

АНАЛІЗ ТИПІВ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ЗАДАЧІ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ ДОРОЖНЬОГО РУХУ

Вінницький національний технічний університет

Анотація

Розглянуто принцип роботи згорткових нейронних мереж, методи виявлення і розпізнавання об'єктів дорожнього руху, та їх особливості.

Ключові слова: штучний інтелект, згорткова нейронна мережа, виявлення образів, розпізнавання об'єктів, дорожній рух.

Abstract

The principle of operation of convolutional neural networks, methods of road traffic objects detection and recognition, and their features are considered.

Keywords: artificial intelligence, convolutional neural network, object detection, object recognition, road traffic.

Вступ

Для задачі розпізнавання об'єктів у зображеннях найкраще підходять згорткові нейронні мережі, проте на сьогоднішній день існує чимала кількість різних методів розпізнавання і є потреба у визначенні найбільш оптимального, з точки зору точності та швидкості, типу згорткової нейромережі для задачі розпізнавання об'єктів дорожнього руху в режимі реального часу [1 – 4].

Метою цього дослідження є аналіз типів згорткових нейронних мереж для задачі розпізнавання об'єктів дорожнього руху.

Результати дослідження

Згорткові нейронні мережі (ЗНМ) мають багато спільних рис із іншими нейромережами: вони утворені шарами та нейронами, що мають параметри у вигляді ваг і біасів, які можна навчати. Характерна особливість ЗНМ полягає в тому, що вони здатні добре обробляти зображення, які подані у вигляді масивів даних, що дозволяє кодувати певні властивості класів об'єктів в архітектурі мережі для розпізнавання конкретних образів у зображеннях.

Кожен шар згорткової нейромережі являє собою різні рівні абстракції. Перші декілька шарів вивчають особливості нижчого рівня, такі як краї, криві або візерунки. Глибокі шари поєднують характеристики для ідентифікації певних об'єктів. Класифікаційний шар зазвичай складається з одного ряду нейронів. Кількість нейронів залежить від кількості бажаних виходів (тобто, кількості класів об'єктів). Наприклад, класи можуть бути пішоходи, велосипедисти або автомобілі, а значить, потрібні три класи. Чим вище вихідне значення для одного з цих нейронів класифікатора, тим більший шанс успішного виявлення пішохода або велосипедиста.

Методи розпізнавання об'єктів у зображеннях можуть бути однією з таких двох категорій: двоетапний (підхід на основі регіональних зон інтересу) або одноетапний детектор (нерегіональний підхід). Початковий крок двоетапного детектора ідентифікує підмножини регіонів зображення, які можуть містити об'єкт. На другому кроці класифікуються об'єкти цих регіонів. Двоетапні методи можуть досягати дуже точних результатів розпізнавання об'єктів, хоча вони, як правило, повільніші, ніж одноетапні. Деякі з методів, що базуються на регіональних зонах інтересу, включають R-CNN, Fast R-CNN та Faster RCNN [5]. Однак проблема цього підходу полягає в тому, що об'єкти, які являють інтерес, можуть мати різні просторові розташування в межах зображення і різні співвідношення сторін. Тому доведеться вибирати величезну кількість регіонів інтересу, і для цього може використовуватись досить багато обчислювальних потужностей.

R-CNN (англ. Region-based Convolutional Neural Networks): щоб обійти проблему вибору величезної кількості областей, Рос Гіршик та ін. запропонували метод, при якому використовується

вибірковий пошук, щоб отримати із зображення лише 2000 областей [6]. Тому замість того, аби намагатися класифікувати об'єкти у великій кількості регіонів, R-CNN працює тільки з 2000 регіонів, які генеруються з використанням алгоритму вибіркового пошуку, що складається з таких кроків:

- а) створити початкову підсегментацію, згенерувати багато регіонів;
- б) використати жадібний алгоритм для рекурсивного об'єднання подібних регіонів у великі;
- в) використати згенеровані регіони для створення 2000 кінцевих регіонів-кандидатів різних розмірів і співвідношень сторін.

Серед основних проблем цього алгоритму можна виділити:

- метод вибіркового пошуку є фіксованим алгоритмом, тому на цьому етапі навчання мережі не відбувається. Це може призвести до генерування поганих пропозицій щодо регіону-кандидата;
- цей підхід неможливо реалізувати для задач розпізнавання об'єктів в режимі реального часу, оскільки для кожного тестового зображення потрібно близько 49 секунд;
- навчання нейромережі займає величезну кількість часу, оскільки необхідно класифікувати 2000 регіонів-кандидатів на зображення.

