МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
УМАНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ САДІВНИЦТВА

Факультет економіки і підприємництва

Кафедра інформаційних технологій

Допущено до захисту
"_____" _____ 2023 р.
Зав. кафедри, к.т.н., доцент
_____ Ліщук Р.І.

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

РОЗРОБКА ІНФОРМАЦІЙНОЇ ПІДСИСТЕМИ РОЗПІЗНАВАННЯ ЖЕСТІВ

Виконав – студент 2 курсу, групи 21к-кн
освітньої програми "Комп'ютерні науки"
спеціальності 122 Комп'ютерні науки
Михайленко Олександр Сергійович

Науковий керівник – к.т.н., доцент
Ліщук Роман Ігорович

Умань – 2023

ЗМІСТ

ВСТУП 4

РОЗДІЛ 1 МАТЕМАТИЧНА МОДЕЛЬ 5

1.1 Нейрони та базова будова нейронні мережі 5

1.2 Класифікація нейронних мереж за способом навчання

1.3 Архітектури нейронних мереж

1.4 Розрахунки для шару нейронної мережі

1.5 Функція активації

1.6 Функція втрат

1.7 Зворотне поширення помилки

1.8 Структура цифрових зображень

1.9 Ядро згортки

1.10 Шари згортки

1.11 Backpropagation згорткового шару

1.12 Pooling шари зі зворотним поширенням

1.13 Класифікація зображення з використанням згорткових нейронних мереж в Keras

РОЗДІЛ 2 РЕАЛІЗАЦІЯ МОБІЛЬНОГО ДОДАТКУ

2.1 Датасет MNIST для розпізнавання жестів

2.2 Побудова моделі CNN

2.3 Матриця помилок (confusion matrix)

2.4 Огляд роботи мобільного додатку

ВИСНОВКИ

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

ДОДАТКИ

ВСТУП

Актуальність розробки систем машинного перекладу жестових мов (від чує до глухому і навпаки) складається як в недостатньому кількості перекладачів жестових мов, так і не завжди бажаному посередництва (медицина, особистісні відносини і т.п.) при комунікаціях глухих і чуючих громадян. Програмні рішення в розпізнаванні мови жестів зустрічаються рідко, тому що практично немає загальнодоступних наборів даних.

У даній роботі розглядається проблема автоматичного розпізнавання сфотографованого жесту. Крім цього, необхідно вивчити існуючі рішення даної задачі, а також вибрати вхідні дані.

Для реалізації мети бакалаврської роботи необхідно вирішити задачі:

- Аналіз предметної області.
- Проектування та розробка інтерфейсу мобільного додатку.
- Моделювання та вибір налаштування згорткової нейронної мережі.
- Розробка нейронної мережі та інтеграція до мобільного додатку.
- Розробка нейронної мережі.
- Тестування мобільного додатку.

У результаті буде розроблено мобільний додаток для визначення мовного жесту на основі штучного інтелекту.

РОЗДІЛ 1 МАТЕМАТИЧНА МОДЕЛЬ

1.1. Нейрони та базова будова нейронні мережі

Вивчення і використання штучних нейронних мереж, почалося вже досить давно – на початку 20 століття, але по-справжньому широку популярність вони отримали дещо пізніше. Пов'язано це, в першу чергу, з тим, що стали з'являтися просунуті (для того часу) обчислювальні пристрої, потужності яких були досить великі для роботи зі штучними нейронними мережами. На даний момент можна легко змоделювати нейронну мережу середньої складності на будь-якому персональному комп'ютері.

Глибокі нейронні мережі зробили революцію в машинному навчанні. В останні роки ця техніка дозволила провести великий прогрес у розпізнаванні текстів, звуків, зображень та відео. Розуміння цих методів викликає питання у зв’язках між математикою та алгоритмікою.

Нейронні мережі - це алгоритми, які обчислюють із входу x (наприклад, зображення) вихід y. Цей результат найчастіше являє собою набір ймовірностей (чим ближче це число до 100%, чим більше це означає, що алгоритм впевнений у результаті передбачення). Ми також обмежуємо тільки зображення, але нейронні мережі також дуже ефективно розпізнають тексти або відео. Математично такий алгоритм визначає функцію f_w (тобто $y=f_w(x)$). Комп'ютерна програма, яка обчислює цю функцію, дуже проста: вона складається з послідовності декількох етапів, і кожен етап виконує елементарні обчислення (додавання, множення та максимум). Для порівняння, комп’ютерні програми, знайдені в операційній системі комп’ютера, набагато складніші. Але в чому велика різниця між «класичним» алгоритмом та нейронною мережею, а саме в тому, що нейронна мережа залежить від параметрів, які є вагою нейронів. Перед використанням нейрона мережі, ці ваги повинні бути модифіковані, щоб алгоритм зміг найкращим чином вирішити поставлену задачу. Це робиться з використанням математичних та алгоритмічних методів,

які будуть розглянуті далі. Цей процес називається «навчанням» нейронної мережі, і це вимагає багато часу, машинних обчислень та ресурсів [1].

Нейрон представляє з себе елемент, який обчислює вихідний сигнал (за певним правилом) із сукупності вхідних сигналів. Тобто основна послідовність дій одного нейрона така:

- Прийом сигналів від попередніх елементів мережі
- Комбінування вхідних сигналів
- Обчислення вихідного сигналу
- Передача вихідного сигналу наступним елементам нейронної мережі

Між собою нейрони можуть бути з'єднані абсолютно по-різному, це визначається структурою конкретної мережі. Але суть роботи нейронної мережі залишається завжди однією і тією ж. За сукупністю сигналів, що надходять на вхід мережі, на виході формується вихідний сигнал (або кілька вихідних сигналів). Тобто нейронну мережу можна представити у вигляді чорного ящика, у якого є входи і виходи. А всередині цього ящика знаходиться величезна кількість нейронів.

Штучна нейронна мережа побудована навколо біологічної основи. Будова первинної зорової кори відносно добре відома, і Хубель, і Візель здобули Нобелівську премію з фізіології за відкриття в 1962 р. Організації нейронів у перших кортикальних шарах. Таким чином, в надзвичайно спрощеному поданні мозку нейрони організовані шарами, кожен нейрон отримує інформацію з попереднього шару, виконує дуже просте обчислення і передає свій результат нейронам наступного шару. Однак слід пам’ятати, що це лише метафора та джерело натхнення: біологічні мережі мають набагато складніші зв’язки, а математичні рівняння, що ними керують, також є більш складними (їх відкрили Алан Ходжкін та Ендрю Хакслі в 1952 і за це вони виграли Нобелівську премію). Тому залишається складним точно зв'язати іноді дивовижні показники штучних нейронів із когнітивними можливостями мозку. Наприклад, методи навчання штучних мереж, сильно відрізняються від способу навчання дитини [2].

Детальний опис прикладу такої штучної мережі ми можемо бачити на рис. 1.1. Цей тип нейронів був запроваджений в 1943 р. МакКаллоком та Піттсом. Для спрощення він тут складається лише з двох шарів нейронів (перший шар між x та u, другий між u та y), але найефективніші мережі сьогодні можуть мати кілька десятків шарів, їх називають глибокими. У нашому прикладі входи x - це пікселі зображення. Зображення, як правило, містить мільйони пікселів, і на рисунку показана лише невелика їх кількість: справжня нейронна мережа є більш складною. Крім того, кожен піксель x, насправді складається з 3 значень (по одному для кожного основного кольору червоного, зеленого та синього).

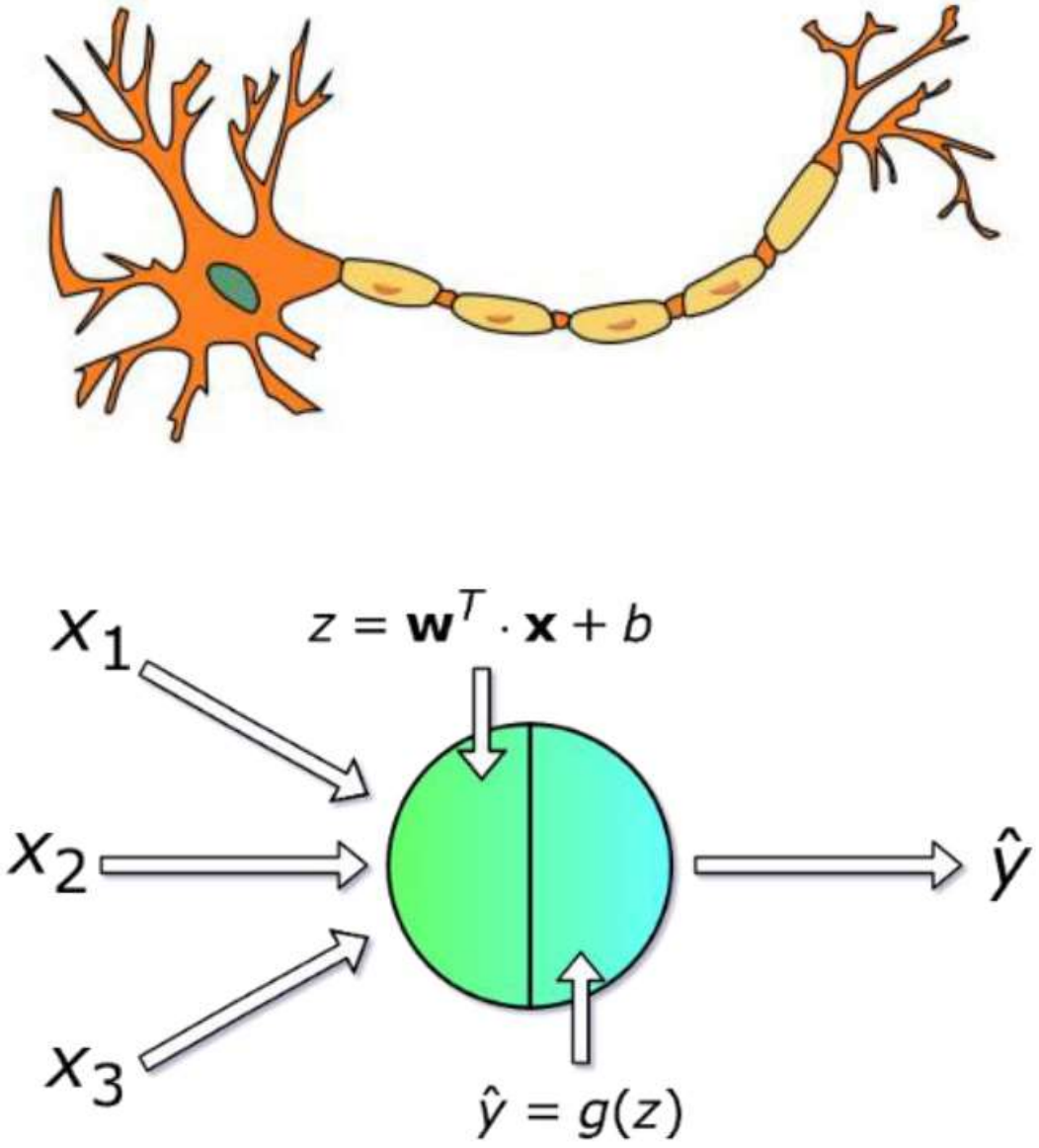


Рис. 1.1 Біологічні та штучні нейрони.

Перехід від одного шару (наприклад, x-шару входів) до іншого (наприклад, другого шару $\hat{y}$, який є “прихованим” шаром в центрі мережі) здійснюється через набір штучних нейронів. Штучний нейрон зображений на Рис. 1.1 обчислює вихідний сигнал $\hat{y}$. Цей нейрон з'єднує певну кількість елементів першого шару (тут три: x_1, x_2, x_3, але може бути і більше) з одним елементом другого, тому тут. Формула, за якою нейрон розраховує вихідний сигнал:

$$z=w_1x_1+w_2x_2+w_3x_3+\cdots+ w_nx_n=w^T*x \qquad (1.1)$$

Таким чином, нейрон виконує зважену суму трьох входів з трьома вагами w_1, w_2, w_3, і ми також додаємо w_3, яка є зміщенням. Потім нейрон обчислює максимум між цією сумою та нулем. Можна також використовувати функцію, відмінну від максимальної, але ця є найбільш популярною. Це порогова операція. Ми можемо порівняти це з біологічними нейронами, які передають інформацію чи або не передають її через нейрон, залежно від того, достатньо вони збуджені чи ні. Отже, якщо зважена сума $w_1x_1+w_2x_2+w_3x_3+\cdots+ w_nx_n$ менше 0, то нейрон повертає значення $\hat{y} = 0$, інакше він повертає значення цієї суми і записує її в $\hat{y}$.

Такі нейронні мережі були введені Розенблатом в 1957 році, який назвав їх "персептронами". Перші персептрони містили лише один шар. Такі одношарові архітектури занадто прості, щоб мати можливість вирішувати складні задачі. Завдяки додаванню декількох шарів ми можемо розрахувати більш складні функції. Таким чином, глибокі нейронні мережі використовують дуже велику кількість шарів. В останні роки ці архітектури дали змогу отримати дуже вражаючі результати для розпізнавання зображень та відео, а також для автоматичного перекладу текстів. Саме це дослідження глибинних мереж дозволило французькому досліднику Яну Ле Куну, а також Джеффрі Хінтону та Йошуа Бенджо отримати премію Тьюрінга в 2019 році. Ця премія вважається еквівалентом Нобелівської премії з інформатики.

Джордж Цибенко довів у 1989 р., що нейронна мережа fw з двома шарами може наближатись настільки точно, наскільки будь-яка безперервна функція $f*$ (тому в якомусь сенсі він може вирішити будь-яке завдання, представлене

невідомою функцією f^*, яка змогла б розпізнати об’єкти на будь-якому зображенні), доки розмір внутрішнього шару u (отже, кількість нейронів) довільно великий. Це не означає, що така мережа з двома рівнями працює добре на практиці. Щоб застосувати теорему Цибенка, необхідно мати потенційно нескінченну кількість навчальних даних, що на практиці далеко не так [3].

Кінцевою метою навчання є не мінімізація помилки навчання E_w а можливість якомога точнішого прогнозування нових даних. Маючи лише обмежений обсяг даних, ви не зможете навчати достатньо точно, а тому прогнози щодо невідомих даних у майбутньому будуть поганими. Функція f_w насправді буде дуже далекою від ідеальної функції f, яку хотілося б навчити.

Для того, щоб досягти найбільш точних прогнозів з обмеженою кількістю навчальних даних, ми, таким чином, шукаємо найбільш підходящі мережеві архітектури, які можуть ефективно вловлювати інформацію, що присутня в даних. Глибокі нейронні мережі (з великою кількістю шарів), але з відносно невеликою кількістю зв’язків між шарами, виявилися дуже ефективними для дуже «структурованих» даних, таких як текст, звук та зображення. Наприклад, для зображення пікселі мають взаємозв'язок між суміжними, і можна встановити конкретні зв'язки (архітектуру) і не з'єднувати нейрон з усіма іншими, а лише з сусідами (інакше зв'язків було б занадто багато). Крім того, можна припустити, що ваги, пов'язані з нейроном, є однаковими з вагами, пов'язаними з іншим нейроном. Цей тип мереж називають згортковими мережами. На даний момент не існує математичного аналізу, який би пояснив цю ефективність глибоких згорткових мереж. Тому існує потреба в нових математичних досягненнях, щоб зрозуміти поведінку та обмеження цих глибинних мереж.

1.2 Класифікація нейронних мереж за способом навчання

Класифікація нейронних мереж за характером навчання ділить їх на:

- нейронні мережі, що використовують навчання з вчителем;

- нейронні мережі, що використовують навчання без вчителя.

Навчання нейронної мережі полягає у виборі «найкращих» можливих ваг набору нейронів, що складають мережу (наприклад, зокрема ваг нейронів w_1, w_2 та w_3, показаних на рис. 1.1).



Рис. 1.2 Приклади зображень, що використовуються для навчання.

Таким чином, необхідно вибрати значення цих ваг, щоб найкраще вирішити задачу, яка вивчається за набором навчальних даних. Для розпізнавання об'єктів на зображеннях це проблема вирішується наступним чином: ми маємо як зображення x, так і пов'язані з ними y (ймовірність присутності на зображенні кота та/або собаки). На рис. 1.2 наведено кілька прикладів зображень, що використовуються для навчання мережі, для яких ми знаємо, що вони містять (клас котів та клас собак). Тому перед початком етапу навчання необхідно, щоб люди виконували довгу і нудну роботу, позначаючи тисячі чи навіть мільйони зображень.

