

DOI [10.28925/2663-4023.2020.8.97112](https://doi.org/10.28925/2663-4023.2020.8.97112)

УДК 004.08; 005.4

Сватюк Данило Романович

інженер-програміст

ELEKS, Львів, Україна

ORCID: 0000-0002-6006-4810

*danylosva@gmail.com***Сватюк Оксана Робертівна**

к.е.н., доцент, доцент кафедри менеджменту та соціально-гуманітарних дисциплін

Львівський інститут ДВНЗ«Університет банківської справи», Львів, Україна

ORCID: 0000-0003-0099-2532

*svatuk@gmail.com***Белей Олександр Ігорович**

к.е.н., доцент, доцент кафедри Систем автоматизованого проектування

Національний університет «Львівська політехніка», Львів, Україна

ORCID: 0000-0003-4150-7425

Oleksandr.I.Belei@lpnu.ua

ЗАСТОСУВАННЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ БЕЗПЕКИ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ У ВІДЕОПОТОЦІ

Анотація. Стаття присвячена аналізу методів розпізнавання зображень та пошуку їх у відеопотоці. Проаналізовано еволюцію структури згорткових нейронних мереж, що використовуються в області діагностики комп'ютерних відеопотоків. Здійснено оцінку ефективності алгоритмів діагностики відеопотоків та розпізнавання номерних знаків автомобілів. Описана методика розпізнавання номерних знаків автомобілів, що знаходяться у відеопотоці транспортних нейронних мереж. В дослідженні приділено увагу створенню комбінованої системи, яка поєднує в собі технологію штучного інтелекту та комп'ютерного зору на основі нечіткої логіки. Для вирішення проблеми розпізнавання зображень номерних знаків у відеопотоці транспортної системи запропоновано метод розпізнавання зображень у безперервному відеопотоці з його реалізацією на основі складу традиційних методів обробки зображень та нейронних мереж із згортковими та періодичними шарами. Описано структуру та особливості функціонування інтелектуальної розподіленої системи безпеки міського транспорту, особливістю якої є використання мобільних пристроїв, підключених до єдиної мережі. Запропоновано та здійснено практичну реалізацію програмного застосування для розпізнавання автомобільних номерних знаків мобільними пристроями на платформі операційної системи Android. Розроблено різні сценарії розпізнавання номерних знаків автомобілів у реальному часі та збереження їх у базі даних для подальшого аналізу та використання. В запропонованому застосуванні використовуються дві різні спеціалізовані нейромережі: одна - для детектування об'єктів у відеопотоці, інша – для розпізнавання тексту з виділеного зображення. Проведене випробування та аналіз програмного застосування на платформі операційної системи Android для розпізнавання автомобільних номерних знаків у реальному часі підтвердив працездатність запропонованого математичного забезпечення і може використовуватися для безпечного аналізу номерних знаків автомобілів у сканованому відео потоці шляхом порівняння з номерними знаками в існуючій базі даних. Авторами реалізовано функціонування метод згорткових нейронних мереж виявлення та розпізнавання номерних знаків, персоналу та критичних ситуацій у відеопотоці з камер мобільних пристроїв у режимі реального часу. Продемонстрована можливість його застосування у сфері безпечної ідентифікації номерних знаків автомобілів.

Ключові слова: нейромережі; згорткові мережі; відео потік; програма мобільних пристроїв.



1. ВСТУП

Методи розпізнавання образів, засновані на штучному інтелекті, застосовуються в системах комп'ютерного зору (СКЗ) для автоматичної класифікації образів об'єктів або ситуацій з кадрів відеопотоку і віднесення їх до певного класу.

Варто відзначити, що на кадрах відео потоку можуть бути присутніми фізичні і технічні об'єкти, де кожен з них має певний набір властивостей, а взаємодія об'єктів характеризує ситуацію. Тоді завданнями алгоритмів, застосовуваних в СКЗ, є знаходження об'єктів на зображенні і витяг їх ключових ознак з наступною класифікацією і виробленням керуючих впливів [1, 2].

Останнім часом розвиток отримують інтелектуальні транспортні системи: від систем автомобільної навігації, регулювання світлофорів, систем контролю трафіку і розпізнавання автомобільних номерів до систем, що інтегрують інформаційні потоки від учасників транспортної мережі.

Зі збільшенням швидкостей транспортних засобів та інтенсивності руху особливого значення для сучасної транспортної системи набуває підвищення її безпеки. При цьому особливо важливе значення набуває виявлення і розпізнавання критичних ситуацій в режимі реального часу (РРЧ).

Критичними ситуаціями для транспортних систем можуть бути: ймовірність дорожньо-транспортної пригоди, якщо на шосе сторонній предмет; наявність ДТП, заторів для попередження транспортної системи; погана видимість, викликана природними катаклізмами, атмосферними явищами; несанкціонований доступ до автомобілів на парковці, у дворах; наявність штучних і природних перешкод на дорозі для їх ідентифікації; неадекватна поведінка транспортного засобу, викликане як людським фактором, так і об'єктивними причинами.

Розпізнавання цих ситуацій виконується на основі візуальних даних, водієм або оператором міської транспортної мережі, а значить для вирішення цього завдання також можуть застосовуватися СКЗ. Об'єкти розпізнавання для безпілотних автомобілів (UGV) або систем допомоги водієві DA: об'єкти фону, транспортні засоби, люди, різні дії і непрямі ознаки критичних ситуацій.

Зростання кількості автономних транспортних засобів, безпеки розумного відеоспостереження, систем виявлення обличчя та ідентифікації осіб стимулюють все нові і нові дослідження у цій сфері. Такі системи передбачають не тільки розпізнавання та класифікацію кожного об'єкта в зображенні, але і локалізацію кожного з них, накладаючи відповідне обмежувальне поле навколо них. Це робить виявлення об'єктів суттєво важчим завданням, ніж шляхом простої класифікації зображень. Саме тому безпека виявлення об'єктів є однією з важливих задач у сфері інформаційних технологій, яка прогресує надзвичайно швидко. З появою перших згорткових нейронних мереж у 2012 році вони стрімко знайшли своє місце застосування у багатьох сферах.

