Проблематика реалізації паралелізму в сучасних алгоритмах детектування та розпізнавання зображень

Ігор Лазарович, Сергій Яковин

Кафедра інформаційних технологій
ДВНЗ "Прикарпатський національний університет імені Василя Стефаника"
м.Івано-Франківськ, Україна

***Анотація*—В роботі проаналізовано особливості реалізації задачі детектування об'єктів на зображеннях засобами сучасних та ефективних методів з метою визначення можливості виділення операцій, які можуть бути виконані паралельно. Здійснено практичну реалізацію моделей та виконано порівняльний аналіз ефективності розпаралелення.**

Keywords—паралельний алгоритм, паралелізм, детектування зображень, згорткова нейронна мережа, машинне навчання, навчальна вибірка.

I. Вступ

Завдяки популярності використання технологій штучного інтелекту, зокрема комп'ютерного бачення, машинного навчання, відбувається швидкий розвиток методів обробки та розпізнавання зображень. Разом із постійно зростаючою роздільною здатністю сучасних цифрових камер, зростає і обчислювальна складність реалізації алгоритмів роботи із зображеннями. Тому, не зважаючи на зниження "вартості" обробки інформації, коли йде мова про реалізацію подібних задач в енергонезалежних пристроях (смартфонах, мобільних роботах), обчислювальна складність алгоритму та його швидкодія є одними із стримуючих факторів імплементації теорії в практичні рішення.

II. Аналіз останніх досліджень і постановка задачі

Протягом останніх років значний розвиток отримали методи детектування і розпізнавання зображень на основі згорткових нейронних мереж CNN (Convolutional Neural Networks) [1-3]. Відомі платформи для машинного навчання TensorFlow і комп'ютерного зору OpenCV [4] містять засоби для програмних реалізацій цих мереж. Дослідження вищенаведених алгоритмів показує можливість розпаралелення значної кількості обчислень, що дозволяє зменшити час роботи алгоритму. Проте відсутній порівняльний аналіз ефективності саме паралельних реалізацій нейромереж. Тому проведення таких досліджень є актуальною задачею.

III. Виклад основного матеріалу

При використанні нейронних мереж для детектування і класифікації об'єктів основними параметрами, які визначають вибір конкретного алгоритму є: точність і швидкодія . Точність визначає якість детектування – на скільки точно знайдений об'єкт буде описаний обмежуючим контуром, інваріантність до розміру шуканого об'єкта, можливість розділення схожих об'єктів і т.д. Точність детектування залежить як від вибору типу моделі, так і від особливостей самого зображення і його параметрів, зокрема для невеликих об'єктів зазвичай важко забезпечити високу точність . Швидкодія моделі визначається часом опрацювання

кадру зображення, або більш точно – у вигляді сукупної кількості операцій для такого опрацювання. Якщо пов'язувати час виконання і кількість операцій, то час можна скоротити, якщо алгоритм буде розпаралелено.

Серед сімейства згорткових нейронних мереж за параметрами балансу точності та швидкодії можна виділити наступні типи мереж [5]: SSD (Single Shot MultiBox Detector) та YOLO v2(You only Look Once). Оскільки ці мережі дозволяють здійснити детектування скінченного набору класів, то із певним наближенням можна говорити і про розпізнавання об'єктів. Основна ресурсоємка робота в цих алгоритмах полягає в багаторазових обчисленнях операції згортки, яка піддається паралельній реалізації.

Для перевірки ефективності розпаралелення було розроблено програмні моделі детекторів засобами мови C++ та бібліотеки OpenCV із підтримкою багатопотоковості на базі технології CUDA. Для тренування нейромереж використано набір тестових зображень PascalVOCdatabase [6], об'єм навчальної вибірки складає 5000 зображень. Після отримання натренованих моделей було проведено дослідження залежності часу детектування об'єкту від кількості паралельних потоків, які виконувалися засобами GPU Nvidia Geforce 940MX, в якій може паралельно працювати до 384 паралельних конвеєрів. Проводилось також тренування і дослідження роботи моделей з використанням CPU Intel I5 7200U (2 ядра, 4 потоки).

IV. Результати

Результати тестування доводять ефективність паралельної роботи моделей. Час обробки одного кадру SSD та YOLO v2 для однопотокової реалізації на CPU складає відповідно 63мс(15,7fps) і 205мс (4.9fps). Для цих самих вхідних даних і моделей, але багатопоточної реалізації засобами GPU отримано наступі показники часу 18.9мс (53fps) та 47мс (21.3fps), прискорення для SSD складає 3,33рази, а для YOLO v2 - 4,36 рази.

V. Висновки

Проведені теоретичні та практичні дослідження методів детектування засобами моделей SSD та YOLO v2 підтвердили ефект від їх паралельної реалізації. Перспективним є дослідження низькорівневої паралельної реалізації наведених моделей для високопродуктивних багатоядерних мікропроцесорів.

Література

[1] B. A. Godinho de Oliveira, F. Magalhães Freitas Ferreira and C. A. P. d. S. Martins, "Fast and Lightweight Object Detection Network: Detection and Recognition on Resource Constrained Devices," in *IEEE Access*, vol. 6, pp. 8714-8724, 2018, doi: 10.1109/ACCESS.2018.2801813.

[2] Z. Jiang and D. Q. Huynh, "Multiple pedestrian tracking from monoc-ular videos in an interacting multiple model framework," IEEE *Trans.Image Process*, vol. 27, pp. 1361–1375, 2018.

[3] M. Kozlenko, V. Tkachuk, and M. Dutchak, "Software implementation of microcomputer based intrusion detection and prevention system with binary neural network," in *Proceedings 2nd International Scientific-Practical Conference on Problems of Cyber Security of Information and Telecommunication Systems (PCSITS)*, O. Oksiiuk et al, Eds. Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, Apr. 11-12, 2019, pp. 371-373.

[4] W. G. Hatcher and W. Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends," in *IEEE Access*, vol. 6, pp. 24411-24432, 2018, doi: 10.1109/ACCESS.2018.2830661.

[5] CV-Tricks.com Learn Machine Learning, AI & Computer vision. URL: https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/ (дата звернення: 25.10.2019).

[6] PascalVOCdatabase. URL: https://host.robots.ox.ac.uk/pascal/VOC (дата звернення 14.05.2018)