Таким чином, процес тренування полягає у модифікації ваг w таким чином, що для кожного x мережа f_w передбачає якомога точніше пов'язаний y, тобто в кінці тренування, що $y \approx f_w(x)$. Простий вибір - мінімізувати суму E_w квадратів помилок, яку ми математично записуємо як:

$$\min_w E(w)=\Sigma_{(x,y)}(f_w(x)-y)^2 \qquad (1.2)$$

Це відповідає задачі оптимізації, оскільки необхідно знайти набір параметрів, який оптимізує певну кількість інтересів. Це складна проблема, оскільки параметрів дуже багато, і ці параметри, особливо параметри прихованих шарів, мають складний вплив на результат. Але, існують ефективні математичні та алгоритмічні методи для ефективного вирішення цього типу задач оптимізації. Вони ще не повністю досліджені на теоретичному рівні, і це дуже активна область досліджень [4]. Ці методи оптимізації модифікують ваги мережі, щоб покращити її та зменшити похибку навчання E_w. Математичне правило для прийняття рішення щодо стратегії оновлення ваги називається зворотним розповсюдженням, це особливий випадок математичного та алгоритмічного методу, який називається зворотним автоматичним диференціюванням.

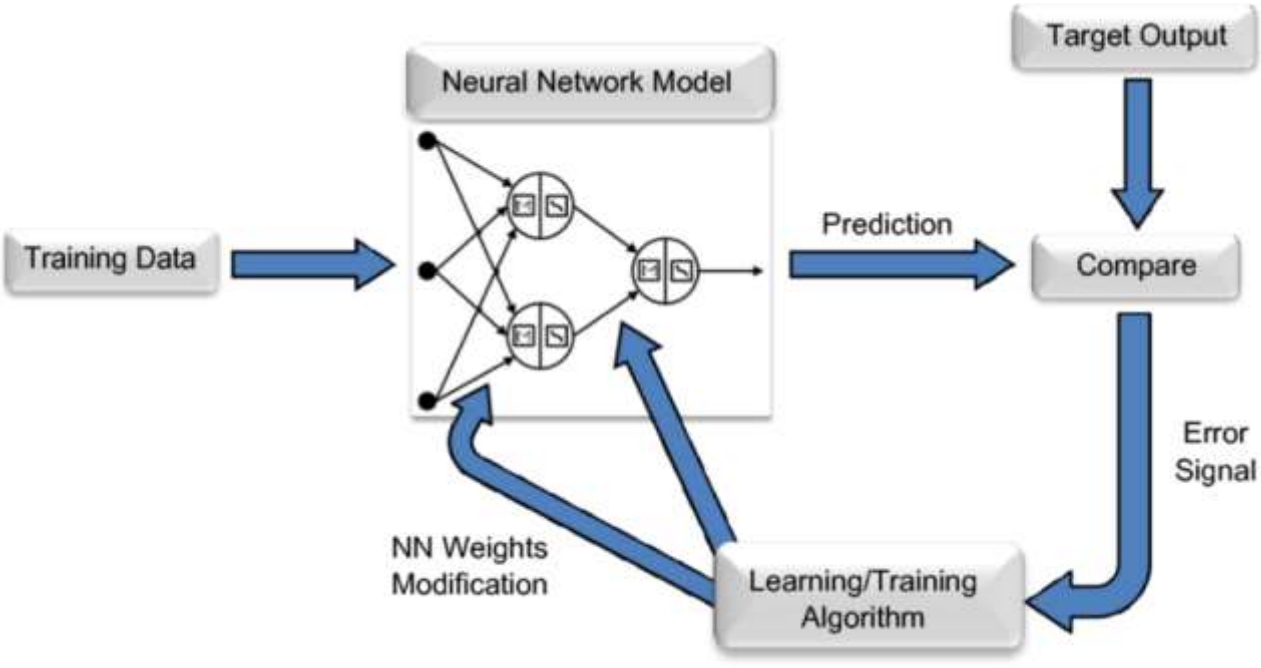


Рис. 1.3 Процес навчання нейромережі.

Ці методи навчання з учителем датуються в основному з 1980-х роками. Але лише в 2012 році в роботі Крижевського, Суцкевера та Хінтона було зроблено прорив, показавши, що глибокі мережі можуть вирішувати складні завдання розпізнавання зображень. Ця революція стала можливою завдяки поєднанню трьох інгредієнтів: нових баз даних, значно більших, ніж раніше;

високої обчислювальної потужність завдяки графічним процесорам («графічні процесори», які раніше обмежувались відеоіграми); впровадження декількох методів оптимізації, що покращують та стабілізують навчання.

Навчання без учителя є набагато більш правдоподібною моделлю навчання з точки зору біологічних коренів штучних нейронних мереж. Розвинена Кохоненом і багатьма іншими, вона не потребує цільового вектору для виходів і, отже, не вимагає порівняння з визначеними ідеальними відповідями.

Навчальна множина складається лише з вхідних векторів. Навчальний алгоритм підлаштовує ваги мережі так, щоб виходили узгоджені вихідні вектори, тобто щоб пред'явлення досить близьких вхідних векторів давало однакові виходи [5]. Процес навчання, виділяє статистичні властивості навчальної множини і групує подібні вектори в класи.

Класифікація нейронних мереж по типу вхідної інформації ділить їх на:

- аналогові – вхідна інформація представлена в формі дійсних чисел;
- виконавчі – вся вхідна інформація в таких мережах представляється у вигляді нулів і одиниць.

1.3. Архітектури нейронних мереж

У повнозв'язних нейронних мережах кожен нейрон передає свій вихідний сигнал іншим нейронам, в тому числі і самому собі. Всі вхідні сигнали подаються всім нейронам. Вихідними сигналами мережі можуть бути всі або деякі вихідні сигнали нейронів після кількох тактів функціонування мережі.

У багатошарових нейронних мережах нейрони об'єднуються в шари. Шар містить сукупність нейронів з єдиними вхідними сигналами.

Число нейронів в шарі може бути будь-яким і не залежить від кількості нейронів в інших шарах. У загальному випадку мережа складається з шарів, пронумерованих зліва направо. Зовнішні вхідні сигнали подаються на входи нейронів вхідного шару (його часто нумерують як нульовий), а виходами мережі є вихідні сигнали останнього шару. Крім вхідного і вихідного шарів в

багатошаровій нейронній мережі є один або кілька прихованих шарів. Зв'язки від виходів нейронів деякого шару q до входів нейронів наступного шару (q + 1) називаються послідовними.

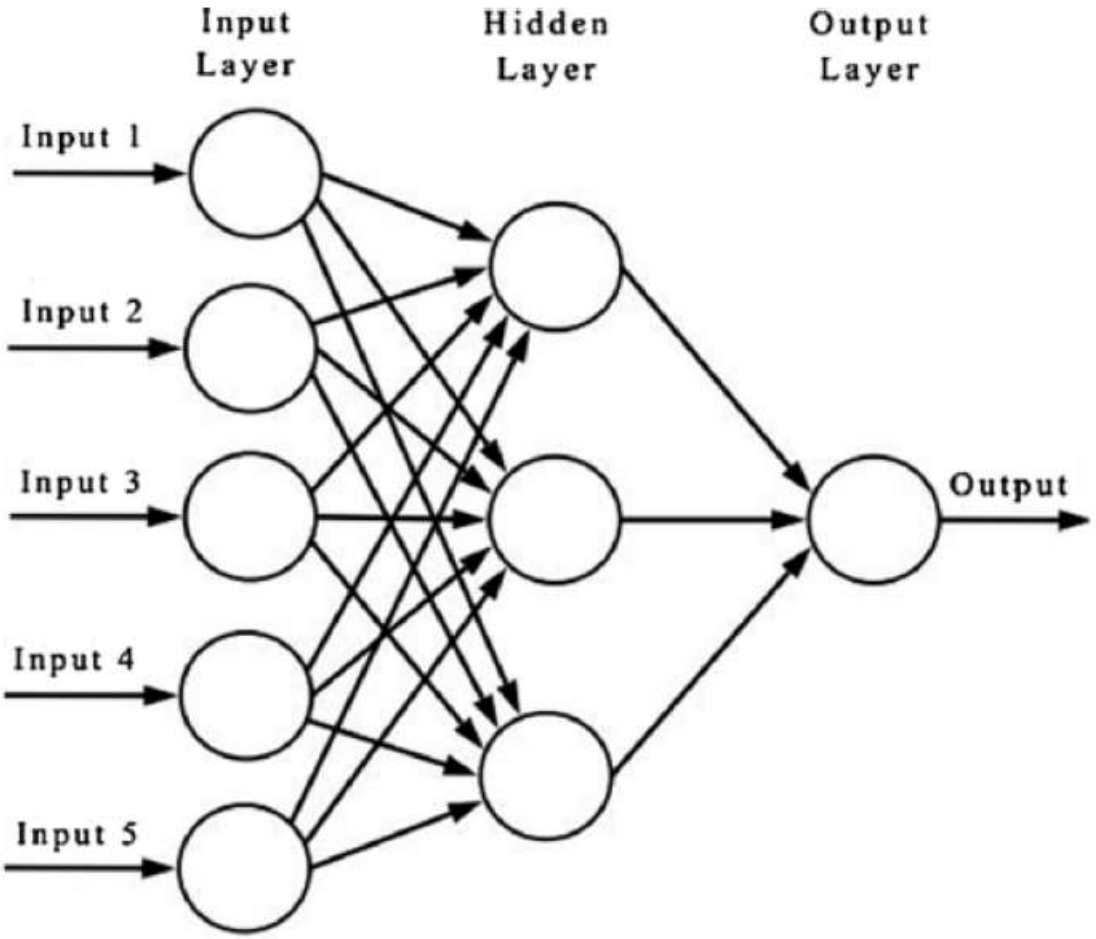


Рис. 1.4 Структура багатошарової нейронної мережі.

Спочатку розглянемо архітектуру згорткової нейронної мережі. Існує кілька варіацій цієї архітектури; вибір, який ми робимо, досить довільний. Однак алгоритми будуть дуже схожі для всіх варіацій, і їх виведення будуть виглядати дуже схожими.

Наша згорткова нейронна мережа складається з натсупних типів шарів:

- Згортковий: згорткові шари складаються з прямокутної сітки нейронів. Це вимагає, щоб попередній шар був також прямокутною сіткою нейронів. Кожен нейрон бере вхідні дані із прямокутного перерізу попереднього шару; ваги цього прямокутного перерізу однакові для кожного нейрона у згортковому шарі. Таким чином, згортковий шар - це

лише згортка зображення попереднього шару, де ваги визначають фільтр згортки. Крім того, у кожному згортковому шарі може бути кілька сіток; кожна сітка бере вхідні дані з усіх сіток попереднього шару, використовуючи потенційно різні фільтри.

- Max-Pooling: Після кожного згорткового шару може бути шар об'єднання. Об'єднуючий шар бере невеликі прямокутні блоки із згорткового шару і подає вибірки для отримання єдиного виводу з цього блоку. Існує кілька способів зробити це об’єднання, наприклад, взяти середнє або максимальне значення, або вивчену лінійну комбінацію нейронів у блоці. Наші шари об’єднання завжди будуть шарами максимального об’єднання; тобто вони приймають максимум блоку, який вони об’єднують.
- Fully-Connected: нарешті, після декількох згорткових та Max-Pooling шарів, обчислення високого рівня в нейронній мережі здійснюються через повністю з'єднані шари. Повністю зв’язаний шар приймає всі нейрони попереднього шару (будь-то повністю зв’язаний, об’єднаний або згортковий) і з’єднує його з кожним окремим нейроном, який він має. Повністю з'єднані шари більше просторово не розташовані (їх можна уявити як одновимірні), тому згорткових шарів після повністю пов'язаного шару бути не може.

Ми не обмежені двовимірними згортковими нейронними мережами. Ми можемо точно так само будувати одно- або тривимірні згорткові нейронні мережі; наші фільтри просто отримають відповідні розміри, а наші шари об'єднання також змінять розмір. Наприклад, ми можемо використовувати одновимірні згорткові мережі на аудіо- або тривимірні мережі на даних МРТ.

Тепер, коли ми описали структуру нашої нейронної мережі, попрацюємо над прямим і зворотним розповсюдженням, щоб зробити прогнозування та градієнтні обчислення в цих нейронних мережах.

1.4. Розрахунки для шару нейронної мережі

Тепер давайте трохи зменшимо масштаб і розглянемо, як виконуються розрахунки для цілого шару нейронної мережі. Ми використаємо наші знання про те, що відбувається всередині одного нейрону, і векторизуємо по всьому шару, щоб об'єднати ці обчислення у матричні рівняння. Для уніфікації позначень рівняння будуть записані для вибраного шару $[l]$. Індекс i позначає індекс нейрона в цьому шарі.

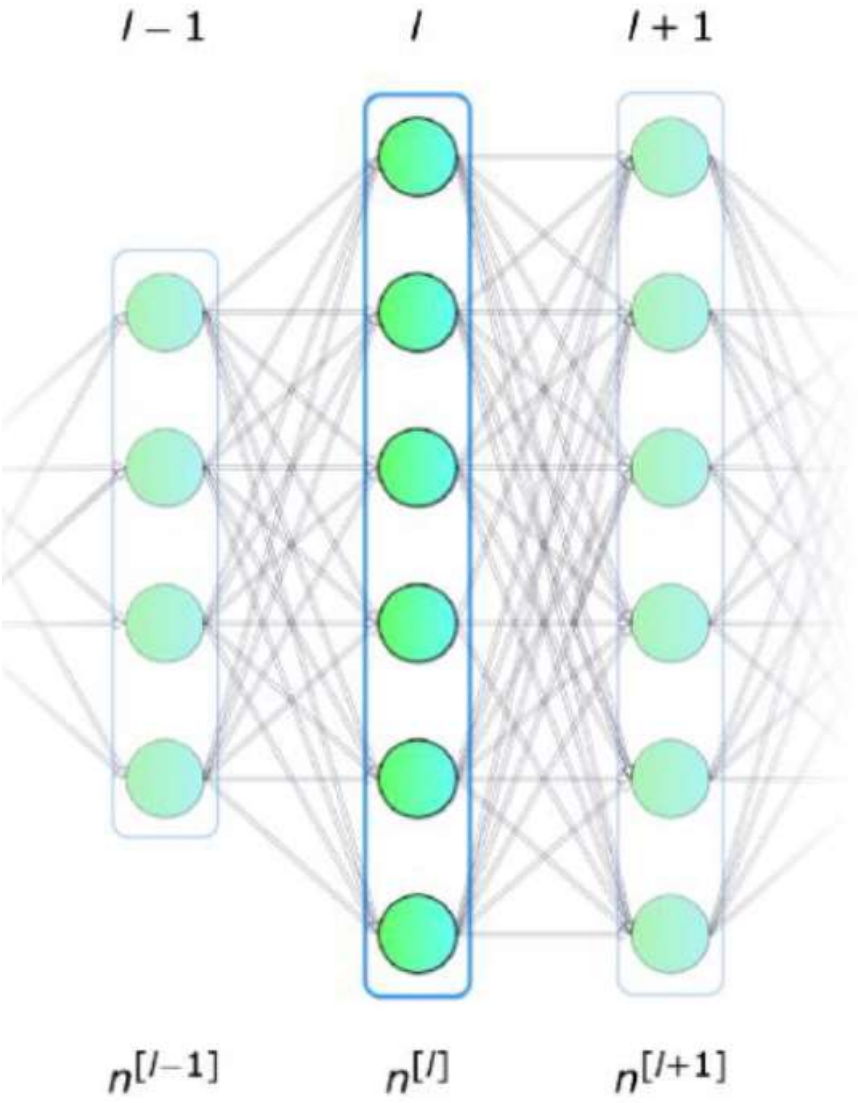


Рис. 1.5 Одношаровий перцептрон.

