

# Розв'язання Задачі Підрахунку Об'єктів при Складних Умовах Отримання Зображень

Роман Кветний  
Кафедра автоматики та інформаційно-вимірювальної  
техніки,  
Вінницький національний технічний університет,  
Вінниця, Україна  
[rkvetny@sprava.net](mailto:rkvetny@sprava.net)

Олег Бісікало  
Кафедра автоматики та інформаційно-вимірювальної  
техніки,  
Вінницький національний технічний університет,  
Вінниця, Україна  
[obisikalo@gmail.com](mailto:obisikalo@gmail.com)

Володимир Коцюбинський  
Кафедра автоматики та інформаційно-вимірювальної  
техніки,  
Вінницький національний технічний університет,  
Вінниця, Україна  
[vkotsyubinsky@sprava.net](mailto:vkotsyubinsky@sprava.net)

Роман Маслій  
Кафедра автоматики та інформаційно-вимірювальної  
техніки,  
Вінницький національний технічний університет,  
Вінниця, Україна  
[romas4580@gmail.com](mailto:romas4580@gmail.com)

Олександр Кириленко  
Кафедра автоматики та інформаційно-вимірювальної  
техніки,  
Вінницький національний технічний університет,  
Вінниця, Україна  
[sasha.kyrylenko@gmail.com](mailto:sasha.kyrylenko@gmail.com)

## Solving the Problems of Counting Objects in Complicated Conditions for Obtaining Images

Poman Kvyetnyy  
Department of automation and information measuring  
devices,  
Vinnytsia national technical university,  
Vinnytsia, Ukraine  
[rkvetny@sprava.net](mailto:rkvetny@sprava.net)

Oleg Bisikalo  
Department of automation and information measuring  
devices,  
Vinnytsia national technical university,,  
Vinnytsia, Ukraine  
[obisikalo@gmail.com](mailto:obisikalo@gmail.com)

Roman Maslii  
Department of automation and information measuring  
devices,  
Vinnytsia national technical university,,  
Vinnytsia, Ukraine  
[romas4580@gmail.com](mailto:romas4580@gmail.com)

Volodymyr Kotsyubinsky  
Department of automation and information measuring  
devices,  
Vinnytsia national technical university,,  
Vinnytsia, Ukraine  
[vkotsyubinsky@sprava.net](mailto:vkotsyubinsky@sprava.net)

Alexandr Kyrylenko  
Department of automation and information measuring  
devices,  
Vinnytsia national technical university,,  
Vinnytsia, Ukraine  
[sasha.kyrylenko@gmail.com](mailto:sasha.kyrylenko@gmail.com)

*Анотація*—в даній роботі досліджуються задача обробки зображень, отриманих у складних умовах, використовуючи згорткові нейронні мережі, для визначення кількості автомобілів з відеозаписів низької якості, які записуються камерами.

*Abstract*—in this paper we investigate the problem of processing images obtained in difficult conditions using convolutional neural networks to determine the number of cars from low quality video recorded by cameras.

*Ключові слова* - виявлення об'єктів, підрахунок кількості об'єктів, глибоке навчання, OpenCV.

*Keywords* - object detection, object counting, deep learning, OpenCV.

## I. ВСТУП

На сьогоднішній день велика кількість міст оснащена сотнями відеокамер на вулицях та перехрестях. Вони цілодобово записують рух транспортних засобів, збираючи велику кількість відеоданих. Ці камери можуть виконувати багато функцій, адже завдяки засобам комп'ютерного зору вони можуть аналізувати потік автомобілів та рахувати кількість перехожих; проте, такі їхні можливості наразі не використовуються у повній мірі.

Підрахунок кількості автомобілів є досить важливою задачею в багатьох сучасних застосуваннях, наприклад, в управлінні трафіком міського транспорту. Проте, водночас це є надзвичайно складною задачею, через обмеження пропускну здатності мереж, недостатній об'єм пам'яті для збереження записів та проблеми приватності. Крім того відео отримані з таких камер характеризуються: низькою роздільною здатністю, низькою частотою кадрів; великим відсотком перекриття автомобілів іншими автомобілями, особливо в години пік; великою перспективою. Все це ускладнює підрахунок автомобілів за допомогою відеокамер.

