

## ЗАДАЧА ВИЯВЛЕННЯ І КЛАСИФІКАЦІЇ ОБ'ЄКТІВ ЗА ДОПОМОГОЮ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Вінницький національний технічний університет

### Анотація

*Розглядається принцип роботи згорткових нейронних мереж, методи виявлення та класифікації об'єктів з використанням згорткових нейронних мереж, та їх особливості.*

**Ключові слова:** Згорткові нейромережі, ЗНМ, класифікація, виявлення, модель мережі.

### Abstract

*The principle of work of convolutional neural networks, methodical identification and classification of objects using stellar neural networks, and their features are considered.*

**Keywords:** Convolutional neural network, CNN, classification, detection, network model.

Комп'ютерне бачення - це міждисциплінарна сфера, яка за останні роки набирала величезний обсяг уваги (починаючи з CNN), а автомобілі, з автопілотом зайняли центральне місце. Іншою невід'ємною частиною комп'ютерного зору є виявлення об'єктів. Засоби виявлення об'єктів відповідно до позиції на зображенні, виявлення транспортних засобів, спостереження за допомогою відеокамер, тощо. Різниця між алгоритмами виявлення об'єктів і алгоритмами класифікації полягає в тому, що в алгоритмах виявлення ми намагаємося намалювати обмежувальну рамку навколо об'єкта, що цікавить, щоб знайти його в зображенні. Крім того, не обов'язково намалювати лише один обмежувальний блок у випадку виявлення об'єкта, може бути багато обмежувальних блоків, що представляють різні об'єкти, що представляють інтерес у зображенні, і невідомо скільки їх може бути на одному зображенні [1].

Згорткова нейронна мережа (скор. ЗНМ) є конкретним випадком нейронних мереж глибокого навчання, які вже використовувалися в кінці 90-х років, але які в останні роки стали надзвичайно популярними, коли досягли дуже вражаючих результатів у розпізнаванні зображень та вплинули на розвиток комп'ютерного зору.

Згорткові нейронні мережі мають багато спільних рис із іншими нейронними мережами: вони утворені нейронами, які мають параметри у вигляді ваг і біасів, які можна навчати. Але характерна особливість ЗНМ полягає в тому, що вони здатні добре обробляти зображення представлені у вигляді масивів даних, що дозволяє нам кодувати певні властивості в архітектурі для розпізнавання конкретних об'єктів у зображеннях [2].

Щоб отримати інтуїтивне уявлення про те, як працюють ці нейронні мережі, варто звернути увагу на те, як люди розпізнають речі. Наприклад, якщо ми бачимо обличчя, ми його розпізнаємо, тому що у нього є вуха, очі, ніс, волосся тощо. Тоді, щоб вирішити, чи є щось обличчям, ми робимо це так, ніби ми маємо деякі уявні зони перевірки характеристик, які ми беремо до уваги. Іноді можна не бачити вуха людини, тому що воно закрито волоссям, але ми також класифікуємо обличчя з певною вірогідністю, тому що ми бачимо очі, ніс і вуста. Власне, саме по такому принципу працюють класифікатори на основі згорткових нейронних мереж.

Але насправді, ми повинні спочатку знати, що таке вухо або ніс, щоб дізнатися, чи є вони на зображенні; це означає, що ми повинні раніше ідентифікувати лінії, границі, контури або фігури, які схожі на ті, які містять вухо або ніс, які ми бачили раніше. І саме цьому довірені шари згорткової нейронної мережі.

Виявлення цих елементів недостатньо для того, щоб сказати, що щось є обличчям. Ми також повинні бути в змозі визначити, як частини обличчя стикаються один з одним, їх відносні розміри, тощо; інакше, обличчя не буде схоже на те, до чого ми звикли. Візуально можна представити, що перші шари згорткової мережі містять у собі набір контурів і форм, які в наступних шарах зв'язують окремі частини зображення і шукають чи відповідають розташування окремих частин зображення бажаним (рис.1). Кожен шар відповідає за свій рівень абстракції.