Коли ми писали рівняння для однієї одиниці, ми використовували x та y, які були відповідно вектором стовпця ознак та передбачуваним значенням. При

переході до загальних позначень шару ми використовуємо вектор a - що означає активацію відповідного шару. Отже, вектор x є активацією для шару 0 - вхідного шару. Кожен нейрон у шарі виконує подібні обчислення згідно з наступними рівняннями:

$$z_i^{[l]} = w_i^T a^{[l-1]} + b_i \qquad (1.3)$$

$$a_i^{[l]} = g^{[l]}(z_i^{[l]}) \qquad (1.4)$$

Для прикладу запишемо рівняння, наприклад, для шару 2:

$$z_1^{[2]} = w_1^T a^{[1]} + b_1 \qquad a_1^{[2]} = g^{[2]}(z_1^{[2]}) \qquad (1.5)$$

$$z_2^{[2]} = w_2^T a^{[1]} + b_2 \qquad a_2^{[2]} = g^{[2]}(z_2^{[2]}) \qquad (1.6)$$

$$z_3^{[2]} = w_3^T a^{[1]} + b_3 \qquad a_3^{[2]} = g^{[2]}(z_3^{[2]}) \qquad (1.7)$$

$$z_4^{[2]} = w_4^T a^{[1]} + b_4 \qquad a_4^{[2]} = g^{[2]}(z_4^{[2]}) \qquad (1.8)$$

$$z_5^{[2]} = w_5^T a^{[1]} + b_5 \qquad a_5^{[2]} = g^{[2]}(z_5^{[2]}) \qquad (1.9)$$

$$z_6^{[2]} = w_6^T a^{[1]} + b_6 \qquad a_6^{[2]} = g^{[2]}(z_6^{[2]}) \qquad (1.10)$$

Як бачимо, для кожного з шарів ми повинні виконати ряд дуже схожих операцій. Використання for-loop для цієї мети є не дуже ефективним, тому для прискорення обчислення ми будемо використовувати векторизацію [6]. Перш за все, склавши горизонтальні вектори ваг w (транспонувавши), ми побудуємо матрицю W. Аналогічним чином, ми складемо між собою зміщення кожного нейрона в шарі, створюючи вертикальний вектор b. Зараз ніщо не заважає нам будувати єдині матричні рівняння, які дозволяють нам проводити обчислення для всіх нейронів шару одночасно. Також запишемо розміри використаних нами матриць та векторів.

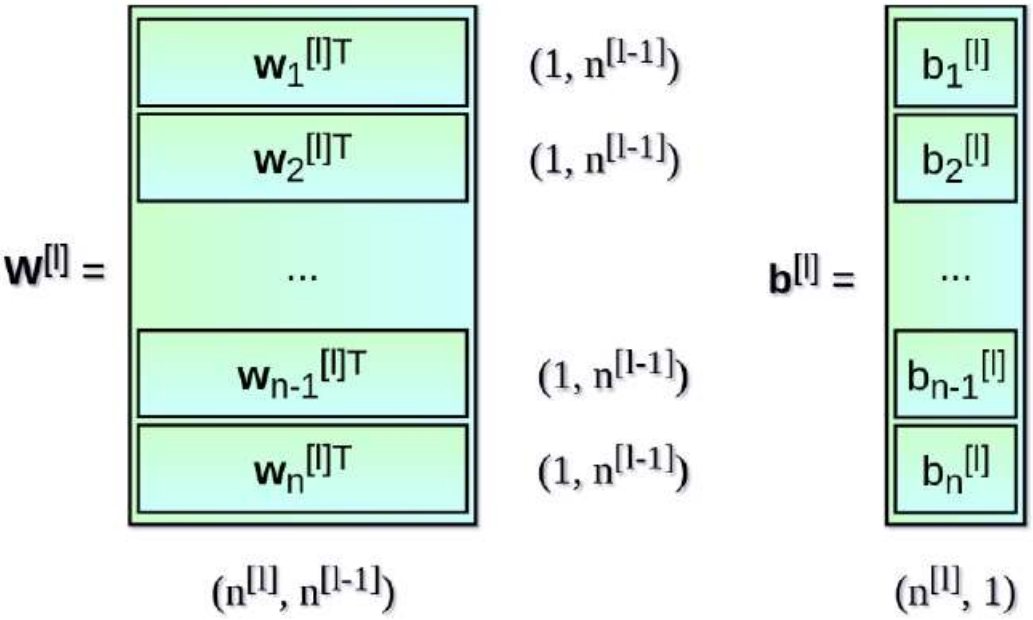


Рис. 1.6 Формальне представлення рівняння обчислення для нейронів шару.

$$z^{[l]} = w^{[l]} + a^{[l-1]} + b^{[l]} \quad (1.11)$$

$$a^{[l]} = g^{[l]}(z^{[l]}) \quad (1.12)$$

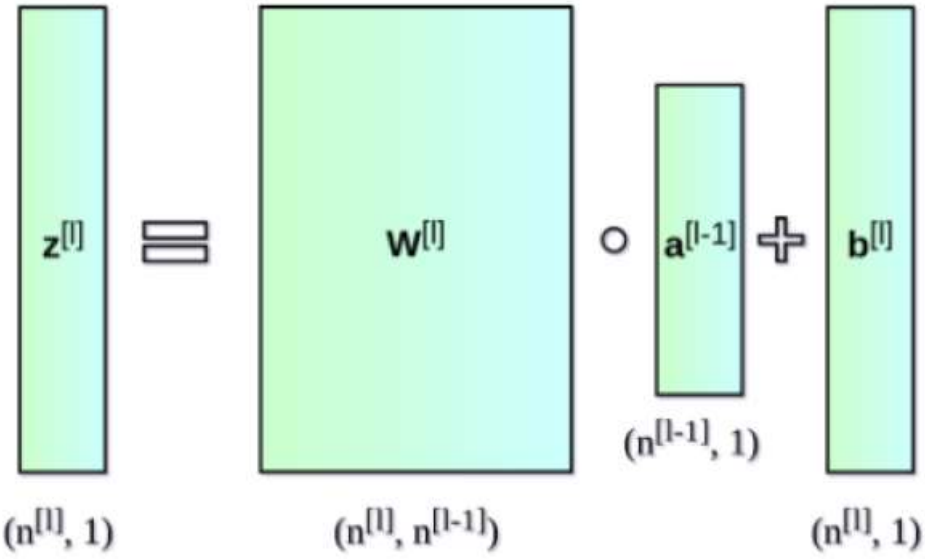


Рис. 1.7 Формальне представлення рівняння обчислення для нейронів шару.

Рівняння, яке ми склали до цього часу, включає лише один приклад. Під час процесу навчання нейронної мережі ми працюємо з великими наборами даних, до мільйонів записів. Отже, наступним кроком буде векторизація на

кількох прикладах. Припустимо, що у нашому наборі даних є m записів із кожною функцією nx. Перш за все, ми складемо вертикальні вектори x, a та z кожного шару, створюючи матриці X, A та Z відповідно. Потім ми переписуємо викладене раніше рівняння з урахуванням новостворених матриць.

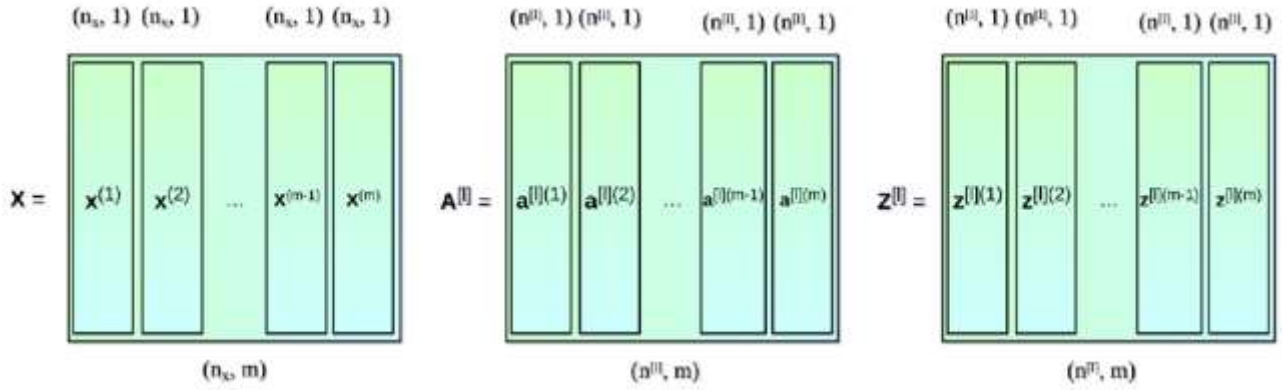


Рис. 1.8 Процес векторизації на кількох прикладах.

$$Z^{[l]} = W^{[l]} + A^{[l-1]} + b^{[l]} \quad (1.13)$$

$$A^{[l]} = g^{[l]}(Z^{[l]}) \quad (1.14)$$

1.5. Функція активації

Функції активації є одним з ключових елементів нейронної мережі. Без них наша нейромережа стала б комбінацією лінійних функцій, тому це була б просто сама лінійна функція. Наша модель мала б обмежену експансивність, не більшу за логістичну регресію. Елемент нелінійності забезпечує більшу гнучкість і створення складних функцій під час навчального процесу. Функція активації також суттєво впливає на швидкість навчання, що є одним з головних критеріїв їх вибору [7]. На Рис. 1.9 показані деякі загальновживані функції активації. В даний час найпопулярнішим для прихованих шарів є, ReLU, яка і була використана.

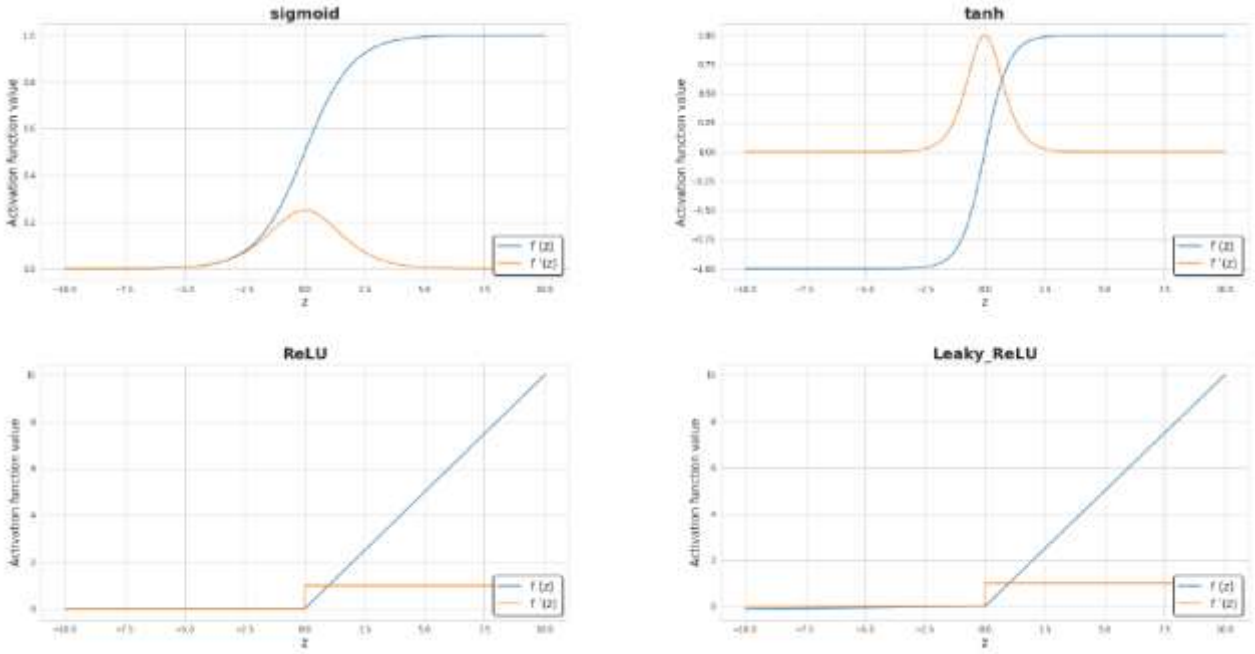


Рис. 1.9 Діаграми найпопулярніших функцій активації разом з їх похідними.

ReLU наполовину випрямлені (знизу). f (z) дорівнює нулю, коли z менше нуля, і f (z) дорівнює z, коли z більше або дорівнює нулю. Діапазон: [від 0 до нескінченності)

Також була використана функція Softmax.

Функція активації Softmax, також відома як SoftArgMax або Нормалізована експоненціальна функція, являє собою функцію активації, яка приймає вектори дійсних чисел в якості вхідних даних і нормалізує їх в розподіл ймовірностей, пропорційних експонентам вхідних чисел. Перед застосуванням деякі вхідні дані можуть бути негативними або більше 1. Крім того, вони можуть не складати 1. Після застосування Softmax кожен елемент буде в діапазоні від 0 до 1, а сума елементів буде дорівнює 1. Це речі, їх можна інтерпретувати як розподіл ймовірностей . Для більшої ясності: чим більше вхідний число, тим більше буде вірогідність.

1.6. Функція втрат

Основним джерелом інформації про хід навчального процесу є значення функції втрат. Взагалі кажучи, функція втрат повинна показати, наскільки далеко ми знаходимось від «ідеального» рішення. У нашому випадку ми використовували двійкову кроссентропію, але в залежності від проблеми ми маємо справу з різними функціями. Функція, яку ми використовуємо, описується наступною формулою. З кожною ітерацією значення функції втрат зменшується, а точність зростає.

$$\mathrm{J}(\mathrm{W},\mathrm{b}) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{L}(\hat{\mathrm{y}}^{(\mathrm{i})}, \mathrm{y}^{(\mathrm{i})}) \qquad (1.15)$$

$$L(\hat{y},y) = -(y \log \hat{y} + (1-y) \log (1-\hat{y})) \qquad (1.16)$$

Процес навчання полягає у зміні значень параметрів W та b, щоб функція втрат була мінімізована. Для досягнення цієї мети ми звернемось за допомогою до обчислення та використаємо метод градієнтного спуску, щоб знайти мінімум функції. У кожній ітерації ми будемо обчислювати значення часткових похідних функції втрат відносно кожного з параметрів нашої нейронної мережі. Похіднамає здатність описувати нахил функції. Завдяки цьому ми знаємо, як маніпулювати змінними, щоб рухатись вниз по графіку. З кожною наступною епохою ми рухаємось до мінімуму. У нашій нейронній мережі це працює однаково - градієнт, розрахований на кожній ітерації, показує нам напрямок, в якому ми повинні рухатися. Головна відмінність полягає в тому, що у нашій зразковій нейронній мережі ми маємо набагато більше параметрів, якими маніпулюємо.

1.7. Зворотне поширення помилки

Зворотне поширення це алгоритм, який дозволяє обчислити дуже складний градієнт. Параметри нейронної мережі коригуються згідно з наступними формулами.

$$W^{[l]} = W^{[l]} - \alpha dW^{[l]} \qquad (1.17)$$

$$b^{[l]} = b^{[l]} - \alpha db^{[l]} \qquad (1.18)$$

У наведених вище рівняннях α представляє швидкість навчання - гіперпараметр, який дозволяє контролювати значення виконаного коригування. Вибір швидкості навчання є вирішальним - ми встановлюємо її занадто низькою, наша нейронна мережа буде вчитися дуже повільно, ми встановлюємо її занадто високою і ми не зможемо досягти мінімуму. dW і db обчислюються за допомогою ланцюгового правила, часткових похідних функції втрат відносно W і b. Розмір dW і db такі ж, як розміри W і b відповідно. На рис. 1.10. показана послідовність операцій у нейронній мережі. Ми чітко бачимо, як пряме та зворотне поширення працюють разом, щоб оптимізувати функцію втрат.

$$\mathrm{dW}^{[\mathrm{l}]} = \frac{\partial \mathrm{L}}{\partial \mathrm{W}^{[\mathrm{l}]}} = \frac{1}{m}\mathrm{dZ}^{[\mathrm{l}]}\mathrm{A}^{[\mathrm{l}-1]\mathrm{T}} \qquad (1.19)$$

$$\mathrm{db}^{[\mathrm{l}]} = \frac{\partial \mathrm{L}}{\partial \mathrm{b}^{[\mathrm{l}]}} = \frac{1}{m}\sum_{i=1}^{m}\mathrm{dZ}^{[\mathrm{l}](\mathrm{i})} \qquad (1.20)$$

$$\mathrm{dA}^{[\mathrm{l}-1]} = \frac{\partial \mathrm{L}}{\partial \mathrm{A}^{[\mathrm{i}-\mathrm{l}]}} = \mathrm{W}^{[\mathrm{l}]\mathrm{T}}\mathrm{dZ}^{[\mathrm{l}]} \qquad (1.21)$$

$$dZ^{[l]} = dA^{[l]} * g'(Z^{[l]}) \qquad (1.22)$$

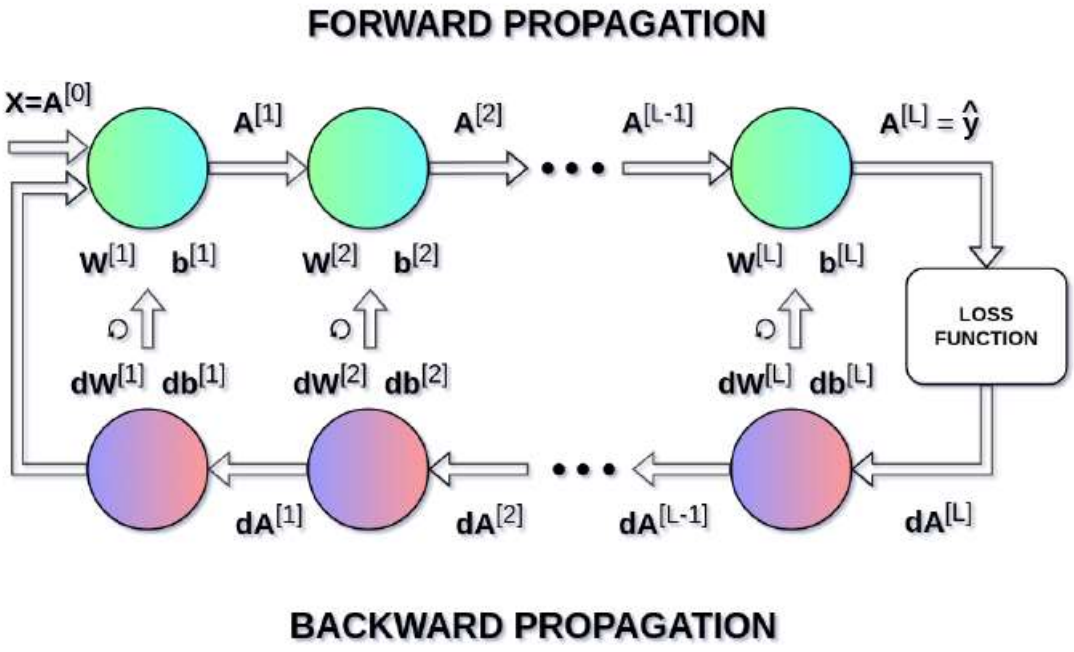


Рис. 1.10 Пряме та зворотне поширення помилки.