Метою даної роботи є аналіз підходів до підрахунку кількості об'єктів у реальному часі в відеопослідовностях та дослідження задачі підрахунку на прикладі автомобілів.

## II. ПІДХОДИ ДО ПІДРАХУНКУ ОБ'ЄКТІВ

В якості об'єктів для підрахунку у відео послідовності були вибрані автомобілі. В результаті аналізу літературних джерел виділимо такі підходи до підрахунку [1]:

frame differencing based (методи основані на різниці у фреймах) – характеризуються легким налаштуванням параметрів, але мають високу чутливість до шумів, різких змін освітлення та змін на задньому фоні;

detection based (методи основані на виявленні об'єктів) - успішно відстежують окремі автомобілі в кожному кадрі, проте не показують високих результатів при низькій роздільній здатності відео та великому відсотку перекриття автомобілів іншими автомобілями;

motion based (методи основані на русі об'єкту) – здійснюють відстеження автомобілів, але не здатні підраховувати кількість автомобілів у реальному часі при низькій частоті кадрів.

density estimation based (методи основані на оцінці щільності ймовірності) – не показують високих результатів при великій перспективі та автомобілів великих розмірів;

методи основані на глибокому навчанні (deep learning based) - останнім часом показують високі результати виявлення та підрахунку автомобілів, що робить їх досить перспективними при застосуванні у складних умовах.

## III. ВИКОРИСТАННЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ ПІДРАХУНКУ ОБ'ЄКТІВ

Застосування багатошарового перцептрона з традиційною структурою при вирішенні реальних задач розпізнавання і класифікації зображень викликає певні труднощі. По-перше, зображення, як правило, мають велику розмірність, внаслідок чого зростає число нейронів і синаптичних зв'язків у мережі. У свою чергу, це вимагає збільшення навчальної вибірки, внаслідок чого збільшується час і обчислювальна складність процесу навчання. По-друге, ігнорується топологія вхідних даних. Компоненти вхідного шару можуть бути представлені в будь-якому порядку, без урахування мети навчання. Однак зображення мають строго двовимірну структуру, в якій існує залежність між просторово-сусідніми пікселями [2].

Від цих недоліків вільні так звані згорткові нейронні мережі, які являють собою особливий клас багатошарових перцептронів, спеціально створених для розпізнавання двовимірних поверхонь з високим ступенем інваріантності до масштабування, зсуву, повороту, зміни ракурсу та інших просторових спотворень, і глибокі НМ побудовані на базі автоенкодерів, попереднє навчання яких здійснюється на базі обмежених машин Больцмана [2].

Глибока нейронна мережа (DNN – Deep Neural Network) це штучна нейронна мережа з декількома прихованими шарами. Подібно до звичайних нейронних мереж, глибокі нейронні мережі можуть моделювати складні нелінійні відносини між елементами. У процесі навчання глибокої нейронної мережі отримувана модель намагається представити об'єкт у вигляді комбінації простих примітивів (наприклад, у задачі розпізнавання осіб такими примітивами можуть бути частини обличчя: ніс, очі, рот і так далі). Додаткові шари дозволяють будувати абстракції все більш високих рівнів, що і дозволяє будувати моделі для розпізнавання складних об'єктів реального світу.

Як правило, глибинні мережі будуються як мережі прямого поширення. Однак останні дослідження показали, як можна застосувати техніку глибинного навчання для рекурентних нейронних мереж. Згорткові нейронні мережі використовуються в області машинного зору, де цей підхід показав себе як ефективний. Також згорткові нейронні мережі були застосовані для розпізнавання мови.

Згорткова нейронна мережа (ЗНМ) – тип багатошарової нейронної мережі, яка свою назву «згорткова мережа» отримала за назвою операції – згортка, вона часто використовується для обробки зображень і може бути описана наступною формулою

$$(f \times g)[m, n] = \sum_{k, l} f[m - k, n - l] \cdot g[k, l],$$

де  $f$  – вихідна матриця зображення;  $g$  – ядро (матриця) згортки.

Неформально цю операцію можна описати наступним чином – вікном розміру ядра  $g$  проходимо з заданим кроком (зазвичай 1) все зображення  $f$  на кожному кроці поелементно множимо вміст вікна на ядро  $g$ , результат сумується і записується в таблицю результату.

Ідея згорткових нейронних мереж полягає в чергуванні згорткових шарів (англ. convolution layers) і субдискретизуючих шарів (англ. subsampling layers, верств підвибірки) [2].

