduplication and Leakage Detection in Large Scale Image Datasets with a focus on the CrowdAI Mapping Challenge Dataset / Y. K. Adimoolam, B. Chatterjee, C. Poullis, M. Averkiou // arXiv preprint. – 2023. – <https://arxiv.org/abs/2304.02296>.
3. Shah R. A Comprehensive Study of Image Deduplication Techniques: Addressing Challenges and Advancements in Storage Optimization / R. Shah, A. K. Shrivastava // *Advances in Health Informatics, Intelligent Systems, and Networking Technologies* / ed. A. Jeyabose, V. E. Balas, S. L. Fernandes. – Springer, 2025. – P. 275–289. – doi: https://doi.org/10.1007/978-981-96-4008-9_21.
4. Aghabagherloo A. Impact of Data Duplication on Deep Neural Network-Based Image Classifiers: Robust vs. Standard Models / A. Aghabagherloo, A. Abadi, S. Sarkar, V. A. Dasu, B. Preneel // arXiv preprint. – 2025. – <https://arxiv.org/html/2504.00638v2>.
5. Roboflow. Roboflow Annotation Tool. URL: <https://roboflow.com> (date of access: 10.03.2025).
6. Alahdal N. M. Real-time Object Detection in Autonomous Vehicles with YOLO / N. M. Alahdal, F. Abukhodair, L.H. Meftah, A. Cherif // *Procedia Computer Science*. – 2024. – Vol. 246. – P. 2792–2801. – <https://www.sciencedirect.com/science/article/pii/S1877050924024293>.
7. Li X. Vehicle–Pedestrian Detection Method Based on Improved YOLOv8 / X. Li, Y. Wang, H. Zhang, J. Liu // *Electronics*. – 2024. – Vol. 13, № 11. – Article № 2149. – <https://www.mdpi.com/2079-9292/13/11/2149>.
8. Tomas J. P. Q. YOLOv8-Based Vehicle Type Detection: An In-Depth Exploration of Deep Learning Techniques for Robust Analysis of Dashcam Footage / J. P. Q. Tomas, A. Aquino, G. E. M. David, R. D. Del Ayre, J. K. L. Alminiana // *Proceedings of the 2024 9th International Conference on Information and Communication Technology*. – 2024. – P. 91–96. – doi: <https://dl.acm.org/doi/10.1145/3670085.3670098>.
9. Balovsyak S. Car Image Recognition using Convolutional Neural Network with EfficientNet Architecture / S. Balovsyak, O. Kroitor, Kh. Odaiska, A.-B. M. Salem, S. Stets // *IntellITSIS 2024: 5th International Workshop on Intelligent Information Technologies & Systems of Information Security (Khmelnyskyi)*. – CEUR Workshop Proceedings / ed. T. Hovorushchenko, O. Savenko, P. T. Popov, S. Lysenko. – 2024. – P. 182–195.
10. Stelmakh O. P. Information Technology of Video Data Processing for Traffic Intensity Monitoring / O. P. Stelmakh, I. V. Stetsenko, D. V. Velyhotskyi // *Control Systems & Computers*. – 2020. – № 3. – S. 50–59. – doi: <https://doi.org/10.15407/usim.2020.03.050>.
11. Alkhowaiter M. Evaluating Perceptual Hashing Algorithms in Detecting Image Manipulation Over Social Media Platforms / M. Alkhowaiter, K. Almubarak, C. Zou // *Proc. 2022 IEEE International Conference on Cyber Security and Resilience (CSR)*. – Rhodes, Greece, 2022. – P. 149–156. – doi: 10.1109/CSR54599.2022.9850288.
12. Lu C. S. Robust Image Deepfake Detection with Perceptual Hashing / C. S. Lu // *ScitePress*. – 2024. – <https://www.scitepress.org/Papers/2024/123178/123178.pdf>.
13. Balovsyak S. V. Automatic Highly Accurate Estimation of Gaussian Noise Level in Digital Images Using Filtration and Edges Detection Methods / S. V. Balovsyak, Kh. S. Odaiska // *International Journal of Image, Graphics and Signal Processing (IJIGSP)*. – 2017. – Vol. 9, № 12. – P. 1–11. – doi: 10.5815/ijigsp.2017.12.01.
14. Balovsyak S. Adjusting the Brightness and Contrast Parameters of Digital Video Cameras Using Artificial Neural Networks / S. Balovsyak, Kh. Odaiska, O. Yakovenko, I. Iakovlieva // *Proc. SPIE: Sixteenth International Conference*. – 2024. – Vol. 12938. – P. 129380I-1–129380I-4. – doi: 10.1117/12.3009429.
15. Kuhn H. W. The Hungarian Method for the Assignment Problem / H. W. Kuhn // *Naval Research Logistics Quarterly*. – 1955. – doi: <https://doi.org/10.1002/nav.3800020109>.
16. Rezatofighi H. Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression / H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, S. Savarese // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2019. – P. 658–666.
17. Ultralytics. YOLOv8 Documentation. URL: <https://docs.ultralytics.com> (date of access: 11.03.2025).