Незважаючи на високі результати точності роботи, автор R-CNN усунув деякі недоліки мережі для побудови більш швидкого алгоритму виявлення об'єктів, що називається Швидким R-CNN (Fast R-CNN) [7]. Підхід аналогічний попередньому методу, але запропоновані певні модифікації для прискорення процесу обчислень. Зокрема, пропускати через ЗНМ не кожен з 2000 регіонів-кандидатів окремо, а все зображення повністю. Як результат, операція згортки виконується тільки один раз на зображення, і з неї генерується карта ознак. Недоліком такого алгоритму залишається використання вибіркового пошуку для пошуку регіонів-кандидатів. Вибірковий пошук є повільним і трудомістким процесом, що впливає на продуктивність нейромережі.

У 2015 р. команда з Microsoft Research запропонувала обчислювати регіони не за початковим зображенням, а знову ж таки за картою ознак, отриманих з ЗНМ. Для цього був доданий модуль під назвою Region Proposal Network (RPN). Нова архітектура отримала назву Faster R-CNN [8]. Подібно до Fast R-CNN, зображення подається на вхід до згорткової мережі, яка генерує карту функцій згортки. Замість використання алгоритму вибіркового пошуку на карті ознак для ідентифікації пропозицій регіону, у Faster R-CNN для прогнозування використовується окрема мережа RPN. Прогнозовані регіони потім масштабуються та об'єднуються за допомогою шару об'єднання, який потім використовується для класифікації зображення в межах запропонованої області та прогнозування значень зміщення для обмежувальних прямокутників.

За кількістю часу, необхідного для розпізнавання об'єктів, Faster R-CNN швидше, ніж R-CNN і Fast R-CNN у 245 та 11,5 разів відповідно (рис. 1), тому потенційно може використовуватись для виявлення об'єктів у режимі реального часу.

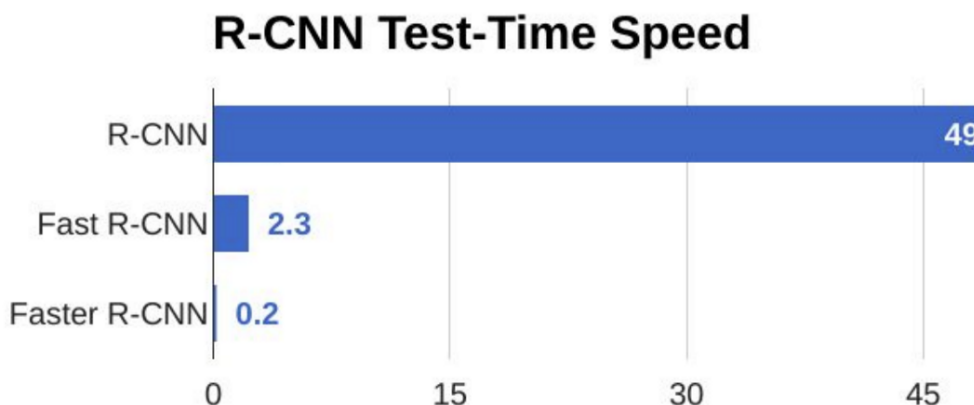


Рисунок 1 – Порівняння швидкості роботи алгоритмів виявлення об'єктів, одиниці вимірювання – секунди

Наскільки перспективними не були би двоетапні детектори, обчислення на другій стадії (тобто, на етапі класифікатора) є досить важкими. Отже, були запропоновані детектори одноетапної дії. Одноетапний детектор класифікує об'єкти у всьому зображенні за допомогою обмежувальних прямокутників, і потім прогнози декодуються для створення остаточних обмежувальних прямокутників для розпізнаних об'єктів. Такі методи можуть бути набагато швидшими, ніж

двоетапні, але вони можуть не досягати однакового рівня точності, особливо для зображень, що містять невеликі об'єкти. Серед нерегіональних підходів можна виділити реалізацію YOLO [9].

You only look once (YOLO) – ЗНМ з одноетапною архітектурою для виявлення об'єктів із повних зображень, яка ділить вхідне зображення на сітку $N \times N$ та для кожної комірки створює M обмежувальних прямокутників. Далі мережа прогнозує для кожного обмежувального прямокутника ймовірність розташування певного класу об'єкта [10]. Останнім кроком виділяються тільки ті прямокутники, в яких ймовірність розташування об'єкта вище порогового значення, наприклад вище 65% (рис. 2). YOLO на порядок швидше, ніж вищезгадані алгоритми виявлення об'єктів і може обробляти до 45 кадрів в секунду, показуючи найкращу взаємозалежність між швидкістю роботи в режимі реального часу та точністю розпізнавання об'єктів. Єдине обмеження алгоритму YOLO полягає в тому, що він не може справлятися з малими об'єктами у зображенні, наприклад, у нього можуть виникнути труднощі у виявленні зграї птахів. Це пов'язано з просторовими обмеженнями алгоритму.