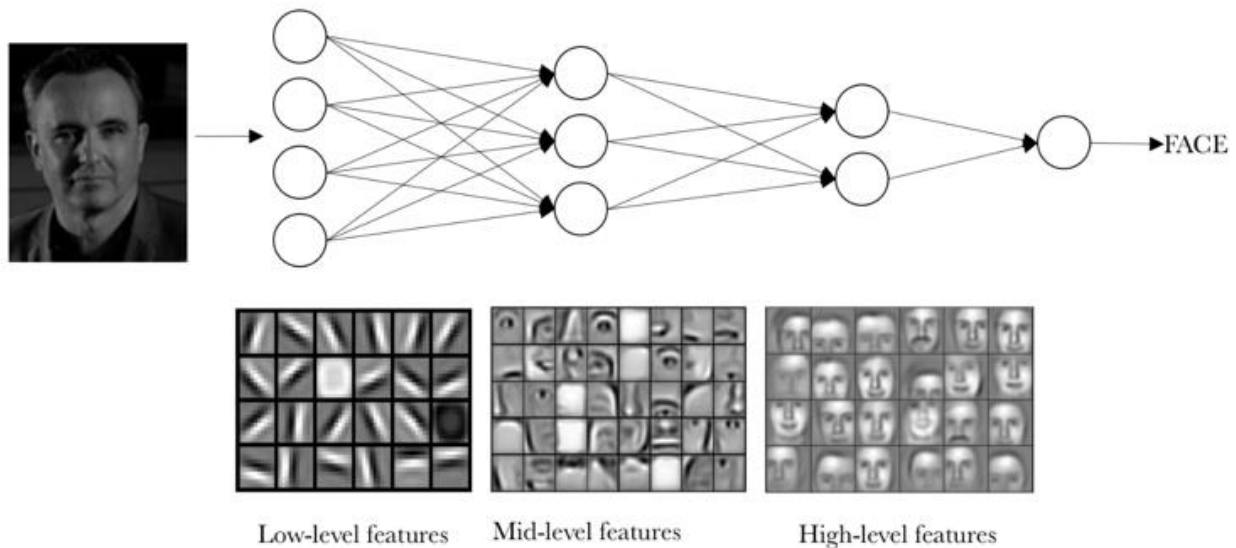


Рисунок 1 – Візуальне представлення перших шарів згорткової нейронної мережі

Але не можна вирішити проблему виявлення і класифікації, побудувавши стандартну згорткову мережу, за якою слідє повністю з'єднаний шар, бо довжина вихідного шару є змінною - не постійною тому, що число об'єктів, що може входити в зображення є не фіксованим. Наївний підхід до вирішення цієї проблеми полягає в тому, щоб взяти з зображення різні області, що представляють інтерес, і використовувати ЗНМ для класифікації присутності об'єкта в межах даної області. Проблема цього підходу полягає в тому, що об'єкти, що представляють інтерес, можуть мати різні просторові розташування в межах зображення і різні співвідношення сторін. Отже, доведеться вибирати величезну кількість регіонів, і це може використовувати багато обчислювальної потужності. Таким чином, алгоритми, такі як R-CNN, YOLO і т.д., були розроблені, щоб знайти зменшити кількість обчислень і підвищити точність.

## R-CNN

Щоб обійти проблему вибору величезної кількості областей, Рос Гіршик та ін. запропонували метод, при якому ми використовуємо вибіркового пошуку, щоб витягти з зображення лише 2000 областей, і він назвав їх регіональними зонами інтересу [3]. Тому зараз, замість того, щоб класифікувати величезну кількість регіонів, можна працювати з 2000 регіонами. Ці 2000 регіональних пропозицій генеруються з використанням алгоритму вибіркового пошуку:

1. Створити початкову підсегментацію, згенерувати багато регіонів-кандидатів;
2. Використовувати жадібний алгоритм для рекурсивного об'єднання подібних регіонів у великі;
3. Використовувати згенеровані регіони для створення кінцевих пропозицій регіону-кандидата.

