

Міністерство освіти і науки України
Західноукраїнський національний університет
Факультет комп'ютерних інформаційних технологій
Кафедра інформаційно-обчислювальних систем та управління

Самостійна робота

З дисципліни: «Методи та системи штучного інтелекту»

на тему: «Класифікація зображень за допомогою згорткової нейронної мережі
(CNN) на наборі даних CIFAR-10»

Студента групи КНШІ-31
Крушельницький С.М.

Що таке CNN?

Convolutional Neural Networks (CNN), або згорткові нейронні мережі, — це тип глибоких нейронних мереж, спеціально розроблений для обробки зображень і розпізнавання об'єктів на них. CNN дозволяють комп'ютерам ідентифікувати різноманітні патерни, текстури та ознаки, що присутні в зображеннях, шляхом автоматичного виявлення важливих особливостей в кожному пікселі. Вони базуються на концепції згортки, де використовуються фільтри для пошуку характерних елементів зображення, що дозволяє моделі ефективно класифікувати зображення.

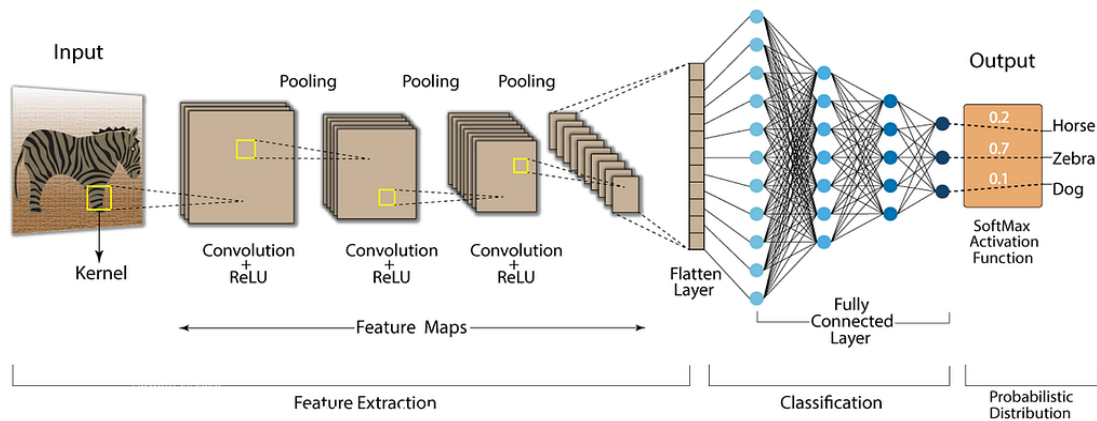


Рис. 1 – Вигляд CNN

Як працює CNN?

1. **Згортка (Convolution):** Це основна операція в CNN, яка використовує спеціальні фільтри (ядерні матриці) для обробки зображень. Фільтри виконують згортку над зображенням, виділяючи важливі ознаки, такі як контури, текстури та форми. Це дозволяє мережі виявляти основні характеристики зображення на різних рівнях абстракції.
2. **Канали (Channels):** Для кольорових зображень кожен канал (червоний, зелений, синій) обробляється окремо. Для кожного каналу застосовуються власні фільтри, що дозволяє створювати окремі карти ознак (feature maps) для кожного кольорового компонента. Це дає змогу мережі ефективно витягувати інформацію з кожного кольорового каналу і комбінувати їх для кращої класифікації.
3. **Крок (Stride) та Заповнення (Padding):** Крок визначає, як швидко фільтр переміщується по зображенню. Великі значення кроку зменшують розмір обробленого зображення. Заповнення допомагає зберегти просторову інформацію на краях зображення, додаючи пікселі до меж зображення, що дозволяє уникнути втрати важливих даних при згортці.
4. **Шари зменшення розмірності (Pooling Layers):** Пулінг зменшує розмірність карти ознак, зберігаючи лише найважливіші ознаки. Найбільш поширеним методом є максимальний пулінг, де вибирається

максимальне значення з кожної ділянки зображення, що дозволяє зменшити обсяг оброблюваних даних і підвищити ефективність мережі.

5. **Плоский шар (Flattening):** Після того, як вхідні дані пройшли через всі шари згортки та пулінгу, вони перетворюються в одновимірний вектор. Це дозволяє передати ці дані в пов'язані шари, де проводиться остаточний аналіз і класифікація на основі витягнутих ознак.