Вище були розглянуті повнозв'язні нейронні мережі. Це мережі, нейрони яких розділені на групи, що утворюють послідовні шари. Кожна така одиниця

пов'язана з кожним нейроном із сусідніх шарів. Приклад такої архітектури наведено на рис. 1.11.

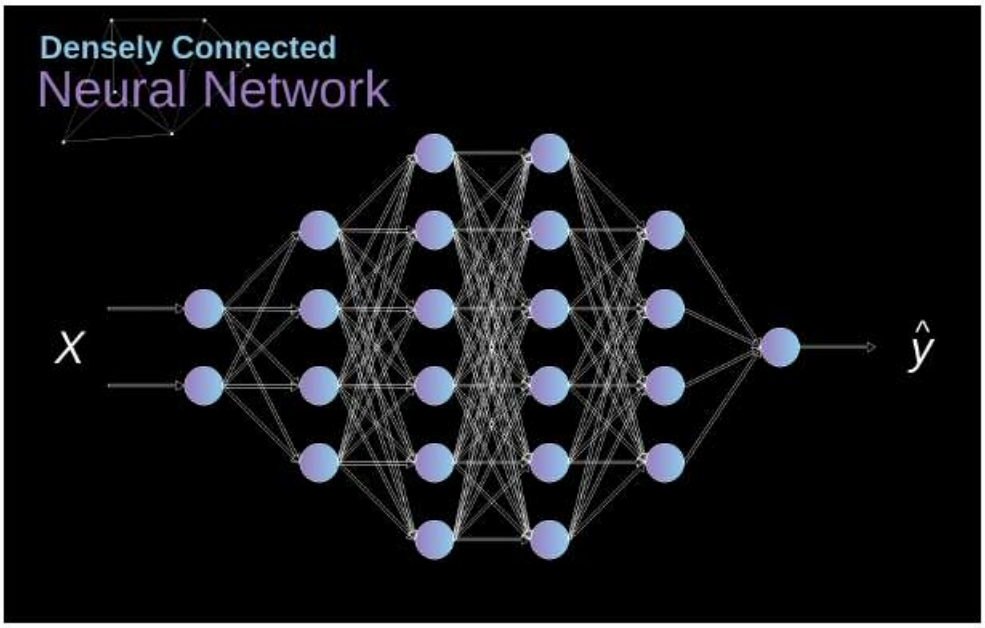


Рис. 1.11 Щільно пов'язана архітектура нейронної мережі.

Цей підхід добре працює, коли ми вирішуємо проблему класифікації на основі обмеженого набору визначених ознак - наприклад, ми прогнозуємо позицію футболіста на основі статистики, яку він реєструє під час ігор. Однак ситуація ускладнюється при роботі з зображеннями. Звичайно, ми могли б розглядати яскравість кожного пікселя як окрему особливість і передавати його як вхід для нашої повнозв'язної мережі. На жаль, для того, щоб це працювало для типової фотографії смартфона, наша мережа мала б містити десятки, а то й сотні мільйонів нейронів. З іншого боку, ми могли б зменшити масштаб нашого зображення, але при цьому ми втратили б цінну інформацію. Відразу ми бачимо, що традиційна стратегія для нас нічого не робить - нам потрібен новий розумний спосіб використовувати якомога більше даних, але в той же час зменшити кількість необхідних розрахунків і параметрів. Саме тоді в дію вступає CNN.

1.8. Структура цифрових зображень

Зображення це величезні матриці чисел. Кожне таке число відповідає яскравості окремого пікселя. У моделі RGB кольорове зображення насправді складається з трьох таких матриць, що відповідають трьом кольоровим каналам - червоному, зеленому та синьому. На чорно-білих зображеннях нам потрібна лише одна матриця. Кожна з цих матриць зберігає значення від 0 до 255 Рис. 1.12. Цей діапазон є компромісом між ефективністю зберігання інформації про зображення (256 значень ідеально вкладаються в 1 байт) та чутливістю людського ока (ми виділяємо обмежену кількість відтінків одного кольору)[8].

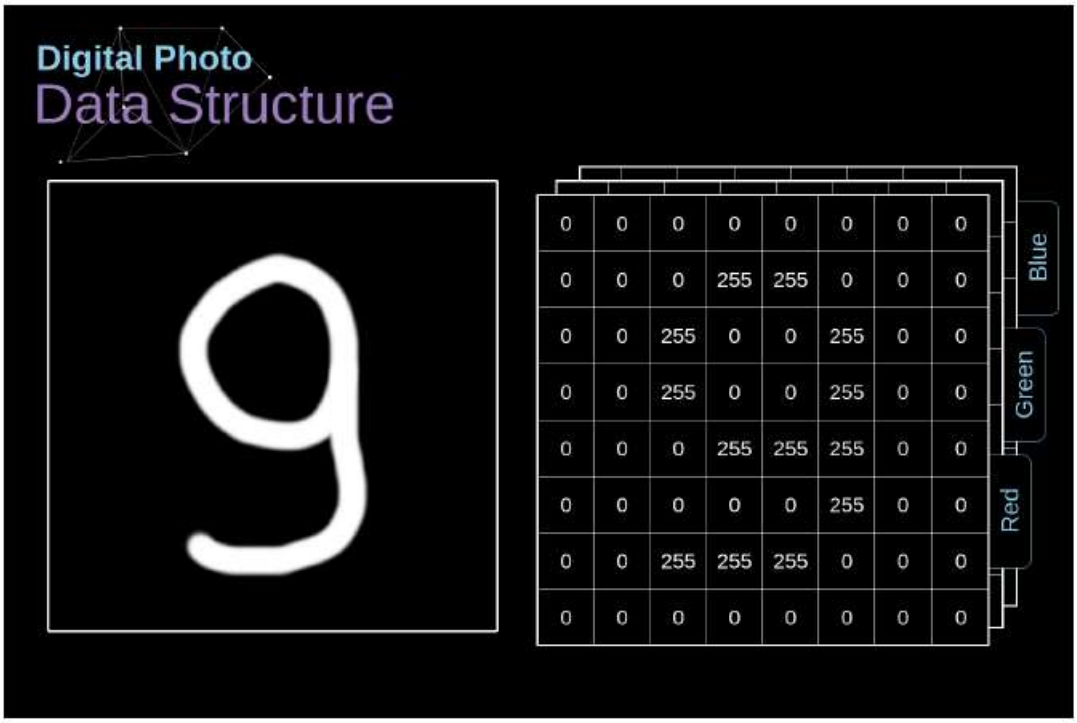


Рис. 1.12 Структура даних за цифровими зображеннями.

1.9. Ядро згортки

Ядра згортки використовується не тільки в CNN, але також є ключовим елементом багатьох інших алгоритмів Computer Vision. Це процес, коли ми беремо малу матрицю чисел (що називається ядром або фільтром), ми передаємо її своєму зображенню та перетворюємо на основі значень із фільтра. Подальші значення карти об’єктів обчислюються за наступною формулою, де

вхідне зображення позначається f, а наше ядро h. Індекси рядків і стовпців матриці результатів позначені m і n відповідно.

$$G[m,n]=(f*h)[m,n]=\Sigma_j\Sigma_k\, h[j,k]\; f[m-j,\, n-k] \qquad (1.23)$$

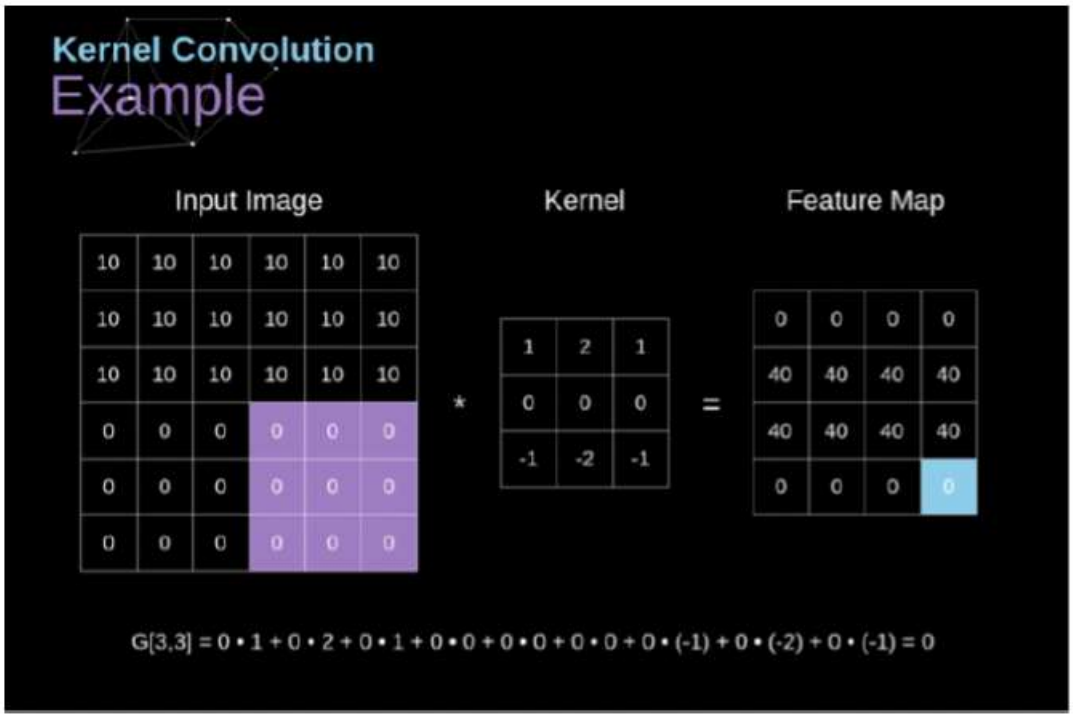


Рис. 1.13 Приклад ядра згортки.

Розмістивши наш фільтр над вибраним пікселем, ми беремо кожне значення з ядра і множимо їх парами з відповідними значеннями із зображення. Нарешті ми підсумовуємо все і розміщуємо результат у потрібному місці на карті вихідних характеристик. Вище ми бачимо, як виглядає така операція в мікро-масштабі.

Коли ми виконуємо згортку зображення 6x6 із ядром 3x3, ми отримуємо карту функцій 4x4. Це тому, що є лише 16 унікальних позицій, де ми можемо розмістити наш фільтр всередині цієї картинки. Оскільки наше зображення зменшується щоразу, коли ми робимо згортку, ми можемо зробити це обмежену кількість разів, перш ніж наше зображення повністю зникне. Вплив пікселів, розташованих на околиці, значно менший, ніж у центрі зображення. Таким чином ми втрачаємо частину інформації, що міститься на зображенні.

Для вирішення обох цих проблем ми можемо заповнити своє зображення додатковою рамкою. Наприклад, якщо ми використовуємо відступ 1px, ми

збільшуємо розмір нашої фотографії до 8x8, так що результат згортки з фільтром 3x3 буде 6x6. На практиці ми заповнюємо додаткові значення нулями. *same* - ми використовуємо межу навколо оригінального зображення, щоб зображення на вході та виході мали однаковий розмір. У другому випадку ширина заповнення повинна відповідати наступному рівнянню, де p - заповнення, а f - розмір фільтра (зазвичай непарний).

$$\mathrm{p} = \frac{\mathrm{f}-1}{2} \qquad (1.24)$$

До цього ми завжди зміщували ядро на один піксель. Однак довжину кроку також можна розглядати як один із параметрів згорткового шару. Розробляючи нашу архітектуру CNN, ми можемо вирішити збільшити крок, якщо хочемо, щоб сприйнятливі поля менше перекривались або якщо ми хочемо менших просторових розмірів нашої карти функцій. Розміри вихідної матриці - з урахуванням заповнення та кроку - можуть бути розраховані за наступною формулою.

$$\mathrm{n}_{\mathrm{out}} = [\frac{\mathrm{n}_{in}+2\mathrm{p}-\mathrm{f}}{s} + 1] \qquad (1.25)$$

Зміна згортки за об'ємом - дуже важлива концепція, яка дозволить нам не тільки працювати з кольоровими зображеннями, але що важливіше застосовувати кілька фільтрів в межах одного шару. Перше важливе правило полягає в тому, що фільтр і зображення, до якого ми хочемо його застосувати, повинні мати однакову кількість каналів. В основному, ми продовжуємо подібно до прикладу з Рис. 1.14, проте цього разу ми множимо пари значень із тривимірного простору. Якщо ми хочемо використовувати кілька фільтрів на одному і тому ж зображенні, ми виконуємо згортку для кожного з них окремо, складаємо результати один на інший і об'єднуємо їх у єдине ціле[9]. Розміри отриманого тензору (як можна назвати нашу 3D-матрицю) відповідають такому рівнянню, в якому: n - розмір зображення, f - розмір фільтра, nc - кількість каналів на зображенні, p - використовуване відступ, s – використаний крок, nf - кількість фільтрів.

$$[\mathrm{n}, \mathrm{n}, \mathrm{nc}] * [\mathrm{f}, \mathrm{f}, \mathrm{nc}] = \left[\frac{\mathrm{n}\ +2\mathrm{p}-\mathrm{f}}{s} + 1\right], \left[\frac{\mathrm{n}\ +2\mathrm{p}-\mathrm{f}}{s} + 1\right], nf \qquad (1.26)$$

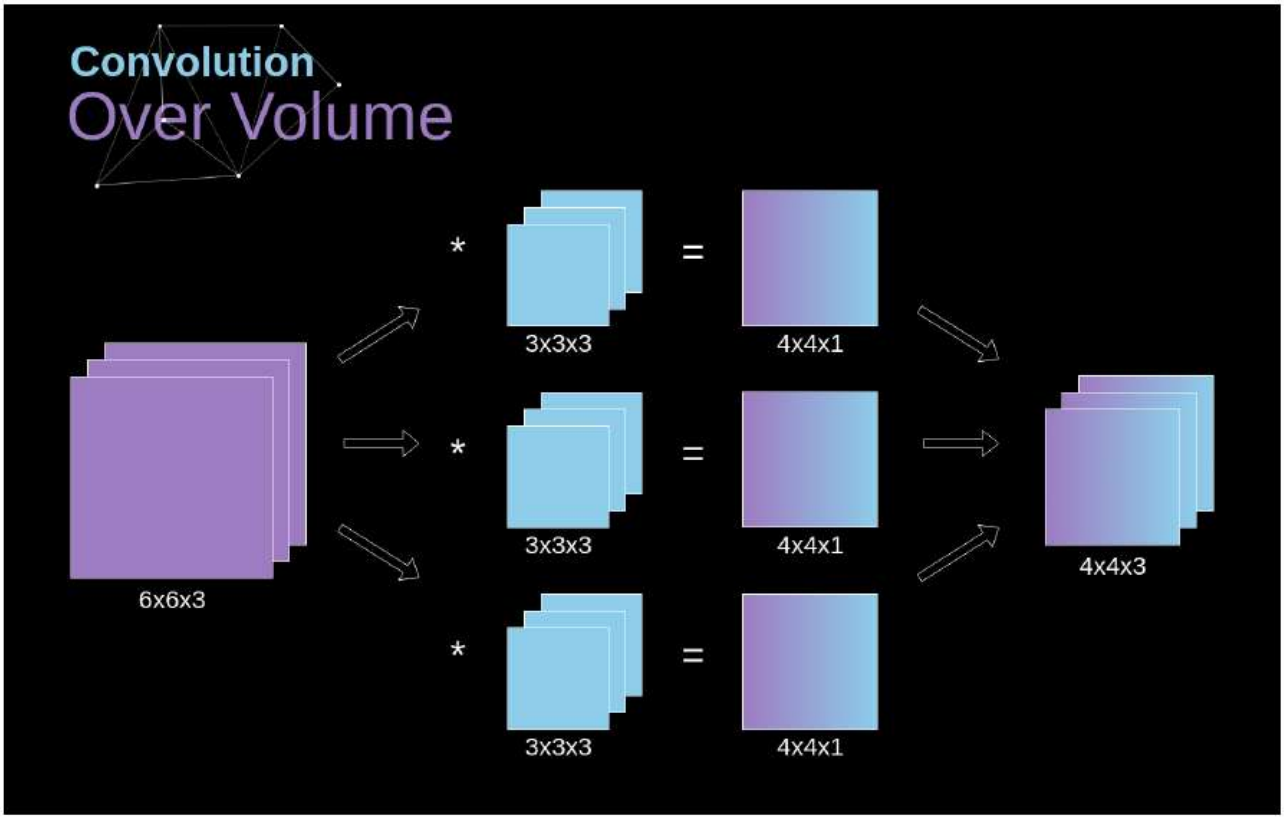


Рис. 1.14 Зміна згортки за розмірністю.