Ці 2000 пропозицій регіонів-кандидатів викривляються у квадрат і надходять у згортку нейронної мережі, яка виробляє 4096-мірний вектор ознак як вихідний. ЗНМ виступає в якості екстрактора ознак і вихідний щільний шар складається з функцій, витягнутих з зображення, і витягнуті ознаки подаються в SVM для класифікації присутності об'єкта в цій пропозиції регіону-кандидата. На додаток до передбачення присутності об'єкта в межах пропозиції регіону, алгоритм також передбачає чотири значення, які є значеннями зміщення для підвищення точності обмежувального регіону. Наприклад, з урахуванням пропозиції регіону, алгоритм передбачав би присутність людини, але обличчя цієї особи в межах цієї пропозиції регіону можна було б скоротити наполовину.

Основні проблеми алгоритму R-CNN:

-Треба багато часу для навчання мережі, оскільки доводиться класифікувати 2000 регіонів по кожному зображенню.

-Метод не може бути реалізовано в реальному часі, оскільки для кожного тестового зображення потрібно близько 47 секунд.

-Алгоритм вибіркового пошуку є фіксованим алгоритмом. Тому на даному етапі навчання не відбувається. Це може призвести до формування поганих пропозицій регіонів-кандидатів.

### **Fast R-CNN**

Автор попередньої роботи (R-CNN) вирішив деякі недоліки R-CNN для побудови більш швидкого алгоритму виявлення об'єктів, що називається швидким R-CNN (Fast R-CNN). Підхід аналогічний алгоритму R-CNN. Але, замість того, щоб подавати регіональні пропозиції до ЗНМ, ми подаємо до мережі вхідне, щоб створити згорткову карту властивостей. На карті згорткових характеристик ми ідентифікуємо області пропозицій і деформуємо їх у квадрати і використовуємо шар об'єднання (RoI), ми їх змінюємо у фіксований розмір, щоб він міг надходити у повністю пов'язаний шар (Fully connected layer). З вектора особливості RoI ми використовуємо шар softmax для прогнозування класу запропонованої області, а також значень зміщення для обмежувального регіону.

Причина того, що «швидкий R-CNN» швидше, ніж R-CNN, в тому, що ви не треба щоразу подавати пропозиції регіону до згорткових нейронних мереж. Замість цього операція згортки виконується тільки один раз на зображення, і з неї генерується карта об'єктів.

### **Faster R-CNN**

Обидва вищезгадані алгоритми (R-CNN & Fast R-CNN) використовують вибіркового пошуку, щоб знайти пропозиції регіону. Вибірковий пошук є повільним і трудомістким процесом, що впливає на продуктивність мережі. Тому Shaoqing Ren et al. придумав алгоритм виявлення об'єктів, який виключає алгоритм вибіркового пошуку і дозволяє мережі вивчати пропозиції регіону.

Подібно до швидкого R-CNN, зображення надається в якості вхідного сигналу для згорткової мережі, яка забезпечує згортку. Замість використання алгоритму вибіркового пошуку на карті об'єктів для ідентифікації пропозицій регіону використовується окрема мережа для прогнозування пропозицій регіону. Прогнозовані області пропозицій потім змінюють, використовуючи шар об'єднання RoI, який потім використовують для класифікації зображення в межах запропонованої області і прогнозування значень зміщення для обмежувальних полів.

### **YOLO (You Only Look Once)**

Всі попередні алгоритми виявлення об'єктів використовують області для локалізації об'єкта в зображенні. Мережа не дивиться на повне зображення. Замість цього, частини зображення, які мають високі ймовірності, містять об'єкт. YOLO або You Only Look Once - це алгоритм виявлення об'єктів, який значно відрізняється від алгоритмів, що базуються на виборі випадкових регіонів. У YOLO єдина згорткова мережа передбачає обмежувальні рамки і вірогідність класу для цих регіонів [4].