Побудова CNN

Модель CNN складається з кількох шарів, кожен з яких виконує певну роль у процесі аналізу та класифікації зображень:

- **Згорткові шари (Convolutional Layers):**
 - **Перший згортковий шар:** використовує 64 фільтри з ядром розміром 5×5 і $\text{padding}=2$ для збереження розмірності зображення. Цей шар виділяє початкові ознаки, такі як контури і текстури.
 - **Другий згортковий шар:** використовує 128 фільтрів з таким же ядром 5×5 і $\text{padding}=2$. Цей шар поглиблює аналіз, дозволяючи моделі виявляти більш складні патерни і форми.
- **Пов'язані шари (Fully Connected Layers):**
 - **Перший пов'язаний шар:** має 1000 нейронів і здійснює перехід до більш абстрактних ознак, що виявлені в згорткових шарах.
 - **Другий пов'язаний шар:** виконує класифікацію, визначаючи, до якого з 10 класів належить вхідне зображення.
- **Шар Dropout:** для запобігання перенавчанню використовується шар Dropout з параметром $p=0.5$, що випадковим чином "відключає" половину нейронів під час тренування, що допомагає мережі загальнити результати.
- **Функція активації:** після кожного згорткового шару застосовується функція активації ReLU (Rectified Linear Unit), яка додає нелінійність і дозволяє моделі ефективніше навчатися на складних даних.

Набір даних

CIFAR-10 — це популярний набір даних для задач комп'ютерного зору, який використовується для розпізнавання об'єктів на зображеннях.

- **Опис:** Набір складається з 60 000 кольорових зображень розміром 32×32 пікселів, поділених на 10 класів об'єктів:
 - Літак, автомобіль, птах, кішка, олень, собака, жаба, кінь, корабель, вантажівка.
- **Дослідники:** Цей набір даних був зібраний і досліджений відомими вченими, такими як Алекс Кріжевський, Вінод Наїр та Джеффри Хінтон.

CIFAR-10 є стандартом у сфері комп'ютерного зору, часто використовується для навчання і тестування моделей глибокого навчання.

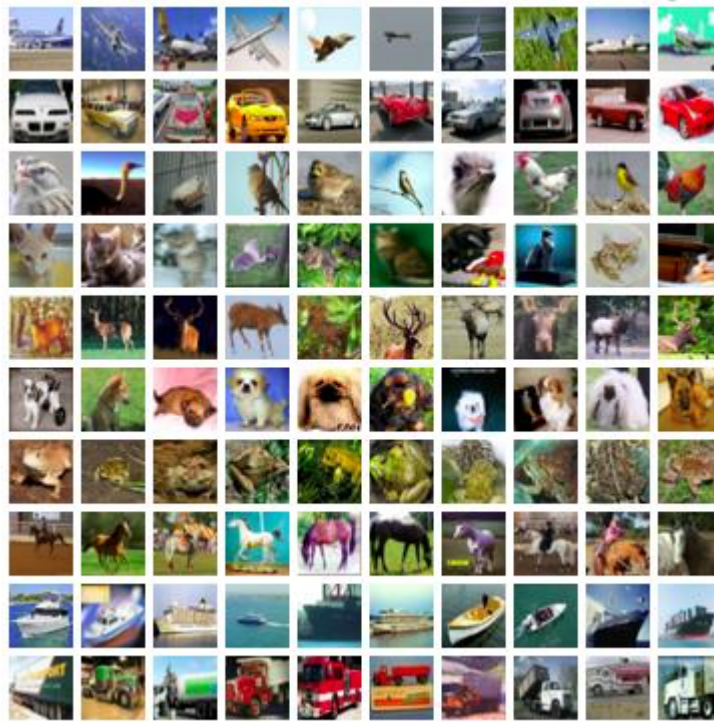


Рис. 2 – Як виглядає датасет

Результати

- **Перший графік:** Червона лінія відображає загальну втрату моделі протягом епох тренування, і ми спостерігаємо її зниження, що свідчить про поступове поліпшення здатності моделі до класифікації зображень. Синя лінія показує точність моделі на валідаційному наборі даних. Спочатку точність зростає, а потім стабілізується на рівні близько 0,75-0,8, що вказує на те, що модель досягла досить високої здатності до класифікації.