Постановка проблеми. У цій роботі буде досліджено безпеку застосування задачі розпізнавання номерних знаків автомобілів. Розпізнавання номерних знаків досі залишається найбільш розповсюдженим комерційним продуктом на основі технологій комп'ютерного бачення. Найбільше застосування ці продукти знайшли у таких сферах як системи контролю правил дорожнього руху та системи ідентифікації на контрольно-пропускних пунктах. Проте нові рішення на основі машинного навчання та нейромереж успішно конкурують з попередніми алгоритмічними підходами на ринку та у більшості



випадків їх перевершують. Досвід останніх років показує, що алгоритми згорткових нейронних мереж (convolutional neural networks – CNN) дозволяють зробити надійні та гнучкі для застосування рішення, які можуть видавати результат у реальному часі на мобільних пристроях з прийнятним рівнем точності.

Існують й інші зручності: при такому підході завжди можна поліпшити надійність рішення на порядок після реального застосування за рахунок повторного навчання. Крім того, такі алгоритми відмінно реалізуються на графічних модулях (GPU), які значно ефективніші з точки зору споживання електроенергії, ніж звичайні процесори.

Не дивлячись на існуючі розробки та дослідження, є цілий ряд завдань, що потребують вирішення і на які є попит: обробка вхідного відеофайлу чи відео потоку; розбиття на окремі кадри за заданою частотою в залежності від швидкодії пристрою; детектування номерного знаку на зображенні; розпізнавання тексту номерного знаку; структурування та збереження отриманого результату. Для кожного з цих завдань існують різні сценарії їх вирішення, які нами наводяться в нашому дослідженні.

Аналіз останніх досліджень і публікацій. Перші роботи, в яких описувалась концепція CNN (convolutional neural networks), з'явилися ще у 60-х роках минулого століття. Проте тоді вони не змогли бути успішно використаними на практиці через брак обчислювальних потужностей та обмеженість даних для навчання. Увага до цього типу архітектур знову зросла у 2012 році, оскільки на цей момент описані вище проблеми можна було ефективно вирішити. З цього часу було розроблено велику кількість архітектур згорткових мереж та підходів для обробки даних, основні з яких проаналізовано нижче.

Стаття [3] вважається однією з найвпливовіших у цій сфері. Дослідники описують створену ними глибоку згорткову нейромережу Alex Net, яка складалась з п'яти згорткових шарів, які чергувались з під вибірковими та трьох повністю з'єднаних шарів. Мережа здатна класифікувати об'єкти серед 100 різних категорій, а загальне число параметрів у мережі складало 6,5 млн. екземплярів. Автори дослідження [4] вдосконалили архітектуру AlexNet, притримуючись принципу 'просто але глибоко'. У запропонованій ними нейромережі використовувались фільтри для згортки розмірності 3x3 на відміну від 7x7 у AlexNet. Ця стаття мала великий вплив на наукову спільноту, оскільки підтвердила переконання, що нейромережі зі збільшенням глибини можуть суттєво покращувати результати, хоча й вимагають більшу кількість даних та обчислювальних потужностей для навчання.

Розроблена мережа GoogLe Net, що була описана в статті [5], складалась з 22 шарів та суттєво відходила від загальноприйнятої на той час архітектури з послідовним з'єднанням згорткових та під вибіркових шарів, тобто модулі введення. Автори статті також наголошують, що використання таких модулів дозволяє суттєво скоротити використання пам'яті та обчислювальної потужності. У 2015 році Microsoft представила нейромережу з 152 шарами, що була описана у роботі "Deep Residual Learning for Image Recognition" [6], що встановила нові рекорди та практично не поступається людині у сфері виявлення об'єктів на зображенні.



Р. Гіршик разом з іншими дослідниками створили Radial Based CNN (R-CNNs) [7]. В статті описується механізм в якому вікна різних розмірів ковзають по зображенню і до кожного знімку з вікна застосовуються описані вище мережі. Стаття [8] присвячена проблемі знаходження можливих місць розташування об'єктів для його розпізнавання. Дослідники представляють метод вибіркового пошуку який поєднує в собі сильні сторони як методу вичерпного пошуку так і сегментації об'єктів.

Перші моделі R-CNN [9] інтуїтивно починалися з пошуку регіонів, до яких потім застосовувалася класифікація. В R-CNN використовується метод вибіркового пошуку для локалізації об'єктів. Модель R-CNN об'єднує вибіркового метод пошуку [10] для виявлення пропозицій регіону та глибокого навчання, щоб виявити об'єкт у цих регіонах [11].

Fast-R-CNN [12] є прямим нащадком R-CNN і нагадує оригінал багатьма способами. FasterR-CNN [13] зараз є канонічною моделлю для виявлення об'єктів на основі глибокого навчання.

В дослідження використовується модель *You Only Look Once (YOLO)* [14]. Вона дозволяє напряму прогнозувати обмежувальні рамки та ймовірності класу з використанням єдиної мережі в межах однієї оцінки. Аналогічно моделі YOLO, була розроблена модель *Single-Shot Detector-SSD* [15], щоб одночасно передбачати всі обмежувальні рамки та ймовірності класів за допомогою єдиної CNN.

Мета статті зводиться до діагностики та аналізу розробленого програмного застосування для розпізнавання автомобільних номерних знаків та його захисту за допомогою мобільних пристроїв на платформі Android. Крім того, у дослідженні здійснено огляд сучасних інструментів для розробки, генерування, тренування, застосування та перевірки пропонованих архітектур нейронних мереж, що доступні у відкритому доступі; проаналізовано сильні та слабкі сторони цих фреймворків та бібліотек.

2. ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Абсолютна більшість сучасних нейромереж застосовується для аналізу зображень. Оскільки будь-яке відео являє собою набір впорядкованих зображень, його необхідно розбити на окремі кадри, або вибирати кадри через певні інтервали, до яких уже застосовуватиметься аналіз нейромережами.