1.10. Шари згортки

Нарешті прийшов час побудувати загальний рівень нашої CNN. Методологія майже ідентична методиці, яку ми використовують для щільно з'єднаних нейронних мереж, єдина відмінність полягає в тому, що замість простого множення матриць цього разу ми будемо використовувати згортку. *Forward propagation* складається з двох етапів. Перший - це обчислення проміжного значення Z, яке отримується в результаті згортки вхідних даних з попереднього шару з тензором W (що містить фільтри), а потім додавання зміщення b. Другий - це застосування нелінійної функції активації до нашого проміжного значення (наша активація позначається g). На рис. 1.15 ми можемо побачити візуалізацію, що описує розміри тензорів, які використовуються в рівнянні.

$$Z^{[l]} = W^{[l]} * Z^{[l-1]} + b^{[l]} \qquad (1.27)$$

$$A^{[l]} = g^{[l]}(Z^{[l]}) \qquad (1.28)$$

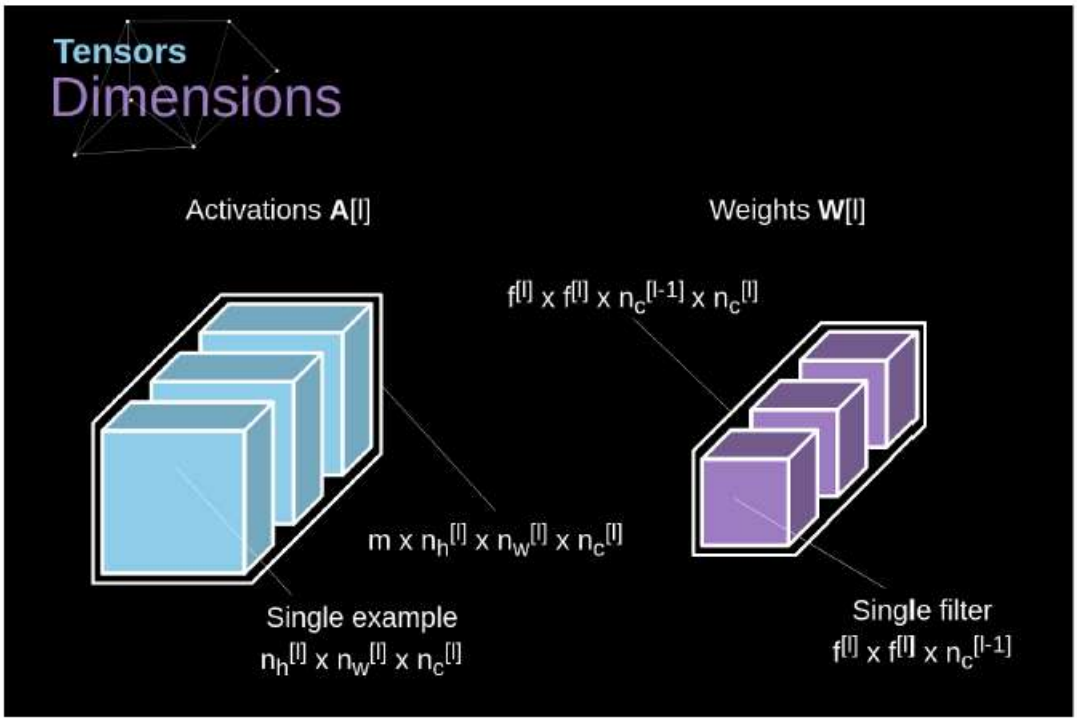


Рис. 1.15 Розміри тензорів.

Як пам’ятаємо, повно-зв’язні нейронні мережі погано працюють із зображеннями через величезну кількість параметрів, які потрібно було б вивчити. Розглянемо, як нам оптимізувати розрахунки. На рис. 1.16 2D-згортка, нейрони, позначені цифрами 1–9, утворюють вхідний шар, який отримує яскравість наступних пікселів, тоді як одиниці A-D позначають обчислені елементи карти об’єктів. І останнє, але не менш важливе, I-IV - це наступні значення з ядра - їх потрібно вивчити.

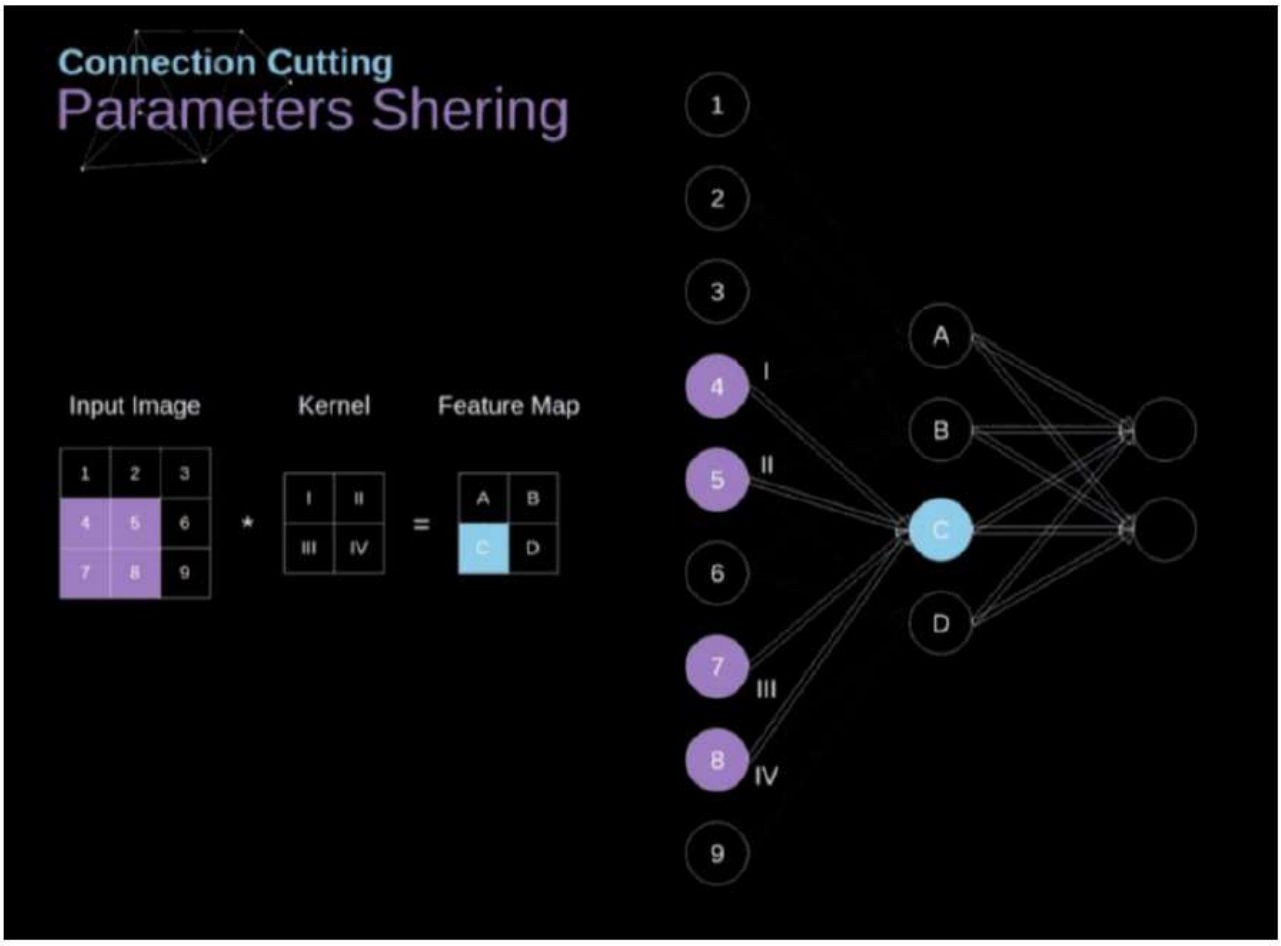


Рис. 1.16 Вилучення з’єднань та обмін параметрами.

Тепер зупинимось на двох дуже важливих атрибутах шарів згортки. Перш за все, ми можемо бачити, що не всі нейрони в двох послідовних шарах пов’язані між собою. Блок 3 впливає лише на значення C. По-друге, ми бачимо, що деякі нейрони мають однакові ваги. Обидві ці властивості означають, що у нас є набагато менше параметрів для вивчення. До речі, варто зазначити, що одне значення з фільтра впливає на кожен елемент карти об’єктів - це буде вирішальним у контексті *backpropagation*.

1.11. Backpropagation згорткового шару

Сьогодні нам не потрібно турбуватися про зворотне розповсюдження - за нас це роблять програми глибокого навчання. Подібно до повнозв’язаних нейронних мереж, наша мета - розрахувати похідні та використовувати їх для

оновлення значень наших параметрів у процесі, який називається градієнтним спуском.

У наших розрахунках ми будемо використовувати ланцюгове правило. Ми хочемо оцінити вплив зміни параметрів на результуючу карту особливостей, а згодом і на кінцевий результат.

$$\mathrm{dA}^{[\mathrm{l}]} = \frac{\partial \mathrm{L}}{\mathrm{dA}^{[\mathrm{l}]}} \qquad (1.28)$$

$$\mathrm{dZ}^{[\mathrm{l}]} = \frac{\partial \mathrm{L}}{\mathrm{dZ}^{[\mathrm{l}]}} \qquad (1.29)$$

$$\mathrm{dW}^{[\mathrm{l}]} = \frac{\partial \mathrm{L}}{\mathrm{dW}^{[\mathrm{l}]}} \qquad (1.30)$$

$$\mathrm{db}^{[\mathrm{l}]} = \frac{\partial \mathrm{L}}{\mathrm{db}^{[\mathrm{l}]}} \qquad (1.31)$$

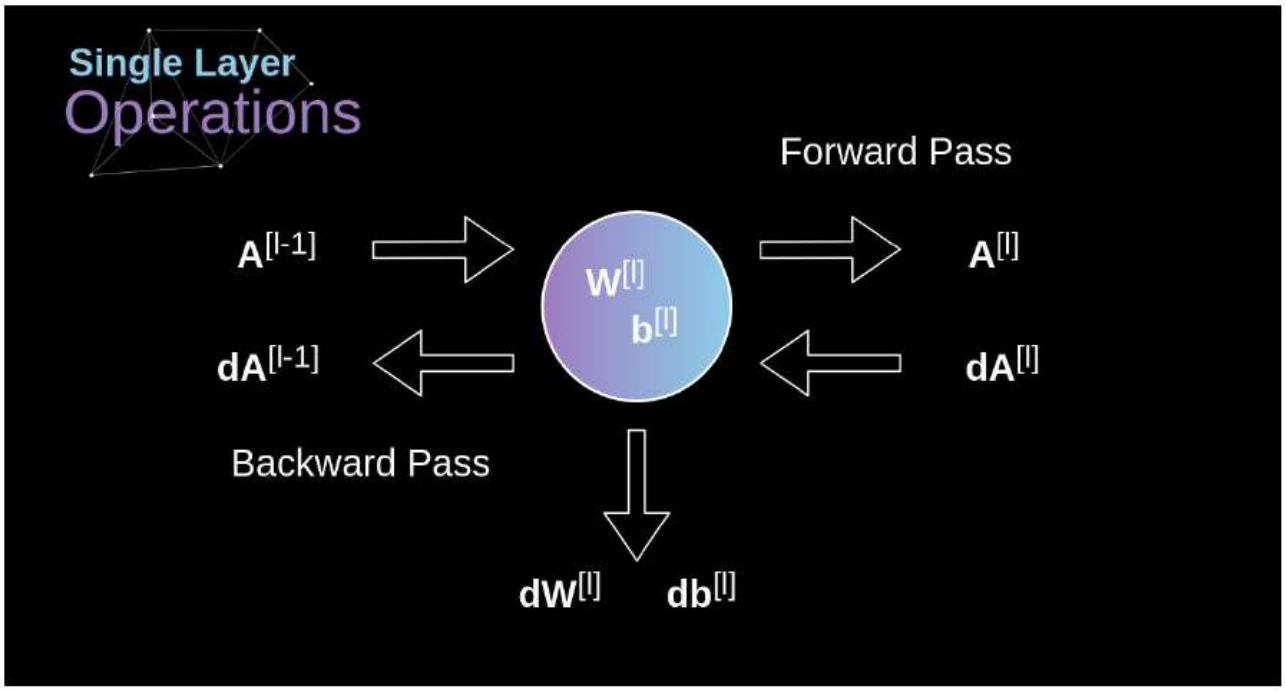


Рис. 1.17 Вхідні та вихідні дані для одного рівня згортки при прямому та зворотному розповсюдженні.

Наше завдання - розрахувати $dW^{[l]}$ і $db^{[l]}$ - які є похідними, пов'язаними з параметрами поточного шару, а також значення $dA^{[l-1]}$ - які будуть передані попередньому шару. Як показано на рисунку, ми отримуємо $dA^{[l]}$ - як вхідні дані. Звичайно, розміри тензорів dW і W, db і b, а також dA і A, однакові. Першим кроком є отримання проміжного значення $dZ^{[l]}$ - застосовуючи похідну нашої функції активації до нашого вхідного тензора. Відповідно до правила ланцюга, результат цієї операції буде використаний пізніше.

$$dZ^{[l]}=dA^{[l]} * g'(Z^{[l]}) \qquad (1.32)$$

Тепер нам потрібно мати справу із зворотним розповсюдженням самої згортки, і для досягнення цієї мети ми будемо використовувати матричну операцію, яка називається повною згорткою - яка наведена нижче. Під час цього процесу ми використовуємо ядро, яке ми попередньо повернули на 180 градусів. Цю операцію можна описати за такою формулою, де фільтр позначається W, а $dZ[m,n]$ - скаляр, який належить до часткової похідної, отриманої з попереднього шару.

$$\mathrm{dA}+= \sum_{m=0}^{n_h} \sum_{n=0}^{n_w} W * dZ[m,n] \qquad (1.33)$$

1.12. Pooling шари зі зворотним поширенням

В шарах цього типу ми не маємо жодних параметрів, які нам доводилося б оновлювати, наше завдання полягає лише в тому, щоб належним чином розподілити градієнти. При прямому розповсюдженні для максимального об’єднання ми вибираємо максимальне значення з кожної області і передаємо їх на наступний шар. Тому очевидно, що під час зворотного поширення градієнт не повинен впливати на елементи матриці, які не були включені в пряме розповсюдження. На практиці це досягається створенням маски, яка запам'ятовує положення значень, використаних на першій фазі, яку ми згодом можемо використовувати для передачі градієнтів.