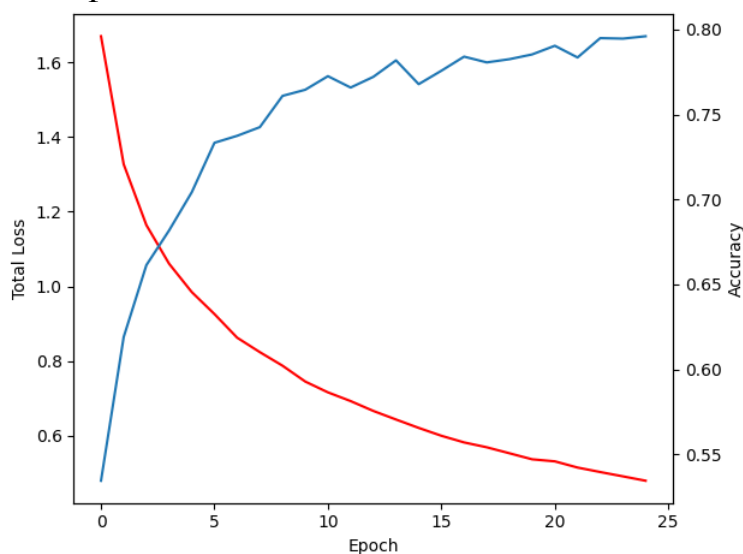


Рис. 3 – Перший графік

- **Другий графік:** Червона лінія вказує на втрати на тренувальному наборі даних, а зелена — на валідаційному. Після певної кількості ітерацій різниця між тренувальними та валідаційними втратами стає менш помітною, що свідчить про те, що модель починає краще узагальнюватися на нові дані.

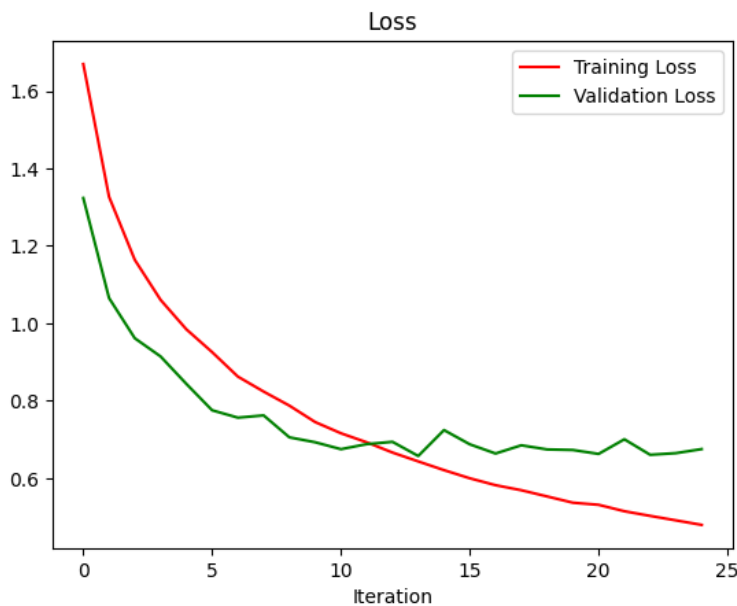


Рис. 4 – Другий графік

- **Максимальна точність:** Модель досягла максимальної точності 79,6%, що є хорошим результатом для даної задачі.

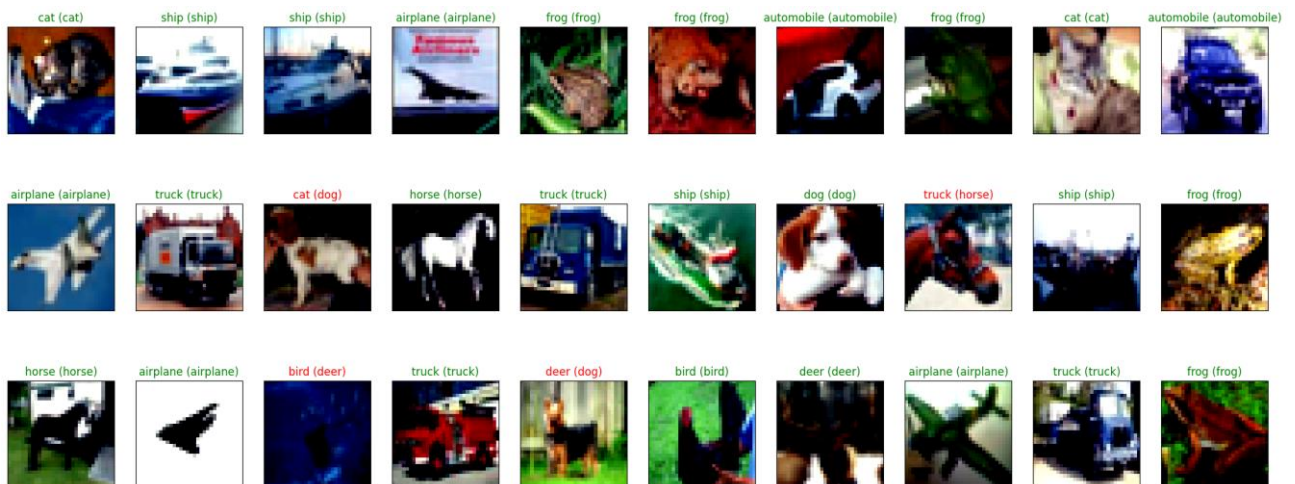


Рис. 5 – Передбачення моделі

Висновки з кореляційної матриці:

- **Загальна точність:** Модель має високу точність, оскільки більшість зображень класифіковані правильно, про що свідчать високі значення на діагоналі матриці плутанини.
- **Найпоширеніші помилки:** Модель часто плутає деякі класи, зокрема «кішку» та «собаку», «літак» і «корабель». Це може бути результатом

візуальної схожості між цими об'єктами, що ускладнює їхнє правильне розпізнавання.

