

## Розпізнавання об'єктів на фото та відео

Вінницький національний технічний університет

### Анотація

*Розглянуто сучасні методи та підходи до розпізнавання об'єктів на фото та відео. Описано різноманітні алгоритми як ненейромережні алгоритми (каркас Viola-Jones, SIFT та метод ознак гістограми орієнтованих градієнтів) так і методи на основі нейромереж, (CNN на основі регіонів, деформовані згорткові мережі та одноступеневі детектори, такі як SSD, RefineDet та RetinaNet). Визначено, що кожен метод описується з врахуванням його принципу дії, переваг та недоліків.*

**Ключові слова:** розпізнавання об'єктів; обробка зображень; аналіз відео; комп'ютерний зір;

### Abstract

*This study explores contemporary methods and approaches to object recognition in photos and videos. Various algorithms were described, including non-neural network algorithms (the Viola-Jones framework, SIFT, and Histogram of Oriented Gradients (HOG)), as well as neural network-based methods (region-based CNNs, deformable convolutional networks, and single-stage detectors like SSD, RefineDet, and RetinaNet). Each method is elucidated considering its principles, advantages, and drawbacks.*

**Keywords:** object recognition; image processing; video analysis; computer vision;

### Вступ

Наразі технології інформаційних систем, які ґрунтуються на розпізнаванні об'єктів на фото та відео зображеннях, набувають все ширшого застосування. Це пов'язано з їхньою здатністю обробляти та аналізувати візуальні дані, що відкриває безліч нових можливостей у різних сферах життя. Прикладами таких застосувань можуть бути системи біометричної ідентифікації, системи відеоспостереження для забезпечення безпеки, автономні транспортні системи, а також індексація та аналіз зображень та відеоматеріалів у базах даних з врахуванням їх змісту.

### Результати дослідження

Проблема розпізнавання об'єктів на фото та відео полягає у впізнанні та ідентифікації об'єктів на цифрових зображеннях та відеофрагментах за допомогою комп'ютерних алгоритмів та методів обробки зображень. Вона створює виклики через складність обробки великої кількості даних, змінність умов освітлення та перспективи, а також наявність різних об'єктів, що можуть бути частково перекриті або знаходитися у русі. Вирішення цієї проблеми має велике значення у багатьох галузях, включаючи комп'ютерний зір, медичну сферу, автоматизоване виробництво, сфери безпеки, навігації, медіа і розваг.

Розпізнавання образів — це віднесення вихідних даних до певного класу за допомогою виділення істотних ознак, що характеризують ці дані, із загальної маси несуттєвих даних. Кожний образ являє собою набір чисел, що описують його властивості і називаються ознаками. Упорядкований набір ознак об'єкта називається вектором ознак. Вектор ознак — це точка в просторі ознак.

Поняття подібності між образами відноситься до оцінки схожості двох або більше зображень або об'єктів на основі їхніх властивостей, характеристик чи змісту. Уводячи поняття подібності між образами можна поставити задачі розпізнавання [1].

Існує багато методів вимірювання подібності між образами, таких як відстань між точками даних у просторі ознак, коефіцієнти кореляції, статистичні міри схожості, а також використання навчених моделей, які враховують складні взаємозв'язки між даними. Визначення подібності між образами є ключовим для багатьох завдань аналізу даних та машинного навчання. Конкретний вид такої постановки залежить від наступних етапів при розпізнаванні відповідно до вибраного оптимального алгоритму [1].

Загальний процес розпізнавання об'єктів на фотографіях та відео можна описати наступним чином: проводиться збір даних, що включає отримання цифрових зображень або відеопотоку, потім дані піддаються передпроцесуванню для зменшення шуму та нормалізації освітлення. Наступний крок — виявлення об'єктів, де застосовуються алгоритми для локалізації потенційних об'єктів на зображенні чи відеопотоці. Після цього проводиться екстракція ознак, що дозволяє виділити характеристики об'єктів для подальшої аналізу та отримання ознак об'єктів і їх класифікація. На заключному етапі результати розпізнавання візуалізуються та інтерпретуються для подальшого аналізу або використання [2].