Задача виявлення об'єктів є одночасно задачею регресії та класифікації. Перш за все, для оцінки просторової точності нам потрібно видалити рамки з низькою впевненістю. Перш за все, щоб оцінити просторову точність, потрібно видалити рамки з низькою впевненістю. Потім використовується ділянка перетину об'єднань рамок (Intersection of Union – IoU) зі значеннями від 0 до 1. Це відповідає зоні перекриття між прогнозним та реальним зображенням у відеопотоці. Чим вище IoU, тим краще прогнозоване розташування рамок для даного об'єкту. Зазвичай зберігаються всі обмежувальні коефіцієнти кандидатів з IoU більшим, ніж деякий поріг.



Рис. 1. Візуалізація роботи алгоритму вибіркового пошуку: на основі [10].

Результатом є декілька пропозицій відеопотоків, які можуть містити об'єкт шляхом об'єднання невеликих відеопотоків. На рис. 1 візуалізовано малі відеопотоки в зображенні та об'єднано їх за ієрархічним принципом. Тобто, остання група – це рамка, що містить все зображення.

Виявлені області (рис. 2) об'єднуються за різними проміжками кольорів та показниками подібності. Кожна пропозиція відеопотоку змінюється так, щоб відповідати входу CNN, з якого ми отримуємо векторні ознаки розміру 4096 bit. Вектор ознак надходить в декілька класифікаторів для отримання імовірностей, що належать до кожного класу.

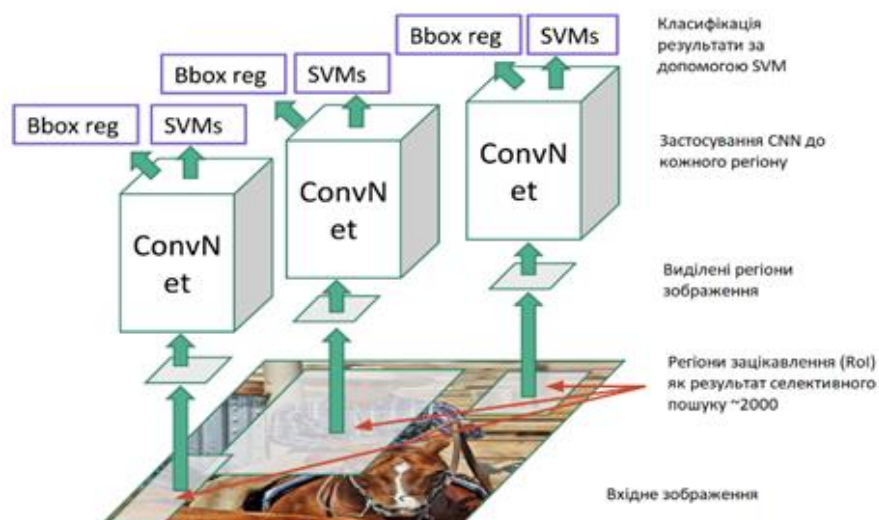


Рис.2. Типова структура роботи R-CNN: на основі [11].

Кожен з цих класів має класифікатор на основі SVM, навчений вивести імовірність виявлення цього об'єкта для певного вектору функцій. Сформовані вектори надходять у модуль лінійної регресії для коригування форм обмежувальної рамки для пропозиції регіону i , таким чином, зменшує помилку локалізації.

Завдяки *Fast-R-CNN* можна покращити швидкість виявлення через дві основні зміни: виконання функції виділення ознак зображення перед застосуванням алгоритму пошуку регіонів, таким чином, зображення проходить через CNN лише один раз; заміна SVM на шар softmax, що дозволило розширити можливості нейронної мережі для прогнозування замість створення та тренування окремих SVM-моделей. Залишалось лише одне проблемне вузьке місце у Fast R-CNN: вибіркового алгоритму пошуку для створення регіонів пропозицій. Але на заміну вибіркового пошуку прийшла Region Proposal Network (RPN), яка напряду генерувала відеопотоки пропозицій, передбачала обмежуючі рамки та виявляла об'єкти. Модель CNN приймає в якості вхідного параметру ціле зображення та створює карти фільтрів (ознак).

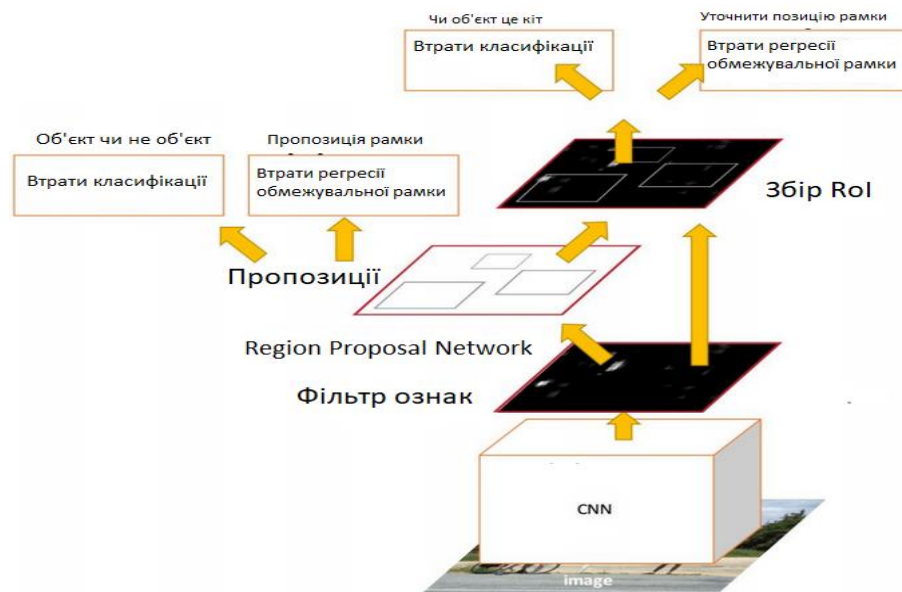


Рис.3. Схема роботи *FasterR-CNN*: на основі [12, 13].

Вікно розміру 3×3 переміщує всі фільтри ознак та виводить вектор ознак, що передається до двох повністю зв'язаних шарів (Fully connected – FC), один для корекції регресії та один для класифікації рамок. Передбачення для запропонованих областей відбувається FC шарами. *Faster R-CNN* використовує RPN, щоб уникнути вибіркового методу пошуку. Це прискорює процес навчання та тестування та покращує ефективність. RPN може використовувати попередньо підготовлену модель над набором даних Image Net для класифікації, і вона точно налаштована на масив даних з VOC PASCAL (Рис. 4).