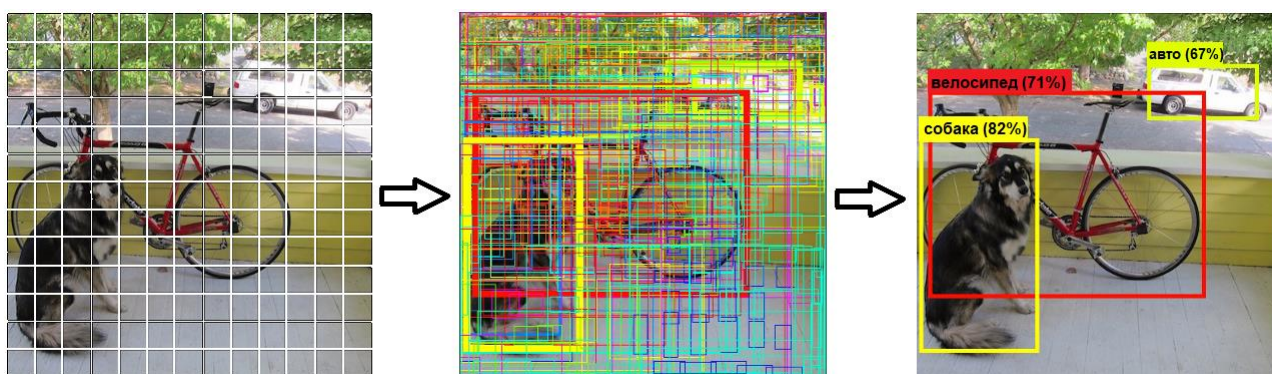


Рисунок 2 – Візуалізація роботи методу виявлення об'єктів YOLO

Висновки

Згідно з проведеним дослідженням, проаналізовано ряд основних реалізацій згорткових нейронних мереж, визначено їх переваги та недоліки. З наведених характеристик можна зробити висновок, що двоетапні методи розпізнавання об'єктів впроваджуються при високій ентропії для підвищення точності роботи нейромережі, тоді як одноетапні методи орієнтовані на загальну швидкість системи, що дозволяє використовувати їх у режимі реального часу. Тому серед розглянутих варіантів, обрано одноетапну згорткову нейромережу YOLO, як найбільш оптимальну з точки зору точності та швидкості розпізнавання об'єктів дорожнього руху в режимі реального часу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. What is Object Detection [Електронний ресурс]. – Режим доступу: <https://www.mathworks.com/discovery/object-detection.html/>. – Назва з екрану.
2. Convolutional Neural Networks [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/convolutional-neural-networks-for-beginners-using-keras-and-tensorflow-2-c578f7b3bf25>. – Назва з екрану.
3. Демчук М. Ю. Дослідження нейромережевого підходу для розпізнавання об'єктів дорожнього руху [Електронний ресурс] / М. Ю. Демчук, Л. В. Крилик // L Науково-технічна конференція підрозділів Вінницького національного технічного університету (Вінниця, 2021). – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fitki/all-fitki-2021/paper/view/12348>. – Назва з екрану.
4. Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. CNN features off-the-shelf: an astounding baseline for recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2014; pp. 806-813.
5. R-CNN, Fast R-CNN, Faster R-CNN, YOLO – Object Detection Algorithms [Електронний ресурс]. – Режим доступу: <https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e>. – Назва з екрану.
6. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 24-27 June 2014; pp. 580-587.
7. Girshick, R. Fast R-CNN. In Proceedings of the IEEE international conference on computer vision, 2015; pp. 1440-1448.
8. Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, 2015; pp. 91-99.
9. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016; pp. 779-788.

10. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017; pp. 6517-6525.

Демчук Микола Юрійович — студент групи 2КН-20м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця,
email: demchuknick6@gmail.com.

Крилик Людмила Вікторівна — к.т.н, доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця.

Demchuk Mykola Yu. — Department Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, email: demchuknick6@gmail.com.

Krylik Lyudmila V. — PhD (Eng.), Associate Professor of Department for Computer Science, Vinnytsia National Technical University, Vinnytsia.