Виділяють два основних підходи до виявлення об'єктів у комп'ютерному баченні: ненеуронні методи та методи, засновані на нейронних мережах. Ненеуронні підходи потребують ручного визначення характеристик, які є інформативними для ідентифікації об'єктів. Найпоширеніші методи ненеуронного розпізнавання є каркас Viola-Jones, SIFT та метод ознак гістограми орієнтованих градієнтів [2].

Каркас Viola-Jones — це алгоритм для швидкого та точного виявлення облич у режимі реального часу. Наг-признаки для аналізу зображень визначають різницю між сусідніми регіонами на зображенні, такі як різниця у середньому значенні пікселів. Каскадна архітектура використовує послідовне застосування класифікаторів, які швидко відкидають негативні області зображення. Метод AdaBoost допомагає підвищити ефективність, шляхом навчання послідовної серії класифікаторів, кожен з яких намагається виправити помилки попереднього. Цей алгоритм застосовується в комп'ютерному баченні, зокрема для розпізнавання облич та відстеження людей. Він має переваги у швидкості, точності та простоті, але обмежений у гнучкості та залежний від розміру. Viola-Jones стимулював розробку інших алгоритмів детектування об'єктів і застосовується в різних галузях [3].

Метод SIFT (Scale-Invariant Feature Transform) — це алгоритм виявлення та опису ознак у зображеннях. Він відомий своєю стійкістю до змін масштабу, обертання та освітлення, що робить його популярним у комп'ютерному баченні. SIFT працює за такими основними принципами: спочатку він виявляє ключові точки у зображенні, потім обчислює для кожної ключової точки описувач ознак, заснований на гістограмах градієнтів, та нарешті використовує метод підбору відповідностей для зіставлення ключових точок між двома зображеннями. Переваги SIFT полягають у його стійкості до змін, високій точності та ефективності. Однак алгоритм може бути чутливим до шуму та мати високу обчислювальну складність, що може обмежувати його застосування у деяких випадках [4].

Гістограми орієнтованих градієнтів (HOG) — це метод аналізу зображень, що базується на визначенні градієнтів яскравості та їхніх орієнтацій у малих блоках зображення. Використовуючи гістограми градієнтів, HOG може описати форму та текстуру об'єктів у зображеннях, що робить його ефективним для задач відстеження та детектування об'єктів. Цей метод застосовується у комп'ютерному баченні, зокрема у визначенні пішоходів на дорозі, відстеженні облич та визначенні об'єктів у відео. HOG відомий своєю стійкістю до змін освітлення та контрасту, що робить його популярним у різних областях комп'ютерної науки та індустрії.[4]

Нейронно-мережеві підходи відзначаються можливістю автоматичного вивчення характеристик з даних, що усуває необхідність у ручному проектуванні ознак. Зазвичай ці методи ґрунтуються на згорткових нейронних мережах (CNN), які є особливо ефективними у розпізнаванні та аналізі зображень. Вони володіють здатністю автоматично виявляти важливі особливості об'єктів у великих наборах даних, що робить їх дуже потужними інструментами для завдань комп'ютерного зору та обробки зображень.

Методи CNN на основі регіонів (R-CNN, Fast R-CNN, Faster R-CNN, Cascade R-CNN) відомі своєю здатністю ефективно виявляти та класифікувати області-кандидати на зображенні. Вони спочатку пропонують потенційні області, які містять об'єкти, і після цього класифікують ці області для визначення наявності та класу об'єкта. R-CNN використовує послідовний підхід для кожної області, Fast R-CNN здійснює класифікацію та локалізацію об'єктів більш ефективно, Faster R-CNN додає модуль Region Proposal Network для автоматичного виявлення областей, тим самим прискорюючи процес, а Cascade R-CNN використовує каскади класифікаторів для збільшення швидкодії. Ці методи знаходять широке застосування у відстеженні об'єктів, розпізнаванні облич, аналізі зображень та багатьох інших завданнях комп'ютерного зору [5].