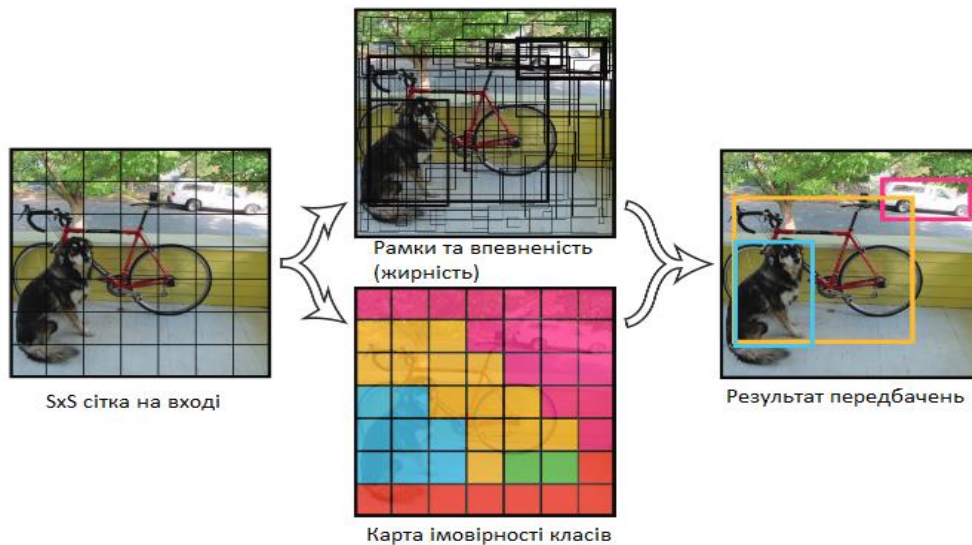


Рис.4. Модель візуалізації роботи передбачень YOLO: на основі [14].

Простота моделі YOLO дозволяє створювати прогнози в реальному часі. Спочатку модель приймає зображення як вхідний параметр. Вона розділяє його на сітку розміром $S \times S$. Кожна клітинка цієї сітки прогнозує в обмежувальних рамках з оцінкою впевненості для кожної клітинки. Використана у YOLO CNN створена на основі GoogLe Net, яка складається з Insertion модулів (Рис. 5).

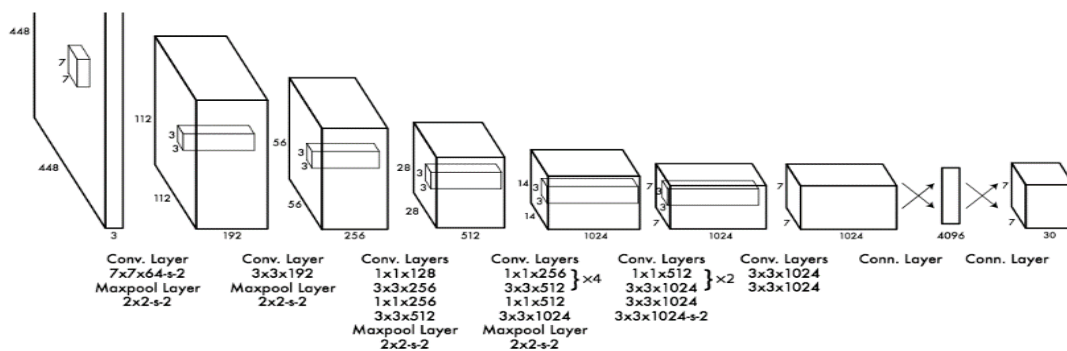


Рис.5. Схема CNN в YOLO: на основі [14].

Модель SSD приймає зображення на вході, яке проходить через декілька шарів з різними розмірами фільтрів (10x10, 5x5 та 3x3). Особливі карти зі згорткових шарів в різних місцях мережі використовуються для прогнозування обмежувальних рамок. Вони обробляються специфічними згортковими шарами з фільтрами розміром 3x3, що називаються додатковими функціональними шарами, для створення набору обмежувальних рамок, подібних до якірних рамок у FastR-CNN.

Кожна рамка має 4 параметри: координати центру, ширина і висота. У той же час вона створює вектор імовірності, що відповідає впевненості належності об'єкта для кожного з класів (Рис. 6).

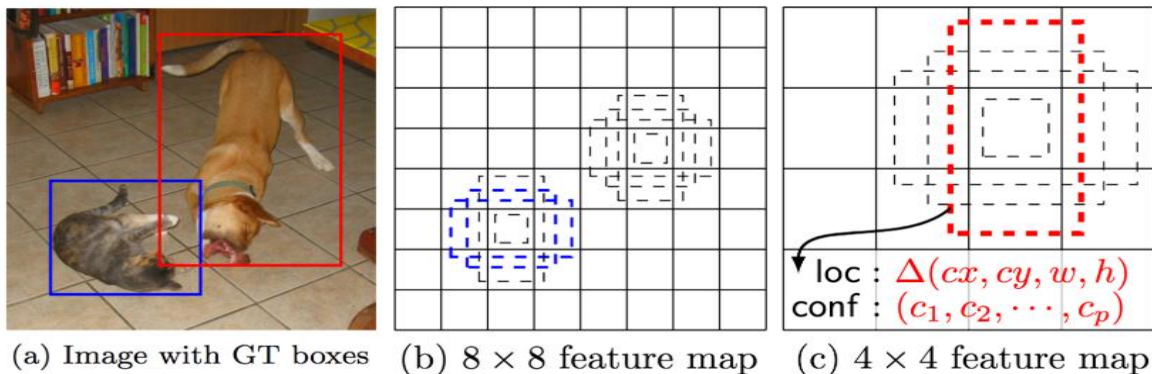


Рис. 6. Модель побудови передбачень у SSD: на основі [15].

В моделі SSD використовується метод немаксимального придушення (Non-Maximum Suppression) для зберігання найбільш відповідних обмежувальних коробок. Крім того, застосовується метод жорсткого видалення негативів для видалення негативних рамок. Найкращі коробки вибираються залежно від співвідношення між негативними і позитивними зображеннями, що не повинно перевищувати 1/3.