Структура мережі – односпрямована (без зворотних зв'язків), багат шарова (рис. 1). Модель згорткової мережі складається з трьох типів шарів: згорткові (convolutional) шари, субдискретизуючі (subsampling, підвибірка) верстви і прошарки «звичайної» нейронної мережі – перцептрона.

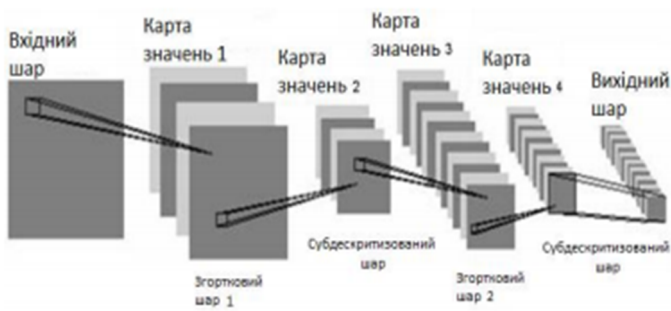


Рис. 1. Структура згорткової мережі.

Архітектура згорткових нейронних мереж реалізує три ідеї, які забезпечують інваріантність мережі до невеликих зрушень, змін масштабу і спотворень [2]:

- кожен нейрон отримує вхідний сигнал від локального рецептивного поля (local receptive fields) у попередньому шарі, що забезпечує локальну двовимірну зв'язність нейронів;
- кожен прихований шар мережі складається з безлічі карт ознак, на яких всі нейрони мають загальні ваги (shared weights), що забезпечує інваріантність до зміщення і скорочення загальної кількості вагових коефіцієнтів мережі;
- за кожним шаром згортки слідує обчислювальний шар, який здійснює локальне усереднення та підвибірку, що забезпечує зменшення розширення для карт ознак.

Робота згорткової нейронної мережі забезпечується двома основними елементами.

Фільтри (filters) (визначники ознак).

Кarti ознак (feature maps).

Кarti ознак (feature maps).

Фільтр – це невелика матриця, що представляє ознаку, яку необхідно знайти на вихідному зображенні. За допомогою верхнього фільтра визначаються частини вихідного зображення з вертикальними лініями, нижній фільтр служить для визначення частин зображення з горизонтальними лініями.

Безпосередньо процес визначення заснований на операції згортки фільтром оригінального зображення. Результати згортки, які визначають місце розташування ознак вихідного зображення, називаються картами ознак [2].

Мета процесу згортки – зменшити розмірність карти ознак до такої міри, щоб з повним набором ознак могла працювати мережа прямого поширення (в більшості випадків багат шаровий перцептрон). Згортковий шар реалізує ідею локальних рецептивних полів, тобто кожен вихідний нейрон з'єднаний тільки з певною (невеликою) областю вхідної матриці і таким чином моделює деякі особливості людського зору [2].

Недоліками згорткових нейронних мереж (ЗНМ) є:

- висока складність архітектури;
- повнозв'язаність;
- фіксована площа вікна шару згортки.

З метою підвищення ефективності роботи ЗНМ необхідно знайти оптимальні значення наступних параметрів:

- кількість карт ознак;
- щільність зв'язків між картами ознак;
- розмір вікна;
- площа перекриття;
- початкова ініціалізація ваг [2].

#### IV. ПРОЕКТУВАННЯ ТА НАВЧАННЯ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ВИРІШЕННЯ ЗАДАЧІ ПІДРАХУНКУ ОБ'ЄКТІВ

DIGITS (Deep Learning GPU Training System) - система підготовки глибокого навчання при використанні графічних процесорів NVIDIA [3]. Дана система використовується вченими та інженерами для швидкої підготовки та навчання глибоких нейронних мереж для задач класифікації зображень, сегментації та виявлення об'єктів.

За допомогою даної системи, було спроектовано згорткову нейронну мережу та здійснено її навчання на наборі даних KITTI [4], результати навчання нейронної мережі зображено на рисунку 2.

