

УДК 004.8

Кусаков А.В.

асп. НУ «Запорізька політехніка»

**ОПТИМІЗАЦІЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДЛЯ ЗАДАЧ
РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ В УМОВАХ ОБМЕЖЕНИХ
РЕСУРСІВ**

Сучасний розвиток технологій комп'ютерного зору та глибокого навчання призвів до широкого застосування згорткових нейронних мереж (Convolutional Neural Networks, CNN) у задачах розпізнавання зображень.

Такі підходи активно використовуються у системах біометричної ідентифікації, розпізнавання облич, класифікації об'єктів, а також у мобільних та вбудованих додатках. Крім того, CNN застосовуються у системах відеоспостереження, автономному транспорті та медичній діагностиці. Завдяки здатності автоматично виділяти інформативні ознаки зображень CNN значно перевершують традиційні методи комп'ютерного зору, зменшуючи залежність від ручного конструювання ознак.

Ранні підходи до розпізнавання зображень базувалися на ручному виділенні ознак, таких як гістограми орієнтованих градієнтів (HOG), локальні бінарні шаблони (LBP) та інші дескриптори. Проте такі методи мали обмежену ефективність у складних умовах, зокрема при зміні освітлення, масштабу або фону, а також вимагали значних зусиль для налаштування. З появою глибоких нейронних мереж, зокрема архітектур VGG, ResNet та їх похідних, стало можливим досягнення значно вищих показників точності за рахунок автоматичного навчання багаторівневих представлень ознак. Водночас такі моделі характеризуються значною кількістю параметрів, високими вимогами до обчислювальних ресурсів та тривалим часом навчання. У зв'язку з цим виникає задача оптимізації архітектур згорткових нейронних мереж для забезпечення ефективної роботи в умовах обмежених обчислювальних ресурсів. Це особливо актуально для мобільних пристроїв, систем реального часу та задач із невеликими навчальними вибірками, де використання повноцінних глибоких моделей є недоцільним. Основною метою є зменшення складності моделі при збереженні достатнього рівня точності розпізнавання, а також забезпечення стабільності роботи моделі в різних умовах.

У даній роботі запропоновано підхід до побудови полегшеної архітектури CNN, заснований на модифікації класичних глибоких моделей. Основна ідея полягає у зменшенні глибини мережі, скороченні кількості згорткових фільтрів, а також спрощенні повнозв'язних шарів. Додатково застосовується використання операцій підвибірки (pooling) та функцій активації типу ReLU для збереження нелінійності моделі при зменшенні її складності. Також оптимізація включає зменшення розмірності вхідних ознак та раціональне розташування шарів, що дозволяє уникнути надлишкових обчислень. Такий підхід дозволяє істотно знизити кількість параметрів без суттєвої втрати якості моделі.

Процес роботи моделі включає етапи попередньої обробки зображень, навчання та тестування. На етапі попередньої обробки виконувалося масштабування зображень до єдиного розміру, нормалізація даних та усунення шумів. Додатково може застосовуватися аугментація даних для підвищення узагальнюючої здатності моделі. Навчання моделі здійснювалося на наборі зображень, що містить декілька десятків класів об'єктів із

варіаціями умов освітлення та фону. Для оцінки ефективності використовувався поділ вибірки на навчальну та тестову частини, що дозволяє об'єктивно оцінити здатність моделі до узагальнення.

Результати експериментального дослідження показали, що запропонована полегшена модель забезпечує точність класифікації на рівні близько 93–94 %, що є вищим за показники базової моделі (~78–80 %) в умовах обмеженого обсягу даних. Додатково було проаналізовано такі метрики, як precision, recall та F1-score, які також продемонстрували покращення або збереження на прийнятному рівні. Отримані результати свідчать про те, що зменшення складності моделі не призводить до суттєвої втрати якості, а в окремих випадках навіть сприяє кращій узагальнюючій здатності. Крім того, було зафіксовано зменшення часу навчання та зниження вимог до обчислювальних ресурсів, що є важливим фактором для практичного використання. Це дозволяє застосовувати запропонований підхід у системах із обмеженими апаратними можливостями без значного зниження ефективності. Отримані результати свідчать про доцільність використання оптимізованих архітектур CNN у практичних задачах комп'ютерного зору. Запропонований підхід може бути ефективно застосований у мобільних додатках, вбудованих системах та інших середовищах із обмеженими ресурсами.

Таким чином, оптимізація згорткових нейронних мереж дозволяє досягти компромісу між точністю та обчислювальною складністю, що є важливим фактором для подальшого розвитку інтелектуальних систем обробки зображень та їх широкого впровадження у практичні задачі.