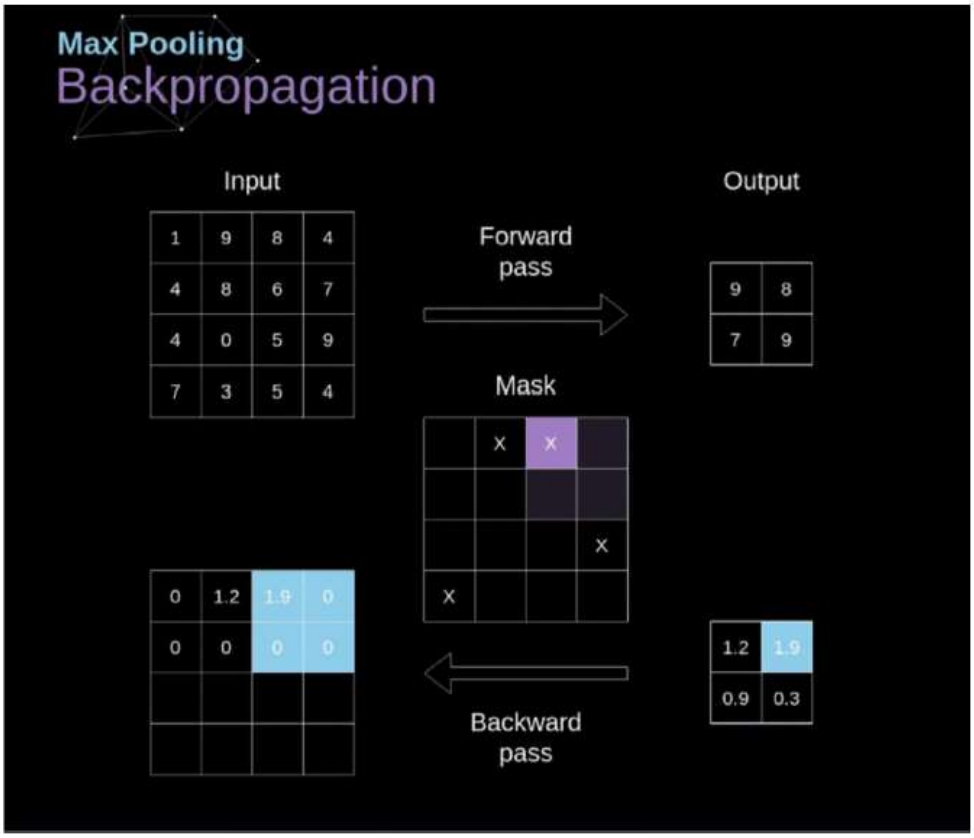


Рис. 1.18 Max pooling зворотний прохід.

1.13. Класифікація зображення з використанням згорткових нейронних мереж в Keras

На рис. 1.19 наведена схема типового CNN. Перша частина складається з шарів згортки та максимального пулу, які виступають в якості екстрактора ознак. Друга частина складається з повнозв'язного шару, який виконує нелінійні перетворення витягнутих ознак і діє як класифікатор.

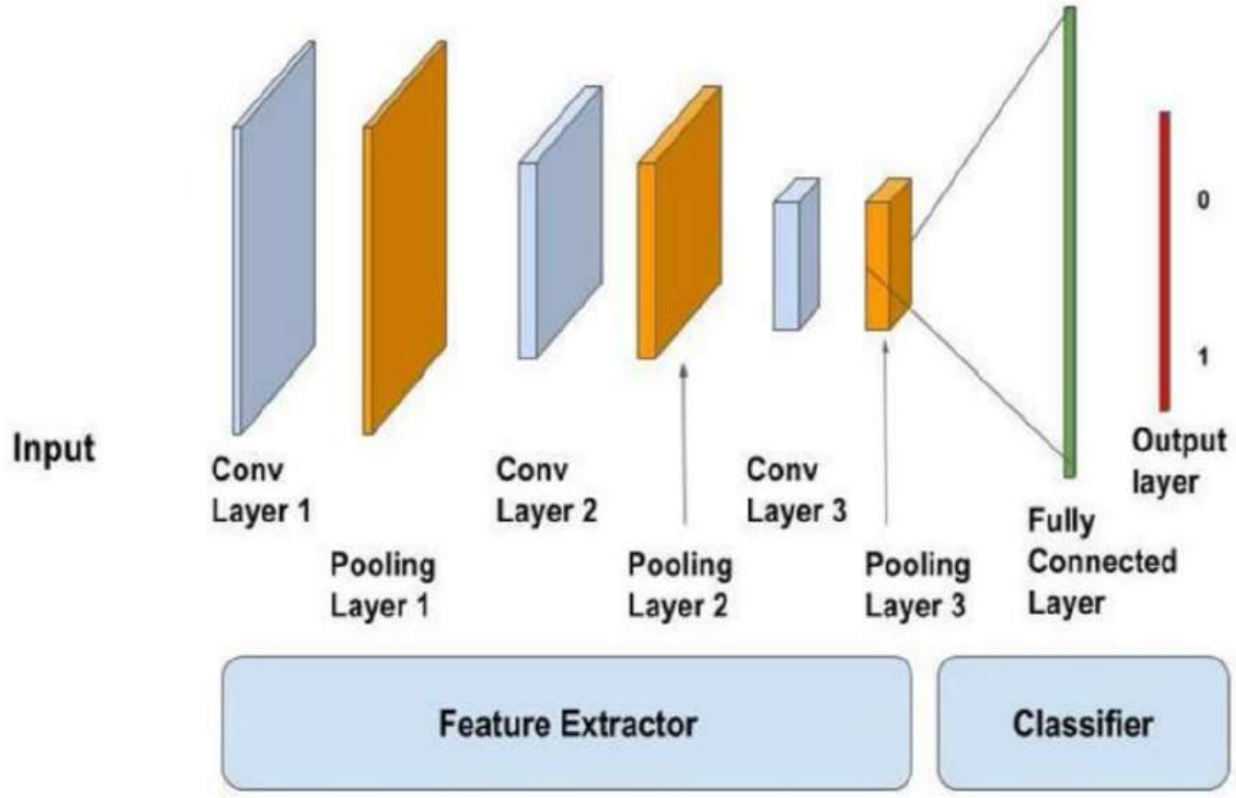


Рис. 1.19 Схема типового CNN.

На наведеній вище діаграмі вхід подається в мережу послідовних шарів Conv, Pool і Dense. Вихідний сигнал може бути шаром softmax, який вказує результат класифікації. Також, в якості вихідного може бути використаний сигмоідний шар, на виході якого буде результат класифікації у вигляді ймовірності.

Згортковий шар ще можна розглядати, як очі згорткової нейронної мережі. Нейрони в цьому шарі шукають певні особливості. Згортку можна розглядати як зважену суму між двома сигналами або функціями. Ядро згортки ковзає по всій матриці для отримання карти активації.

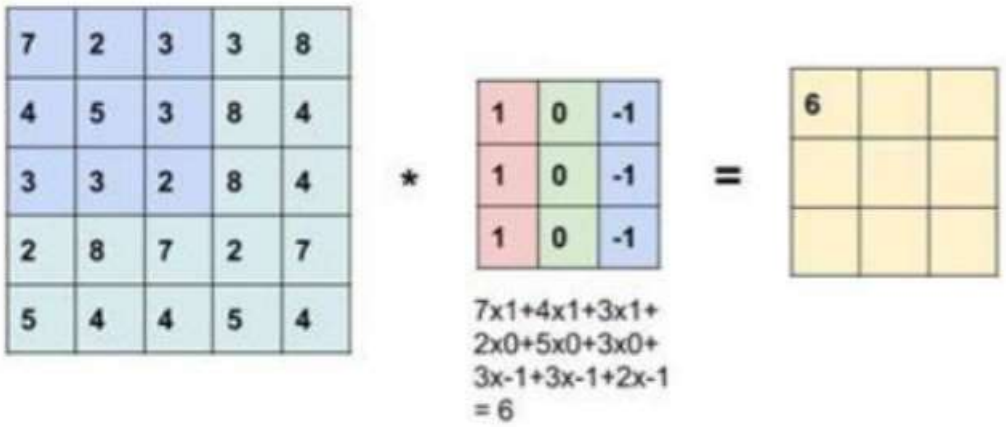


Рис. 1.20 Операція згортки на матриці розміром 5×5 з ядром розміром 3×3.

Припустимо, що вхідне зображення має розмір 32×32×3, тобто це тривимірний масив з глибиною 3. Будь-який фільтр згортки, який ми визначаємо на цьому шарі, повинен мати глибину, рівну глибині введення. Тому ми можемо вибрати фільтри згортки глибини 3 (наприклад, 3×3×3 або 5×5×3 або 7×7×3 і т. д.). Виберемо фільтр згортки розміром 3×3×3, тобто згорткове ядро буде кубом замість квадрата.

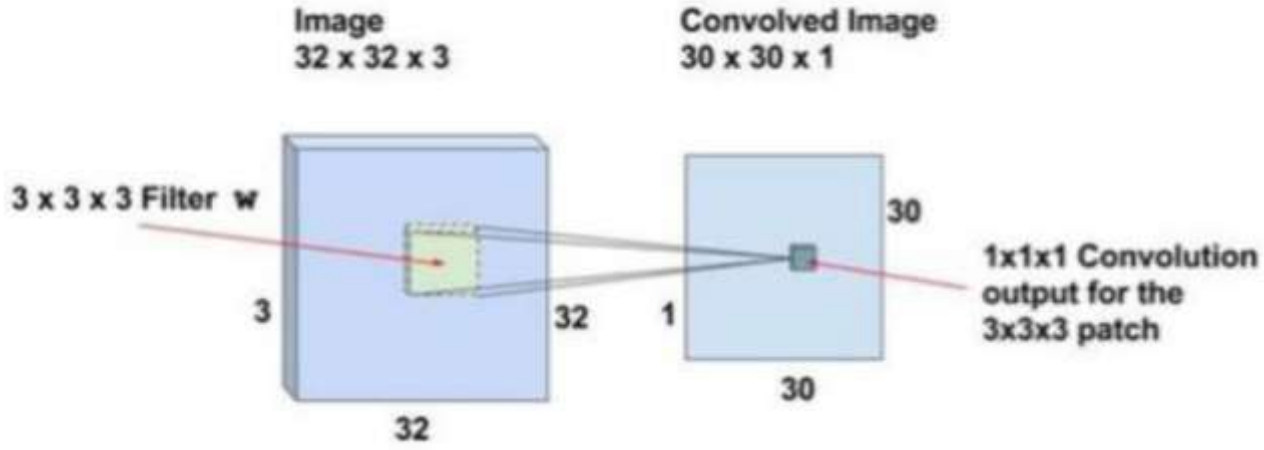


Рис. 1.21 Операція згортки для зображення розміром 32×32×3.

Якщо ми зможемо виконати операцію згортки, зсунувши фільтр 3×3×3 на всі зображення розміром 32×32×3, то ми отримаємо зображення з розмірністю 30×30×1. Це пов'язано з тим, що операція згортки неможлива для смуги шириною 2 пікселя навколо зображення. Фільтр завжди знаходиться всередині

зображення і тому 1 піксель віддаляється від лівої, правої, верхньої та нижньої частини зображення.

Для вхідного зображення 32×32×3 і розміру фільтра 3×3×3 у нас є 30×30×1 розміщення, і для кожного місця існує нейрон. Тоді виходи 30×30×1 або активації всіх нейронів називаються картами активації. Карта активації одного рівня виступає входом для наступного шару. У наведеному прикладі є 30×30 = 900 нейронів, тому що є багато місць, де може застосовуватися фільтр 3×3×3. На відміну від традиційних нейронних мереж, де ваги і зміщення нейронів незалежні один від одного, в випадку згорткових нейронних мереж, нейрони, що відповідають одному фільтру в шарі, мають однакові ваги і зміщення. У наведеному вище випадку ми зміщуємо вікно на 1 піксель за раз. Ми також можемо змістити вікно більш ніж на 1 піксель. Це число називається кроком. Як правило, використовують більше одного фільтра в одному шарі згортки. Якщо ми використовуємо 32 фільтра, у нас буде карта активації розміром 30×30×32. Всі нейрони, пов'язані з одним і тим же фільтром, мають однакові ваги та зміщення. Таким чином, кількість ваг при використанні 32 фільтрів - це 3×3×3×32 = 288, а число зміщень – 32.

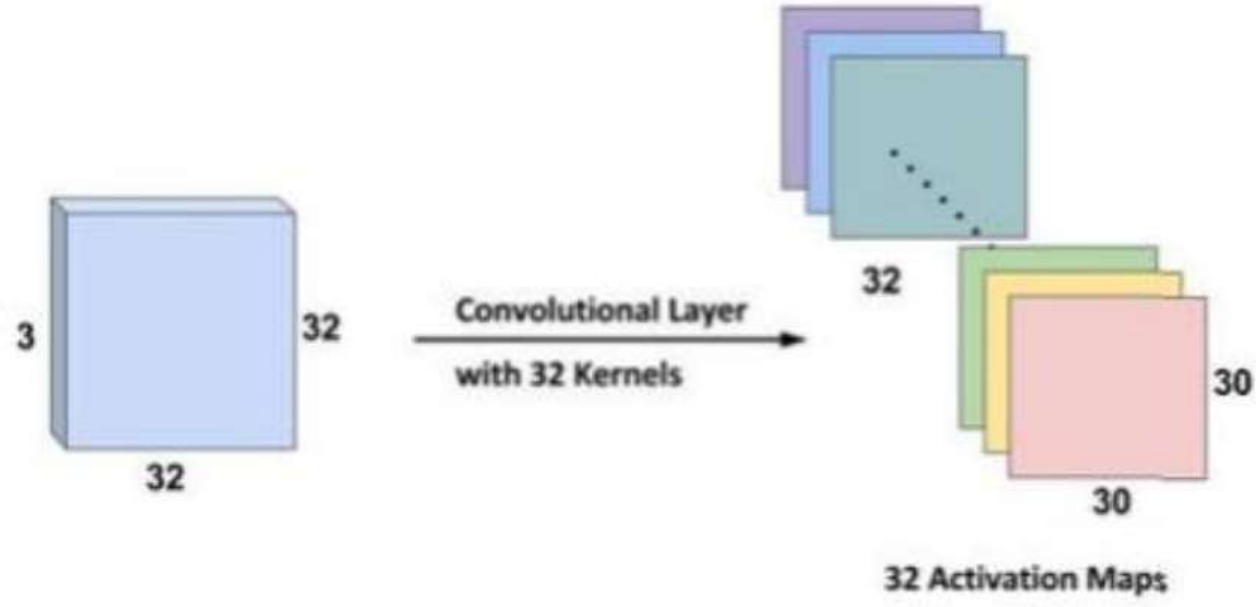


Рис. 1.22 Карти активації, отримані від застосування згорткових ядер.

Після кожної згортки результат зменшується за розміром (так як в цьому випадку ми переходимо від 32×32 до 30×30). Для зручності стандартна

практика полягає в тому, щоб накладати нулі на межу вхідного шару таким чином, щоб вихід був такого ж розміру, які вхідний. Отже, в цьому прикладі, якщо ми додаємо доповнення розміром 1 по обидві сторони від вхідного шару, розмір вихідного рівня буде 32×32×32, що спростить реалізацію. Розглянемо, як згорткові нейронні мережі аналізують зображення.

Початкові шари аналізують менші області зображення і можуть виявляти тільки прості ознаки, такі як межі/кути і т. д. По мірі того як ми йдемо глибше в мережу, нейрони отримують інформацію з більших частин зображення і від різних інших нейронів. Таким чином, нейрони на наступних шарах можуть вивчити більш складні функції [10].

Шар пуллінгу в основному використовується відразу після згорткового шару для зменшення просторового розміру (тільки по ширині і висоті, а не по глибині). Це зменшує кількість параметрів, тому обчислення зменшуються. Використання меншої кількості параметрів дозволяє уникнути перенавчання. Перенавчання – це умова, коли навчена модель відмінно працює з даними навчання, але не дуже добре з тестовими даними. Найбільш поширеною формою пуллінгу є максимальний пулінг, в якому ми беремо фільтр деякого розміру і застосовуємо максимальну операцію max з деякою частиною зображення.



Рис. 1.23 Максимальний пул з розміром фільтра 2×2 і кроком 2.

На рис. 1.23 виходом є максимальне значення в області 2×2, показаної з використанням виділених цифр. Найбільш поширена операція пулінгу

виконується з фільтром розміром 2×2 з кроком 2. Це істотно зменшує розмір введення на половину.