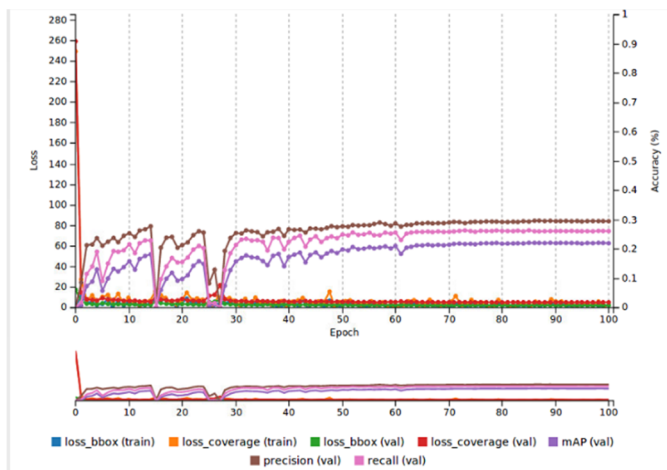


Рис. 2. Результат навчання нейронної мережі в DIGITS

В результаті навчання нейронної мережі на 100 епохах було досягнуто таких показників вірогідності виявлення на валідаційному наборі даних: precision (val) = 82%, recall (val) = 73%, mAP (val) = 62%.

Приклад виявлення автомобілів у складних умовах освітлення на тестовому зображенні при використанні навченої нейронної мережі зображено на рисунку 3.

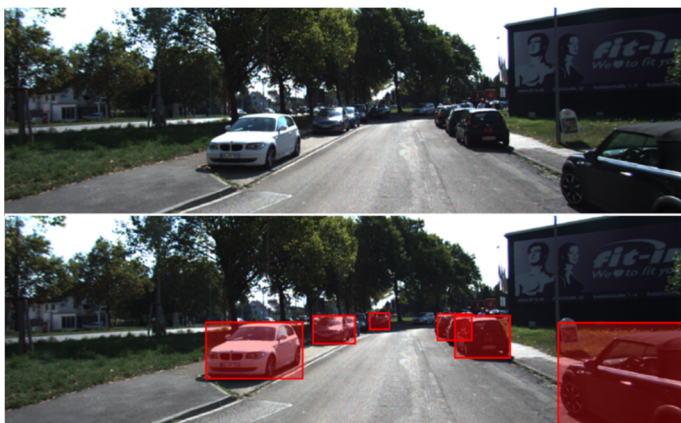


Рис. 3. Приклад виявлення автомобілів

На основі моделі, отриманої в результаті глибокого навчання нейронної мережі при використанні NVIDIA DIGITS, було розроблено програму, що здійснює виявлення та підрахунок автомобілів у відеопослідовностях. Програма розроблена на мові Python з використанням бібліотеки OpenCV. Результат роботи програми зображено на рисунку 4.



Рис. 4. Результат роботи програми

## ВИСНОВКИ

У роботі проведений аналіз підходів до підрахунку кількості об'єктів при складних умовах отримання зображень. Розглянуто переваги та недоліки методів. Перспективним для подальших досліджень визначено підхід, що використовує глибоке навчання, так як існуючі методи з цього підходу показують найкращі результати по критерію вірогідності, в тому числі при складних дорожніх умовах, низькій роздільній здатності та низькій частоті відеокадрів.

Розроблено програму, що здійснює виявлення та підрахунок автомобілів у відеопослідовностях, на основі моделі отриманої в результаті глибокого навчання нейронної мережі при використанні NVIDIA DIGITS.

## ЛІТЕРАТУРА REFERENCES

- [1] Zhang Sh. FCN-rLSTM: Deep Spatio-Temporal Neural Networks for Vehicle Counting in City Cameras / Sh. Zhang, G. Wu, J. P. Costeira, J. M. F. Moura // IEEE International Conference on Computer Vision (ICCV). – 2017.
- [2] Синеглазов В. Глибокі нейронні мережі для вирішення завдань розпізнавання і класифікації зображення [Електронний ресурс] В. Синеглазов, О. Чумаченко. – 2017. – Режим доступу до ресурсу: <http://itcm.comp-sc.if.ua/2017/Sineglazov.pdf>.
- [3] NVIDIA DIGITS [Електронний ресурс] – Режим доступу до ресурсу: <https://developer.nvidia.com/digits>. – Назва з екрану.
- [4] Fritsch J. A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms / J. Fritsch, T. Kuehnl, A. Geiger // International Conference on Intelligent Transportation Systems (ITSC) – 2013.