Оскільки використовується бібліотека Open CV алгоритм для детектування зображень може бути замінений на використання загортової нейромережі. Модуль оптимізації приймає на вході відео потоки зі зображеннями, виявлені на фазі детектування та виконує оптимізації цих зображень засобами Open CV: вирівнювання, поворот, збільшення контрастності, висвітлення та іними. Модуль оптичного розпізнавання тексту (OCR) використовує бібліотеку сучасну система для розпізнавання тексту на основі рекурентних нейромереж типу LSTM (longshort-termmemory). Вибрана система Tesseract 4.00. Tesseract 4.00 має відкритий вихідний код та відноситься до найкращих систем оптичного розпізнавання тексту. Для застосування у прикладній програмі розпізнавання номерних знаків було обрано нами середовище Open ALPR через спеціалізацію та можливість модифікації з мовою програмування C#.

3. МЕТОДИКА ДОСЛІДЖЕННЯ

Пропонована згортокова нейромережа може бути застосована для розпізнавання різнотипних образів з різних предметних областей. Однак найбільш показовим і актуальним є використання його при розпізнаванні критичних ситуацій для транспортних систем по відеозображення, що надходять з камер UGV.

На рис. 7 приведена схема інтелектуальної розподіленої міської системи транспортної безпеки, що містить безліч камер відеоспостереження. пропонується використовувати обчислювальний модуль NVidia Jetson TX2 для мобільного пристрою з GPS.

Розглянемо особливості застосування згортової нейромережі для розпізнавання зображень номерних знаків у відеопотоці.

1. Виділення з безперервного відеопотоку $V = (I_1, \dots, I_t, \dots, I_r)$ кадру I_t розміром $\omega^{I_t} \times h^{I_t}$, де t - номер поточного кадру.



Рис. 7. Структурна схема інтелектуальної розподіленої міської системи транспортної безпеки.

2. Пошук об'єктів на кадрі $f_1 : I_t \rightarrow G_t$, де G_t - масив елементів, з-тримає параметри n об'єктів в кадрі відеопотоку. При необхідності уточнення додаткових властивостей шуканого об'єкта о виконується перехід до наступного етапу, в іншому випадку береться наступний кадр. Створюється структура масиву G_t . При наявності об'єкта «людина» необхідно виконати ідентифікацію особистості по обличчю. Незважаючи на те, що модель YOLO розрізняє кілька класів транспортних засобів, потрібно розпізнавати його орієнтацію щодо камери.

3. Виділення області інтересу першого рівня $R^{(1)} = crop(I_t, x^0, y^0, \omega^0, h^0)$, де, x^0, y^0 - координати центру 0-го об'єкта, ω^0, h^0 - його розміри, $crop$ - операція вирізання з підматриці I_t по координатах $(x^0 - \omega^0 / 2, y^0 - h^0 / 2), (x^0 + \omega^0 / 2, y^0 + h^0 / 2)$.

4. Уточнення області інтересу для деталізації інформації про зображення $f_2 : R^{(1)}, t \rightarrow R^{(2)}$. При знаходженні особи на зображенні як алгоритм локалізації може бути використаний алгоритм HOG або SSD MultiBox. Тоді результатом роботи алгоритму є матриця $2R$, яка містить зображення обличчя людини.

5. Виконання попередньої обробки області інтересу $R^{(2*)} = f_3(R^{(2)}, M, g)$, де M - матриця геометричних лінійних і афінних перетворень $R^{(2)}$, g - набір матричних функцій та їх параметрів для яскравості і контрастних перетворень $R^{(2)}$. Результатом афінних перетворень є матриця $R^{(2*)}$.

6. Виділення інформативних ознак $R^{(2*)}$ шляхом кодування кадру глибокої нейронною мережею зі згортковими шарами: $\Phi^{CNN} : R^{(2*)} \rightarrow \tilde{X}$, де \tilde{X} - область інтересу, транспонована в просторі ознак CNN. Як простору ознак для зображень пропонується використовувати ознаки, отримані з модифікованої архітектури нейромережі MobileNet v2 без двох останніх шарів або Inception v3 для відеопотоків.

7. Віднесення вектора ознак до одного з класів $f_4: \tilde{X} \rightarrow p_{\tilde{X}}$, де $p_{\tilde{X}}$ - вектор розміром $c \times 1$, що містить ймовірності класифікації; c - кількість класів.

8. За параметрами об'єктів G_t , визначаємо штатну або критичну ситуацію у відповідності до:

$$s = F_4(G_t) = \begin{cases} 1, \text{ якщо } G_t \in D^{normal} \\ 0, \text{ якщо } G_t \notin D^{normal} \end{cases} \quad (1)$$

де $s = \{0,1\}$ - штатна або позаштатна ситуація. Наявність відхилення одного з параметрів об'єктів від закладених в D^{normal} можливих комбінацій значень переводить систему в режим критичної ситуації.

Для розпізнавання номерних знаків автомобілів пропонується використовувати архітектуру загорткових нейромереж, яка представлена на рис. 8.