Зображення розбивається на сітку  $S \times S$ , в межах кожної сітки беремо  $m$  обмежувальні поля. Для кожного обмежувального поля мережа виводить значення ймовірностей класу і зміщення для обмежувального поля. Обмежувальні рамки, що мають ймовірність класу вище порогового значення, вибираються і використовуються для знаходження об'єкта в зображенні.

YOLO на порядок швидше (45 кадрів в секунду), ніж інші алгоритми виявлення об'єктів. 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.

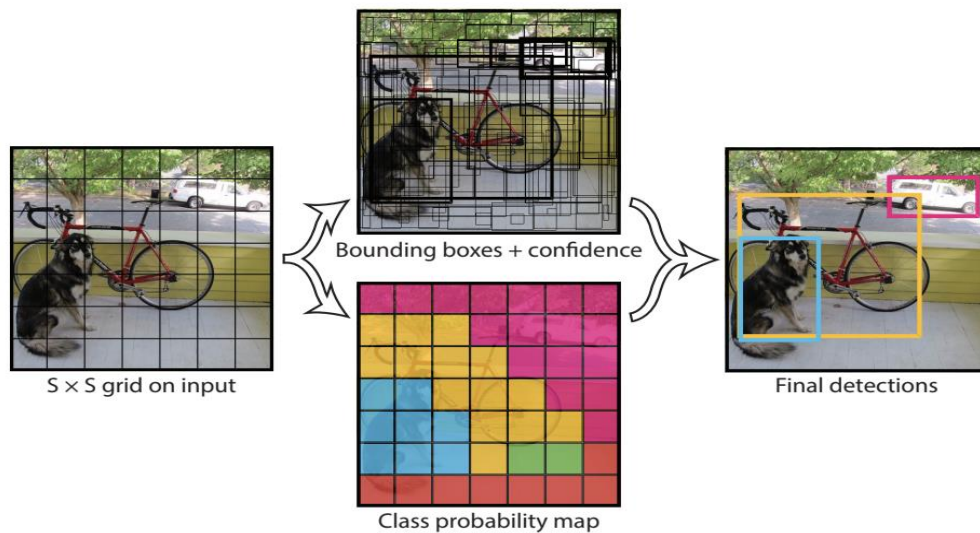


Рисунок 2 – Візуалізація розбиття зображення на зони методом YOLO

Конференції з комп'ютерного зору кожен рік переглядають нові радикальні концепції, і щороку згорткові мережі все ширше використовуються для вирішення задач комп'ютерного зору.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Колесницький О.К. Моделювання імпульсної нейронної мережі у задачі розпізнавання багатовимірних імпульсних послідовностей / О.К. Колесницький, С.М. Богатчук, М.В.Крещенецька, С.С.Яремчук // Вісник Вінницького політехнічного інституту. — Вінниця: УНІВЕРСУМ-Вінниця, 2008. — №5, С.62-66.
2. Convolutional Neural Networks — [Електронний ресурс]. — Режим доступу: <https://towardsdatascience.com/convolutional-neural-networks-for-beginners-practical-guide-with-python-and-keras-dc688ea90dca>
3. 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>
4. YOLO: Real-Time Object Detection — [Електронний ресурс]. — Режим доступу: <https://pjreddie.com/darknet/yolo/>.

**Мазур Максим Віталійович** — студент групи ІКН-18м, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця.

Науковий керівник — **Колесницький Олег Костянтинович** — канд. техн. наук, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет.

**Mazur Maksym V.** — student, faculty of information technology and computer engineering, Vinnitsa National Technical University, Vinnitsa.

Supervisor — **Kolesnytskyj Oleh K.** — Cand. tech Sciences, Associate Professor, Department of Computer Science, Vinnitsia National Technical University, Vinnitsia.