РОЗДІЛ 2 РЕАЛІЗАЦІЯ МОБІЛЬНОГО ДОДАТКУ

2.1. Датасет MNIST для розпізнавання жестів

Створення та навчання нейронної мережі було виконано на ресурсі Kaggle Notebooks - хмарне обчислювальне середовище, що дозволяє створювати та тестувати нейронні мережі. Це тип Jupyter notebooks, середовище складається з послідовності комірок, де кожна комірка відформатована або в Markdown (для введення тексту), або в мову програмування на вибір (для написання коду).

Формат набору даних має шаблон, який відповідає класичному MNIST. Кожен тренувальний і тестовий кейс представляє ярлик (0-25) як індивідуальну карту для кожної букви алфавіту A-Z (і виключає випадки для 9 = J або 25 = Z через рухи жестів). Дані для навчання (27 455 випадків) та тестові дані (7 172 випадки), подані у вигляді pixel1, pixel2 … pixel784, які представляють одне зображення розміром 28x28 пікселів із значенням у градаціях сірого між 0-255.

Імпортуємо тестовий та тренувальний набори даних:

```
train_df = pd.read_csv("../input/sign-language-mnist/sign_mnist_train/sign_mnist_train.csv")
test_df = pd.read_csv("../input/sign-language-mnist/sign_mnist_test/sign_mnist_test.csv")
```

Виведемо перші 5 елементів тренувального набору:

```
train_df.head()
```

	label	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	pixel9	...	pixel775	pixel776	pixel777	pixel778	pixel779	pixel780	pixel781	pixel782	pixel783	pixel784
0	3	107	118	127	134	139	143	146	150	153	...	207	207	207	207	206	206	206	204	203	202
1	6	155	157	156	156	156	157	156	158	158	...	69	149	128	87	94	163	175	103	135	149
2	2	187	188	188	187	187	186	187	188	187	...	202	201	200	199	198	199	198	195	194	195
3	2	211	211	212	212	211	210	211	210	210	...	235	234	233	231	230	226	225	222	229	163
4	13	164	167	170	172	176	179	180	184	185	...	92	105	105	108	133	163	157	163	164	179

5 rows × 785 columns

Рис. 2.1 Елементи тренувального набору

Набір даних train_df складається з 1-го стовпця, що представляє мітки від 1 до 24. Мітка завантажується в окремий фрейм даних, а стовпець 'label' випадає з вихідного навчального кадру даних, який тепер складається лише з 784 піксельних значень для кожного зображення.

Виведемо частоту елементів з набору даних для кожної мітки.

```
plt.figure(figsize = (10,10))
sns.set_style("darkgrid")
sns.countplot(x = train_df['label'].values)
plt.title("Frequency of each label")
```

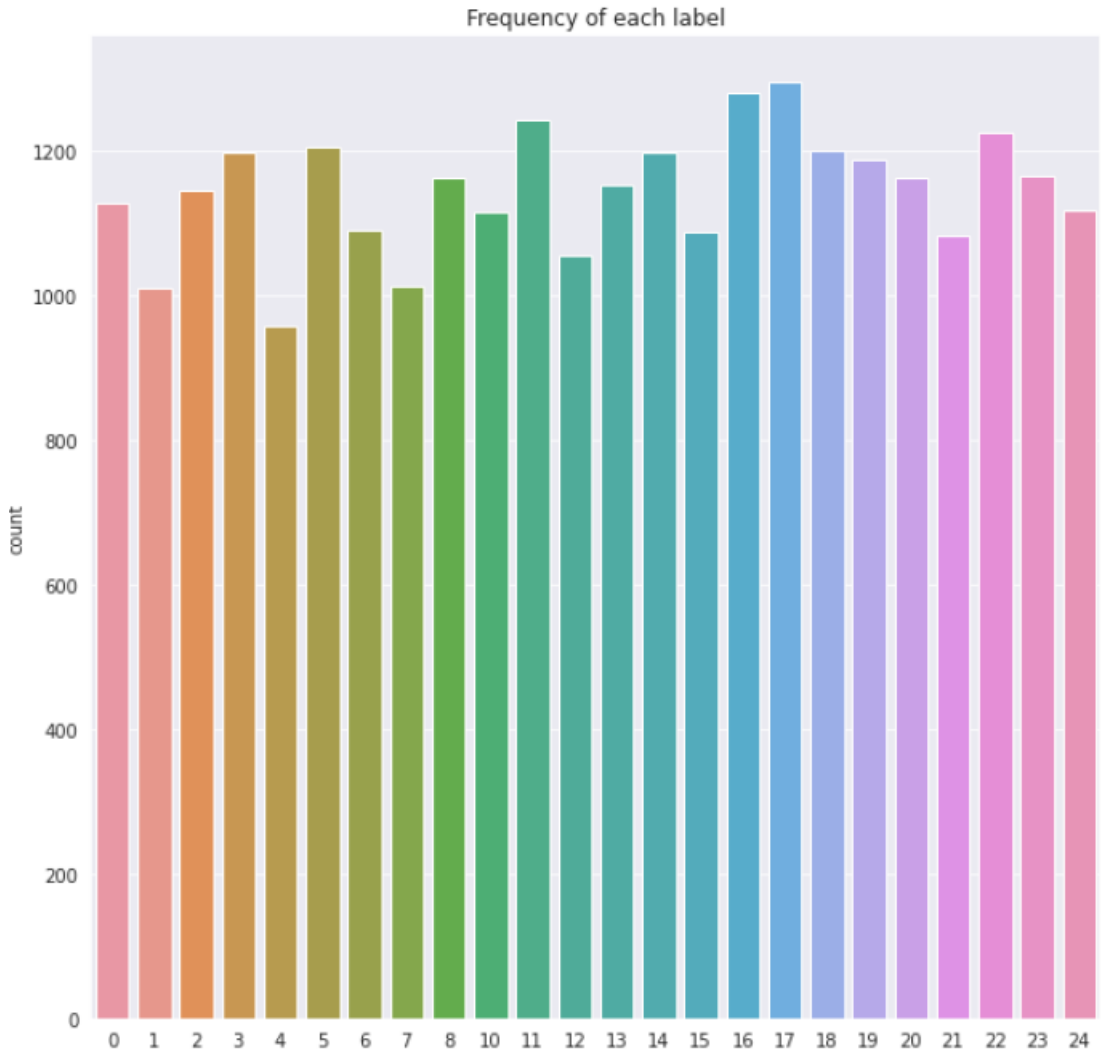


Рис. 2.3 Частота елементів з набору даних для кожної мітки.

Як бачимо, кожен з них розподіляється майже однаково і відкидає стовпчик міток з навчального набору.

Виконаємо перетворення кадру даних у тип масиву numpy, який буде використовуватися під час навчання CNN. Масив перетворюється з 1D в 3D, що є необхідним входом на перший рівень CNN. Ми переформуємо всі дані в 3D-матриці 28×28×1.

Далі бібліотеці Keras потрібна додаткова розмірність, яка відповідає каналам. Зображення мають масштаб сірого, тому вони використовують лише

один канал. Подібну попередню обробку проводимо з тестовим фреймом даних.

```
#Reshaping the data from 1-D to 3-D as required through input by CNN's
x_train = x_train.reshape(-1,28,28,1)
x_test = x_test.reshape(-1,28,28,1)
```

Візуалізуємо зображення тренувального набору даних:

```
f, ax = plt.subplots(2,5)
f.set_size_inches(10, 10)
k = 0
for i in range(2):
        for j in range(5):
                ax[i,j].imshow(x_train[k].reshape(28, 28) , cmap = "gray")
                k += 1
        plt.tight_layout()
```

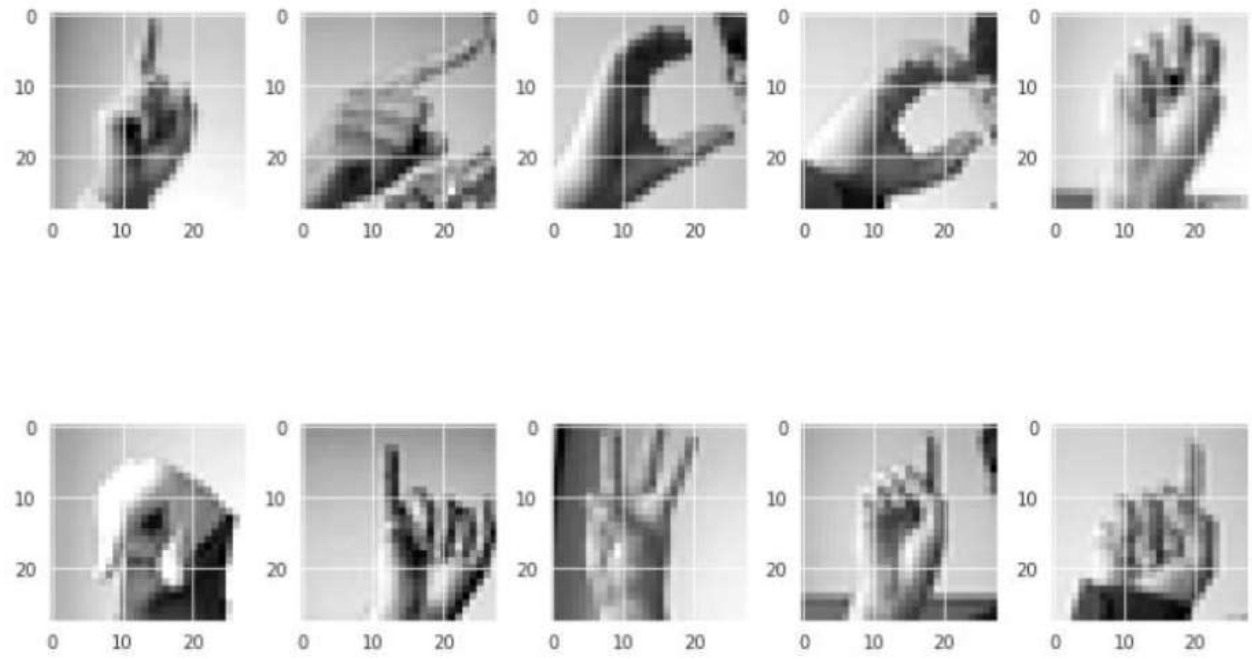


Рис. 2.4 Візуалізація зображень тренувального набору даних.

Пакет ImageDataGenerator від keras.preprocessing.image дозволяє додавати різні спотворення до набору даних зображення, забезпечуючи випадкове обертання, збільшення / зменшення, масштабування висоти або ширини, тощо, до зображень пікселів за пікселем. Проведемо збільшення набору виконавши трансформацію існуючих даних:

```
# With data augmentation to prevent overfitting
datagen = ImageDataGenerator(
```

```
        featurewise_center=False,
        samplewise_center=False,
        featurewise_std_normalization=False,
        samplewise_std_normalization=False,
        zca_whitening=False,
        rotation_range=10,
        zoom_range = 0.1,
        width_shift_range=0.1,
        height_shift_range=0.1,
        horizontal_flip=False,
        vertical_flip=False)
datagen.fit(x_train)
```

Набір даних зображення також необхідно нормалізувати, використовуючи параметр масштабування, який ділить кожен піксель на 255 таким чином, що значення пікселів коливаються від 0 до 1. Ми виконуємо нормалізацію градацій сірого, щоб зменшити ефект різниці освітленості. Завдяки цьому, CNN буде працювати швидше:

```
# Normalize the data
x_train = x_train / 255
x_test = x_test / 255
```

2.2. Побудова моделі CNN

Опишемо шари отриманої моделі нейромережі:

- 3 конволюційні шари (Conv2D) названі від операції згортки. В математиці згортання - це операція над двома функціями, котра видає третю функцію, яка є модифікованою (згорнутою) версією однієї з оригінальних функцій. Отримана функція дає інтеграл точкової мультиплікації двох функцій, як функції від величини, в яку переведена одна з вихідних функцій. Звитий шар складається з груп нейронів, які складають ядра. Ядра мають невеликий розмір, але вони завжди мають ту саму глибину, що і вхідні дані. Нейрони і ядра з'єднані з невеликою областю входу, названою сприйнятливим полем, оскільки вкрай неефективно пов'язати всі нейрони з усіма попередніми виходами у випадку входів великих розмірів, наприклад, зображення розміром 100 x 100 має 10000 пікселів, і якщо перший шар має 100 нейронів, це призведе до 1000000 параметрів.

Замість того, щоб кожен нейрон мав вагу для повного виміру вхідного сигналу, нейрон зберігає вагу для розмірності введення ядра. Ядра ковзають по ширині та висоті вводу, витягують функції високого рівня та створюють двовимірну карту активації. Рядок, за яким ядро ковзає, задається як параметр. Вихід згорткового шару проводиться шляхом складання отриманих карт активації, які в свою чергу використовуються для визначення входу наступного шару. В нашому випадку ці шари мають 75, 50, 25 – карт функцій, відповідно, розмір яких дорівнює 3 × 3 і функцій активації relu.

- 3 шари об'єднання (MaxPool) використовуються з одного боку для зменшення просторових розмірів представлення та зменшення кількості обчислень, проведених в мережі. Ми використовували шари об'єднання, які мають фільтри розміром 2 x 2 з кроком 2. Це ефективно зменшує вхід до чверті його початкового розміру.
- Повно зв’язний шар (Dense) з 128 нейронами і функцією активації softmax. Кожен нейрон із повно зв’язного шару пов'язаний з кожним виходом попереднього шару. Операції над згортковим шаром такі ж, як і в повно зв’язному шарі. Таким чином, можливе перетворення між ними двома.
- 2 шари регуляризації Dropout. Він налаштований на випадкове виключення 20-30% нейронів в шарі, щоб зменшити перенавчання.
- Шар Flatten, який перетворює дані двовимірної матриці в вектор(Flatten). Він дозволяє обробляти вихідні дані стандартними повнозв'язними шарами.
- Шар BatchNormalization, що нормалізує вхідні дані. Пакетна нормалізація застосовує перетворення, яке підтримує середнє значення виведення, близьке до 0, а стандартне відхилення виведення близько до 1.
- Вихідний шар з 24 одиницями для 24 різних класів.

```
model = Sequential() // ініціалізуємо модель
model.add(Conv2D(75 , (3,3) , strides = 1 , padding = 'same' , activation = 'relu' , input_shape = (28,28,1))) // додаємо шари до моделі
```

```
model.add(BatchNormalization())
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))
model.add(Conv2D(50 , (3,3) , strides = 1 , padding = 'same' , activation = 'relu'))
model.add(Dropout(0.2))
model.add(BatchNormalization())
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))
model.add(Conv2D(25 , (3,3) , strides = 1 , padding = 'same' , activation = 'relu'))
model.add(BatchNormalization())
model.add(MaxPool2D((2,2) , strides = 2 , padding = 'same'))
model.add(Flatten())
model.add(Dense(units = 512 , activation = 'relu'))
model.add(Dropout(0.3))
model.add(Dense(units = 24 , activation = 'softmax'))
model.compile(optimizer = 'adam' , loss = 'categorical_crossentropy' , metrics = ['accuracy'])
model.summary()
```

Отримана структура шарів моделі нейромережі:

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_3 (Conv2D)            (None, 28, 28, 75)        750
_________________________________________________________________
batch_normalization_3 (Batch (None, 28, 28, 75)        300
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 14, 14, 75)        0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 14, 14, 50)        33800
_________________________________________________________________
dropout_2 (Dropout)          (None, 14, 14, 50)        0
_________________________________________________________________
batch_normalization_4 (Batch (None, 14, 14, 50)        200
_________________________________________________________________
max_pooling2d_4 (MaxPooling2 (None, 7, 7, 50)          0
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 7, 7, 25)          11275
_________________________________________________________________
batch_normalization_5 (Batch (None, 7, 7, 25)          100
_________________________________________________________________
max_pooling2d_5 (MaxPooling2 (None, 4, 4, 25)          0
_________________________________________________________________
flatten_1 (Flatten)          (None, 400)               0
_________________________________________________________________
dense_2 (Dense)              (None, 512)               205312
_________________________________________________________________
dropout_3 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_3 (Dense)              (None, 24)                12312
=================================================================
Total params: 264,049
Trainable params: 263,749
Non-trainable params: 300
_________________________________________________________________
```

Рис. 2.5 Структура нейронної мережі.

Проводимо навчання вибраного налаштування нейронної мережі в 20 епох:
history = model.fit(datagen.flow(x_train,y_train, batch_size = 128) ,epochs = 20 , validation_data = (x_test, y_test) , callbacks = [learning_rate_reduction])