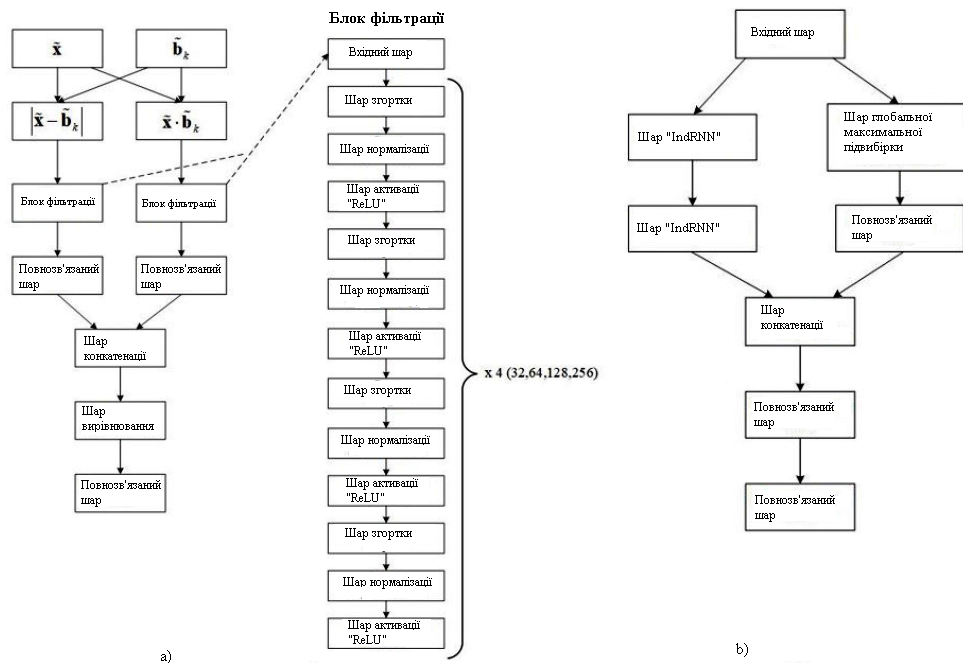


Рис. 8. Архітектури загорткових нейронних мереж: а) для розпізнавання номерних знаків зі згортковими шарами; б) для класифікації відеопотоків з рекурентними шарами.

В архітектурі нейромереж передбачено два входи: на перший вхід надходить еталон у вигляді вектора ознак \tilde{b}_k з обчисленням $h_1^1 = |\tilde{x} - \tilde{b}_k|$, на другий вхід надходить вектор ознак об'єкта \tilde{x} і обчислюється $h_1^2 = |\tilde{x} \times \tilde{b}_k|$.

Кожен з виходів h_1^1 і h_1^2 проходить через блок фільтрації, який являє собою послідовність шарів згортки ядром, нормалізації та активації.

Послідовність шарів повторюється чотири рази, при цьому розмір фільтра приймає в шарі згортки значення 32, 64, 128 і 256 пікселів. Після блоку фільтрації слід повнозв'язну шар розміром 128 нейронів, з функцією активації ReLU.

Як вихід нейромережа використовує повнозв'язаний шар з одним нейроном і функцією активації. Виходом пропонуваної глибокої нейронної мережі зі верточними

шарами є вектор $p_{\tilde{x}}$, що містить ймовірності схожості вхідного вектора і еталонів максимально схожого кластера B_v , інші елементи обнулюються.

В якості класифікатора відеофрагментів f_4 розроблена архітектура глибокої нейромережі зі згортковими та рекурентними шарами f_4^{event} (Нис. 8). Як рекурентні шари використовуються шари IndRNN.

Запропонований підхід був реалізований на мові Python з використанням бібліотеки Tensorflow і Keras. Як обчислювальної платформи застосовувався вбудований модуль GPU Nvidia Jetson TX2.

4. РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Проведемо експерименти та аналіз бібліотек для машинного навчання та роботи з відео потоком. На сьогоднішній день Python є основним інструментом для роботи у сфері машинного навчання. Для Python створено велику кількість бібліотек, більшість з яких має відкритий вихідний код. Хоча, на сьогоднішній день також існує велика кількість середовищ для існуючих популярних бібліотек, що дозволяє запускати та використовувати алгоритми машинного навчання практично на будь-якій платформі.

Автори тестують запропоновані алгоритми та перевіряють ефективність використовуваних архітектур на основі стандартних масивів даних. Розроблена авторами програма призначена для демонстрації застосування згорткових нейронних мереж для розпізнавання у відео потоці. Мобільний додаток в операційній системі Android дозволяє використовувати камеру смартфона для отримання відео потоку в реальному часі. З цього відео потоку виділяються окремі кадри, в яких нейромережа намагається детектувати автомобільні номерні знаки і у випадку детектування – розпізнати серію та номер.

Отримані результати можна побачити у вигляді спливаючого повідомлення в режимі реального часу. Результати детектування також серіалізуються у формат json та зберігаються у пам'яті пристрою. Запуску запропонованої програми здійснювався на смартфоні з камерою та операційною системою Android 7.0 (Рис. 9).

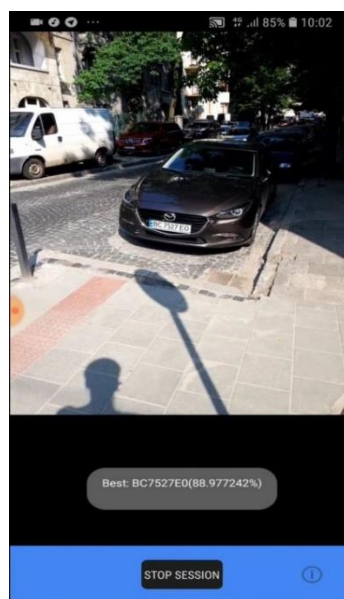


Рис. 9. Розпізнавання автомобільного номера розробленою програмою в ОС Android



Створену для Android програму можна з невеликими зусиллями перенести для використання на iOS. Для обробки та аналізу отриманих кадрів з відеопотоку нами використана бібліотека Open ALPR з модифікованим модулем детектування номерних знаків. Замість використання алгоритму лінійних бінарних шаблонів за допомогою методів Open CV імпортується збережена Tensor Flow модель згорткової нейромережі, яка використовується для знаходження зображень, що ймовірно містять автомобільний номер.

Для підготовки та тренування нейромережі авторами використано Tensor Flow Object Detection API. Для початку роботи була обрана модель SSDLite Mobilenetv2COCO, яка постачається як частина інструментарію Tensor Flow. Вона використовує архітектуру Single-shotdetector та попередньо навчена на наборі даних COCO. До її переваг можна віднести хорошу швидкість та її невеликий розмір (20 мегабайт), що не буде створювати великої затримки при завантаженні її у пам'ять на початковому етапі запуску програми. Оскільки набір даних COCO містить загалом 90 різних класів для класифікації, то ця модель мала 90 вихідних нейронів в останньому повністю з'єднаному шарі нейромережі. Для адаптації цієї моделі під детектування номерних знаків у конфігурації моделі нами було змінено кількість вихідних класів з 90 до 1. Крім того, для моделі була змінена схема розмітки вхідних даних для навчання, яка тепер містить лише один клас "LicensePlate".