Одноступеневі детектори, такі як SSD (Single Shot Multibox Detector), RefineDet і RetinaNet, є методами, які виконують як пропозицію об'єкта, так і класифікацію за один крок, що робить їх швидшими порівняно з підходами, заснованими на R-CNN. Ці методи пропонують області, що містять об'єкти, та одразу ж класифікують їх, уникнувши необхідності використання послідовних етапів, як у R-CNN. Це дозволяє забезпечити ефективний баланс між швидкістю та точністю детектування об'єктів у великих масштабах даних. Такі методи широко використовуються у відстеженні об'єктів, детектуванні облич, розпізнаванні об'єктів у відео та багатьох інших завданнях комп'ютерного зору [5].

Деформовані згорткові мережі (Deformable Convolutional Networks) представляють собою тип згорткових нейронних мереж, який відрізняється від класичних CNN тим, що вони можуть вивчати більш складні форми об'єктів для покращення точності їх детектування. Цей тип мережі враховує можливі деформації об'єктів та дозволяє адаптувати свою архітектуру для врахування цих деформацій. Деформовані згорткові мережі зазвичай використовуються у завданнях детектування об'єктів, де точність та робастність до змін форми об'єктів важливі для досягнення високих результатів. Їхнє використання дозволяє покращити якість детектування об'єктів у зображеннях, зокрема у сценаріях зі значними деформаціями об'єктів або варіаціями їхньої форми [5].

### Висновки

Розпізнавання об'єктів на фото та відео — це динамічно розвиваюча галузь, яка має значний потенціал для покращення нашого життя. Завдяки постійному вдосконаленню методів та алгоритмів, системи розпізнавання об'єктів стають все більш потужними та точними, що відкриває нові можливості для їх застосування в різних сферах.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Дослідження алгоритмів комп'ютерного зору та розпізнавання об'єктів на двовимірних зображеннях: thesis/ В.А.Толбатов та ін. 2016. URL:<http://essuir.sumdu.edu.ua/handle/123456789/46591>.
2. Ivanishyn I. USING VIOLA-JONES AND K NEAREST NEIGHBORS METHODS FOR FACE DETECTION AND RECOGNITION. *International scientific journal "Internauka"*. 2022. № 11(130). URL: <https://doi.org/10.25313/2520-2057-2022-11-8281>.
3. Матлахов В. І. Інтелектуальна система розпізнавання образів у Web-контексті : master's thesis. 2020. URL: <https://essuir.sumdu.edu.ua/handle/123456789/82177>.
4. Грицик В. В., Березький О. М. Методи та високопродуктивні паралельні системи для обробки та розпізнавання зображень у реальному часі. *International journal of computing*. 2014. С. 25–24. URL: <https://doi.org/10.47839/ijc.2.1.159>.
5. Myronyuk D. M., Blahitko B. Y., Zajazchuk I. M. Computer simulation of deep learning for image recognition. *Computer Technologies of Printing*. 2019. Vol. 2, no. 42. P. 57–63. URL: <https://doi.org/10.32403/2411-9210-2019-2-42-57-63>.

**Бондаренко Катерина Олександрівна** – студентка групи ІКІ-20б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця.

**Гнідунець Володимир Олексійович** – студент групи ІКІ-20б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, Вінниця.

**Крупельницький Леонід Віталійович** – кандидат технічних наук, доцент кафедри обчислювальної техніки Вінницького національного технічного університету, Вінниця.

**Bondarenko Kateryna** – a student of group ІКІ-20b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia.

**Hnidunets Volodymyr** – a student of group ІКІ-20b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia.

**Krupelnitskyi Leonid** – candidate of technical sciences, associate professor of the Department of Computer Engineering, Vinnytsia National Technical University, Vinnytsia.