- **Сильні класи:** Модель демонструє відмінні результати в класифікації деяких класів, таких як «автомобіль» та «жаба». Це підтверджується високими значеннями на діагоналі для цих класів, що свідчить про високу точність при їх класифікації.

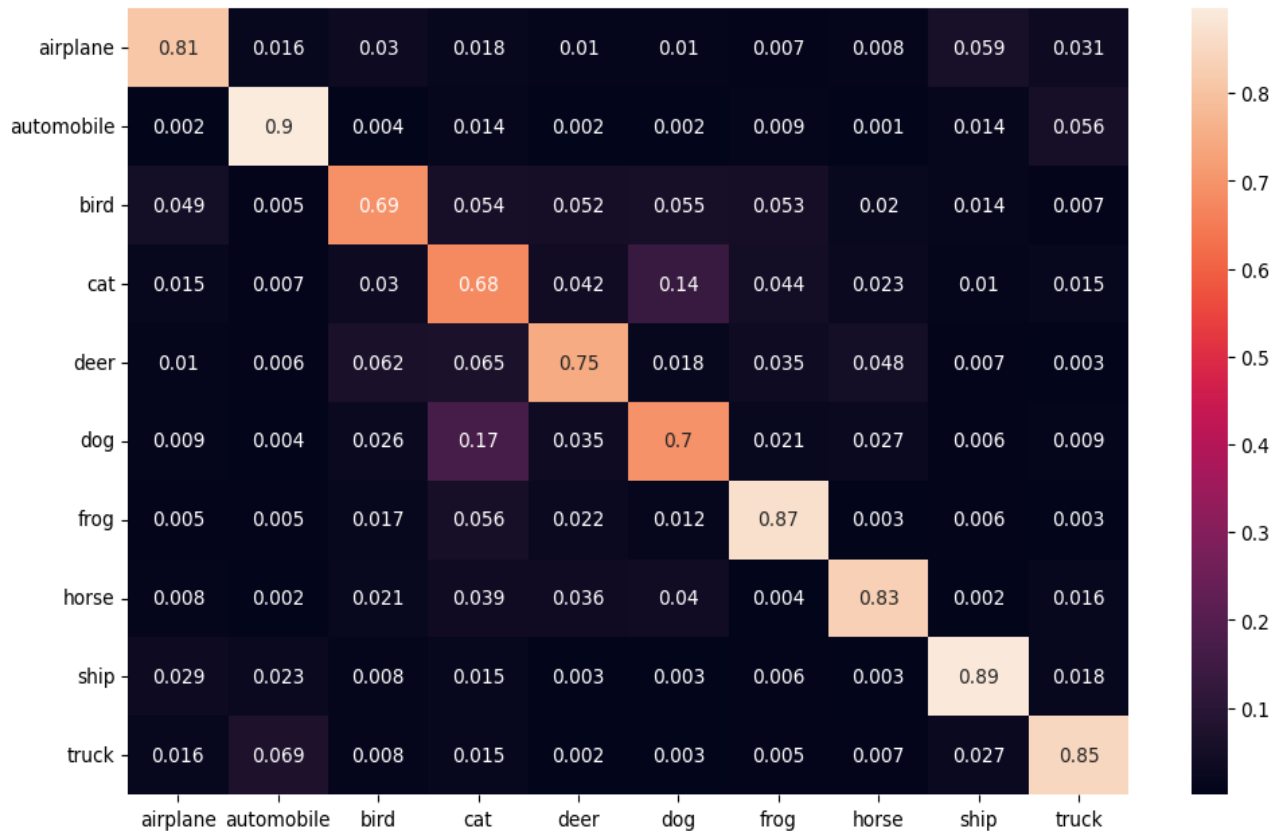


Рис. 6 – Кореляційна матриця

Висновок: Модель на основі згорткових нейронних мереж (CNN) досягла максимальної точності 79,6%, що є добрим результатом для класифікації зображень. Процес навчання показав зменшення загальних втрат і стабілізацію точності на рівні 75-80%. Модель добре класифікує більшість зображень, зокрема для класів «автомобіль» та «жаба», де точність висока. Проте, були проблеми з розпізнаванням схожих об'єктів, таких як «кішка» і «собака» або «літак» і «корабель». Це вказує на можливі складнощі при розрізненні візуально подібних класів.