При формуванні нейромережі були використані дані з декількох джерел:

1. Набір даних від онлайн-сервісу роботи з даними Supervisely, який складається з 10000 зображень, що розмічені та мають усі необхідні анотації для роботи з TensorFlow. Проте ці 10000 зображень номерних знаків є штучно синтезовані, тобто: до набагато меншої кількості зображень окремих номерних знаків були застосовані певні трансформації, а згодом ці зображення були поміщені у різні контексти.

2. Масив даних з фотографіями автомобілів та їхніх номерних знаків, що розповсюджується компанією Recognitory в їхньому власному інтернет-блозі, який складається з 9000 зображень без попередньої обробки. У масиві є велика кількість дублікатів та неякісних фотографій, тому він потребує додаткової ручної обробки.

3. Зображення автомобілів, які були авторами за допомогою смартфона з камерою шляхом фотографування. Загалом було оброблено 300 фотографій.

Розмітка та створення анотацій для власного масиву даних здійснювалась за допомогою інструментарію онлайн-сервісу обробки даних Recognitor. Навчання нейромережі реалізовано шляхом застосування методу Tensor Flow Object Detection API на комп'ютері з двома відеокартами NVIDIA GeForce GT 750m, що працювали у режимі SLI. На мобільному пристрої з процесором моделі Qualcomm Snapdragon 821 діагностовано розроблену систему і виявлено, що вона працює зі швидкістю 3 кадри в секунду. Цього може бути достатньо для задовільної роботи в умовах міської транспортної системи і для діагностування номерних знаків.

7. ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Інтернет речей вимагає дослідження сучасних методів машинного навчання, зокрема, згорткових нейронних мереж. Авторами проаналізовано методи та архітектуру згорткових нейромереж для детектування об'єктів. Було прослідковано еволюцію моделей нейромереж та методів, які дозволяють пришвидшити їх роботу та точність розпізнавання починаючи за рахунок SSD. Проведено аналіз функціональності



сучасних бібліотек для роботи з нейронними мережами та відео потоками, описані їхні переваги та недоліки, обґрунтовано вибір бібліотек для їх використання у подальших дослідженнях та впровадженні на практиці.

Описано та обґрунтовано вибір інструментів, які використовувались при реалізації запропонованих алгоритмів. Зображення з відео потоку камери захоплювалося за допомогою інструментів платформи Хагаїн для роботи з рендерингом інтерфейсу. Задача безпечного детектування номерного знаку на зображенні була вирішена за допомогою системи інтерфейсів бібліотеки Open ALPR та натренованої згорткової мережі типу SSD. Розпізнавання тексту номерного знаку виконувалося за допомогою системи OCR Tesseract 4.00. Отримані результати розпізнавання у форматі json можуть зберігатися в пам'яті пристрою при завершенні роботи програми.

Запропонована методика і розроблена програма може засовуватися органами патрульної поліції для автоматичного аналізу транспортних потоків, відслідковування автомобілів порушників чи автомобілів, що знаходяться у розшуку.

Точність методу при вирішенні задачі розпізнавання номерних знаків склала 0,982 на мобільному пристрої при часі обробки одного кадру, який отримано з відеопотоку, не більше 0,05 с. При вирішенні задачі розпізнавання хаотичних зображень із застосуванням пропонованого обчислювального методу була досягнута точність до 80% і швидкість 1,43 с для відео потоку розміром в 30 с на обчислювальному модулі з GPU.

Застосування згорткових нейромереж з рекурентними шарами і обчисленням на сучасних графічних прискорювачах дозволяють досягти хороших результатів у режимі реального часу при вирішенні задач розпізнавання та ідентифікації зображень. За рахунок застосування операції згортки істотно знижується кількість параметрів, що настроюються в порівнянні з традиційними нейромережами, а чергування згортальних шарів дозволяє вибудувати ієрархію ознак, з досить високою швидкістю і точністю розпізнавання номерних знаків у відеопотоці. Використання рекурентних шарів, в свою чергу, дає можливість виявити тимчасові залежності і ознаки, що характеризують критичні ситуації.

В подальшому планується застосування загорткових нейромереж для розпізнавання та передачі хаотичних сигналів і зображень між пристроями Інтернету речей в безпроводних мережах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] J. Carreira and C. Sminchisescu, "CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1312-1328, July 2012. DOI: 10.1109/TPAMI.2011.231 (in English).
- [2] P. Sermanet, K. Kavukcuoglu, S. Chintala, Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning", *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 3626-3633, 2013. DOI: 10.1109/CVPR.2013.465 (in English).
- [3] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Neural Information Processing Systems*. 25, 2012, pp. 1097-1105. DOI: 10.1145/3065386 (in English).
- [4] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, abs/1409.1556, 2014. DOI: 10.1.1.740.6937 (in English).
- [5] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9. DOI 10.1109/CVPR.2015.7298594 (in English).



- [6] He Kaiming, Z. Xiangyu, R. Shaoqing, "Deep Residual Learning for Image Recognition," *In Proceedings of IEEE conference on computer vision and pattern recognition*, 2016, p. 770–778. DOI: 10.1109/CVPR.2016.90 (in English)
- [7] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 580-587. DOI: 10.1109/CVPR.2014.81 (in English)
- [8] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers *et al.*, "Selective Search for Object Recognition," *Int J Comput Vis* 104, 2013, pp. 154–171. DOI: 10.1007/s11263-013-0620-5 (in English)
- [9] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016.. DOI: 10.1109/TPAMI.2015.2437384 (in English)
- [10] C. Chen, Q. Chen, Q. Huaqi, T. Giacomo, D. Jinming, B. Wenjia, R. Daniel, "Deep Learning for Cardiac Image Segmentation: A Review," *Frontiers in Cardiovascular Medicine*, V. 7, 2020, p. 25. DOI: 10.3389/fcvm.2020.00025
- [11] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440-1448. DOI: 10.1109/ICCV.2015.169. (in English)
- [12] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017. DOI: 10.1109/TPAMI.2016.2577031 (in English)
- [13] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788. DOI: 10.1109/CVPR.2016.91 (in English)
- [14] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," *In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham, 2016. DOI: 10.1007/978-3-319-46448-0_2 (in English)
- [15] T.-Yi Lin, M. Maire Serge, B. James, H. Pietro, P. Deva, R. Piotr, C. Dollár, L. Zitnick., "Microsoft COCO: Common Objects in Context." *European Conference on Computer Vision. ECCV*, Computer Vision – ECCV, 2014, pp. 740-755. DOI: 10.1007/978-3-319-10602-1_48 (in English)

**Danylo Svatiuk**

Senior Software Engineer
ELEKS, Lviv, Ukraine
ORCID: 0000-0002-6006-4810
danylosva@gmail.com

Oksana Svatiuk

Ph.D. in Economics, Associate professor,
Associate professor of Department of Management and Socio-Humanities Department
Lviv Educational and Scientific Institute School of Business Administration «Banking University», Lviv,
Ukraine
ORCID: 0000-0003-0099-2532
svatuk@gmail.com

Oleksandr Belei

Ph.D. in Economics, Associate professor, Associate professor of Department of Computer-Aided Design
Lviv Polytechnic National University, Lviv, Ukraine
ORCID: 0000-0003-4150-7425
Oleksandr.I.Belei@lpnu.ua

APPLICATION OF THE CONVOLUTIONAL NEURAL NETWORKS FOR THE SECURITY OF THE OBJECT RECOGNITION IN A VIDEO STREAM

Abstract. The article is devoted to analyzing methods for recognizing images and finding them in the video stream. The evolution of the structure of convolutional neural networks used in the field of computer video flow diagnostics is analyzed. The performance of video flow diagnostics algorithms and car license plate recognition has been evaluated. The technique of recognizing the license plates of cars in the video stream of transport neural networks is described. The study focuses on the creation of a combined system that combines artificial intelligence and computer vision based on fuzzy logic. To solve the problem of license plate image recognition in the video stream of the transport system, a method of image recognition in a continuous video stream with its implementation based on the composition of traditional image processing methods and neural networks with convolutional and periodic layers is proposed. The structure and peculiarities of functioning of the intelligent distributed system of urban transport safety, which feature is the use of mobile devices connected to a single network, are described.

A practical implementation of a software application for recognizing car license plates by mobile devices on the Android operating system platform has been proposed and implemented. Various real-time vehicle license plate recognition scenarios have been developed and stored in a database for further analysis and use. The proposed application uses two different specialized neural networks: one for detecting objects in the video stream, the other for recognizing text from the selected image. Testing and analysis of software applications on the Android operating system platform for license plate recognition in real time confirmed the functionality of the proposed mathematical software and can be used to securely analyze the license plates of cars in the scanned video stream by comparing with license plates in the existing database. The authors have implemented the operation of the method of convolutional neural networks detection and recognition of license plates, personnel and critical situations in the video stream from cameras of mobile devices in real time. The possibility of its application in the field of safe identification of car license plates has been demonstrated.

Keywords: Neural Networks; Rolling Networks; Security; Video Streaming; Mobile Applications.

REFERENCES

- [1] J. Carreira and C. Sminchisescu, "CPMC: Automatic Object Segmentation Using Constrained Parametric Min-Cuts," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1312-1328, July 2012. DOI: [10.1109/TPAMI.2011.231](https://doi.org/10.1109/TPAMI.2011.231) (in English).



- [2] P. Sermanet, K. Kavukcuoglu, S. Chintala, Y. LeCun, "Pedestrian detection with unsupervised multi-stage feature learning", *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 3626-3633, 2013. DOI: [10.1109/CVPR.2013.465](https://doi.org/10.1109/CVPR.2013.465) (in English).
- [3] A. Krizhevsky, I. Sutskever, G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Neural Information Processing Systems*. 25, 2012, pp. 1097-1105. DOI: [10.1145/3065386](https://doi.org/10.1145/3065386) (in English).
- [4] K. Simonyan, A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, abs/1409.1556, 2014. DOI: [10.1.1.740.6937](https://doi.org/10.1.1.740.6937) (in English).
- [5] C. Szegedy *et al.*, "Going deeper with convolutions," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9. DOI [10.1109/CVPR.2015.7298594](https://doi.org/10.1109/CVPR.2015.7298594) (in English).
- [6] He Kaiming, Z. Xiangyu, R. Shaoqing, "Deep Residual Learning for Image Recognition," *In Proceedings of IEEE conference on computer vision and pattern recognition*, 2016, p. 770-778. DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90) (in English)
- [7] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 580-587. DOI: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81) (in English)
- [8] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers *et al.*, "Selective Search for Object Recognition," *Int J Comput Vis* 104, 2013, pp. 154-171. DOI: [10.1007/s11263-013-0620-5](https://doi.org/10.1007/s11263-013-0620-5) (in English)
- [9] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016. DOI: [10.1109/TPAMI.2015.2437384](https://doi.org/10.1109/TPAMI.2015.2437384) (in English)
- [10] C. Chen, Q. Chen, Q. Huaqi, T. Giacomo, D. Jinming, B. Wenjia, R. Daniel, "Deep Learning for Cardiac Image Segmentation: A Review," *Frontiers in Cardiovascular Medicine*, V. 7, 2020, p. 25. DOI: [10.3389/fcvm.2020.00025](https://doi.org/10.3389/fcvm.2020.00025)
- [11] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, 2015, pp. 1440-1448. DOI: [10.1109/ICCV.2015.169](https://doi.org/10.1109/ICCV.2015.169). (in English)
- [12] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 1 June 2017. DOI: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031) (in English)
- [13] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788. DOI: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91) (in English)
- [14] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," *In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham, 2016. DOI: [10.1007/978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2) (in English)
- [15] T.-Yi Lin, M. MaireSerge, B. James, H. Pietro, P. Deva, R. Piotr, C. Dollár, L. Zitnick., "Microsoft COCO: Common Objects in Context." *European Conference on Computer Vision. ECCV*, Computer Vision – ECCV, 2014, pp. 740-755. DOI: [10.1007/978-3-319-10602-1_48](https://doi.org/10.1007/978-3-319-10602-1_48) (in English)