```
Epoch 1/20
215/215 [==============================] - 57s 261ms/step - loss: 1.8160 - accuracy: 0.4646 - val_loss: 3.3002 - val_accuracy: 0.1390
Epoch 2/20
215/215 [==============================] - 57s 264ms/step - loss: 0.2578 - accuracy: 0.9153 - val_loss: 1.2996 - val_accuracy: 0.5673
Epoch 3/20
215/215 [==============================] - 57s 265ms/step - loss: 0.1188 - accuracy: 0.9607 - val_loss: 0.1341 - val_accuracy: 0.9575
Epoch 4/20
215/215 [==============================] - 57s 264ms/step - loss: 0.0672 - accuracy: 0.9791 - val_loss: 0.1401 - val_accuracy: 0.9505
Epoch 5/20
215/215 [==============================] - 57s 263ms/step - loss: 0.0498 - accuracy: 0.9848 - val_loss: 0.1327 - val_accuracy: 0.9551

Epoch 00005: ReduceLROnPlateau reducing learning rate to 0.0005000000237487257.
Epoch 6/20
215/215 [==============================] - 57s 265ms/step - loss: 0.0278 - accuracy: 0.9924 - val_loss: 0.0795 - val_accuracy: 0.9743
Epoch 7/20
215/215 [==============================] - 57s 264ms/step - loss: 0.0207 - accuracy: 0.9937 - val_loss: 0.0072 - val_accuracy: 0.9993
Epoch 8/20
215/215 [==============================] - 56s 263ms/step - loss: 0.0149 - accuracy: 0.9957 - val_loss: 0.0055 - val_accuracy: 0.9990
Epoch 9/20
215/215 [==============================] - 56s 261ms/step - loss: 0.0155 - accuracy: 0.9959 - val_loss: 0.0126 - val_accuracy: 0.9958

Epoch 00009: ReduceLROnPlateau reducing learning rate to 0.00025000000118743628.
Epoch 10/20
215/215 [==============================] - 57s 263ms/step - loss: 0.0128 - accuracy: 0.9958 - val_loss: 0.0164 - val_accuracy: 0.9964
Epoch 11/20
215/215 [==============================] - 57s 264ms/step - loss: 0.0085 - accuracy: 0.9984 - val_loss: 0.0028 - val_accuracy: 0.9999
Epoch 12/20
215/215 [==============================] - 57s 265ms/step - loss: 0.0085 - accuracy: 0.9981 - val_loss: 0.0047 - val_accuracy: 0.9994
Epoch 13/20
215/215 [==============================] - 57s 266ms/step - loss: 0.0083 - accuracy: 0.9978 - val_loss: 0.0056 - val_accuracy: 0.9999

Epoch 00013: ReduceLROnPlateau reducing learning rate to 0.00012500000059371814.
Epoch 14/20
215/215 [==============================] - 57s 266ms/step - loss: 0.0064 - accuracy: 0.9983 - val_loss: 0.0021 - val_accuracy: 0.9999
Epoch 15/20
215/215 [==============================] - 58s 268ms/step - loss: 0.0057 - accuracy: 0.9985 - val_loss: 0.0060 - val_accuracy: 0.9967

Epoch 00015: ReduceLROnPlateau reducing learning rate to 6.25000029685907e-05.
Epoch 16/20
215/215 [==============================] - 57s 265ms/step - loss: 0.0060 - accuracy: 0.9981 - val_loss: 0.0055 - val_accuracy: 0.9972
Epoch 17/20
215/215 [==============================] - 57s 265ms/step - loss: 0.0050 - accuracy: 0.9984 - val_loss: 0.0032 - val_accuracy: 1.0000
Epoch 18/20
215/215 [==============================] - 57s 264ms/step - loss: 0.0051 - accuracy: 0.9989 - val_loss: 0.0018 - val_accuracy: 0.9997
Epoch 19/20
215/215 [==============================] - 57s 266ms/step - loss: 0.0043 - accuracy: 0.9990 - val_loss: 0.0023 - val_accuracy: 0.9997

Epoch 00019: ReduceLROnPlateau reducing learning rate to 3.125000148429535e-05.
Epoch 20/20
215/215 [==============================] - 57s 265ms/step - loss: 0.0044 - accuracy: 0.9990 - val_loss: 0.0021 - val_accuracy: 0.9999
```

Рис. 2.6 Процес навчання нейронної мережі в Kaggle Notebooks.

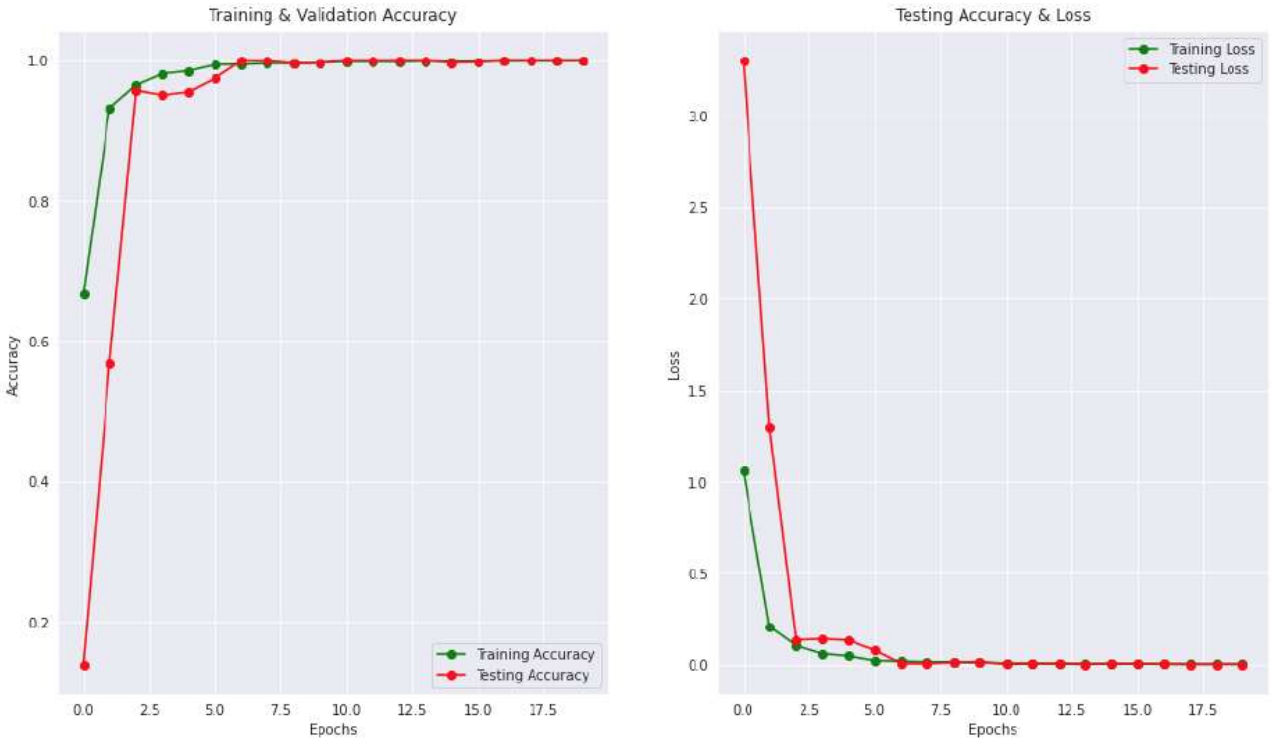


Рис. 2.7 Графік росту точності та помилки

Як бачимо, зі збільшенням кількості епох точність також зростає. Виведемо результуючу точність на тестовому наборі даних:

```
print("Accuracy of the model is - " , model.evaluate(x_test,y_test)[1]*100 , "%")
```

```
225/225 [==============================] - 3s 14ms/step - loss: 0.0021 - accuracy: 0.9999
Accuracy of the model is -  99.98605847358704 %
```

Рис. 2.8. Результуюча точність створеної моделі нейромережі.

2.3. Матриця помилок (confusion matrix)

Нехай дана вибірка X_i кожен об'єкт якої відноситься до одного з C класів і класифікатор α, який ці класи прогнозує. Матрицею помилок для такого класифікатора називається наступна матриця:

$$M = \{m_{ij}\}_{i,j=0}^{C}, m_{ij} = \sum_{k=0}^{N} I[a(x_k) = j]\, I[y_k = i] \quad (2.1)$$

Така матриця показує скільки об'єктів класу i були розпізнані як об'єкти класу j.

```
cm = confusion_matrix(y,predictions)
cm = pd.DataFrame(cm , index = [i for i in range(25) if i != 9] , columns = [i for i in range(25) if i != 9])
plt.figure(figsize = (15,15))
      sns.heatmap(cm,cmap= "Blues", linecolor = 'black' , linewidth = 1 , annot = True, fmt='')
```

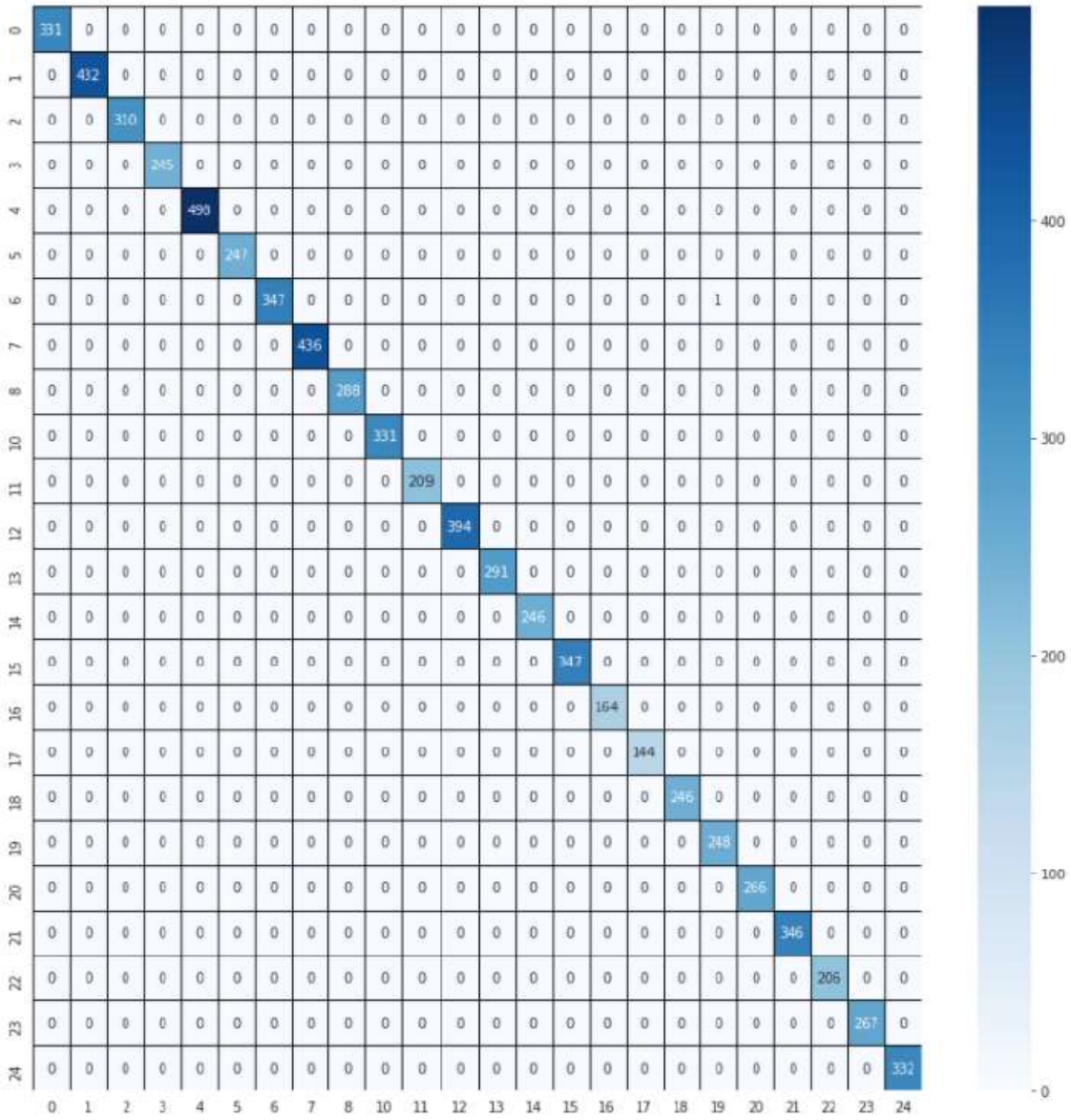


Рис 2.9. Матриця помилок.

Імпортуємо модель в формат json і ваги нейронів в формат bin:

```
tfjs.converters.save_keras_model(model, '/kaggle/working')
```

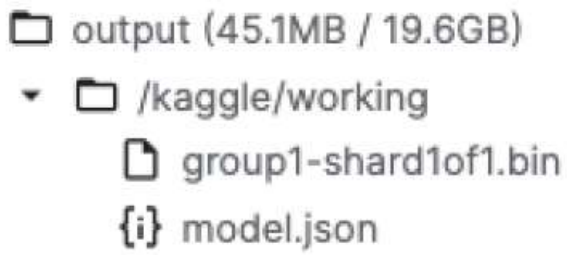


Рис. 2.10 Отримані файли моделі та ваг нейромережі.

2.4. Огляд роботи мобільного додатку

Мобільний додаток складається з двох екранів: Camera та Tutorial. Їх налаштування виконуємо за допомогою бібліотеки react-navigation.



Рис. 2.11 Екран камери.



Рис. 2.12 Екран туторіалу.

```
import * as React from 'react'
import { NavigationContainer } from '@react-navigation/native'
import { createStackNavigator } from '@react-navigation/stack'
import { createBottomTabNavigator } from '@react-navigation/bottom-tabs'
import Icon from 'react-native-vector-icons/Ionicons'
import Camera from '../screens/Camera'
import Tutorial from '../screens/Tutorial'
const Stack = createStackNavigator()
const Tab = createBottomTabNavigator()
const ICON_SIZE = 36
const MyTabs = () => (
        <Tab.Navigator>
        <Tab.Screen
                name='Camera'
                component={Camera}
                options={{
                        tabBarIcon: ({ color }) => {
                                return <Icon name={'ios-camera'} size={ICON_SIZE} color={color} />
                        }
                }}
        />
```

```
        <Tab.Screen
        name='Tutorial'
        component={Tutorial}
        options={{
                tabBarIcon: ({ color }) => {
                        return <Icon name={'ios-information-circle'} size={ICON_SIZE} color={color}
                />
                }
        }}
/>
        </Tab.Navigator>
)
const MainNavigation = () => (
        <NavigationContainer>
                <Stack.Navigator headerMode='none'>
                        <Stack.Screen name='Main' component={MyTabs} />
                </Stack.Navigator>
                </NavigationContainer>
        )
```

Для отримання доступу до камери буда використана бібліотека expocamera. Після отримання зображення з камери нам потрібно провести деякі маніпуляції, перед тим, як відправити його на вхід до моделі нейромережі.

Спочатку потрібно обрізати зображення по рамках, для цього використовуємо бібліотеку expo-image-manipulator.

Данна функція cropPicture на вхід приймає об’єкт зображення, отриманий с камери та maskDimension – це розмір області рамки, яку бачить користувач, у нас це константа MASK_DIMENSION яка відповідає значенню 400 пікселів. Але на вхід нейромережа приймає розмірність вхідних бітових даних 28x28, тож далі ми змінимо розмір зображення відповідно нашим потребам, для цього тут використовуємо константа BITMAP_DIMENSION яка відповідає значенню 28 пікселів.

```
const cropPicture = async (imageData, maskDimension) => {
        const { uri, width, height } = imageData
        const cropWidth = maskDimension * (width / dimension.deviceWidth)
        const cropHeight = maskDimension * (height / dimension.deviceHeight)
        const actions = [
                {
                        crop: {
                                originX: width / 2 - cropWidth / 2,
                                originY: height / 2 - cropHeight / 2,
                                width: cropWidth,
                                height: cropHeight
```

```
                }
            },
            {
            resize: {
                width: BITMAP_DIMENSION,
                height: BITMAP_DIMENSION
            }
        }
]

const saveOptions = {
        compress: 1,
        base64: true
        }
        return await ImageManipulator.manipulateAsync(uri, actions, saveOptions)
})
```

Ми отримали зображення 28x28, яке потрібно для входу нашої нейромережі, але це зображення в форматі rgba, а нейромережа працює з відтінками сірого, тож конвертуємо значення кольорів пікселі отриманого зображення в відтінки сірого. Для цього використовуємо canvas – це елемент HTML5, який застосовується для малювання графіки використовуючи скрипти (переважно JavaScript). Його можна застосувати для малювання графів, створення фотокомпозицій, а також анімації. Спочатку ми створюємо об'єкт canvas нашого зображення, потім отримуємо всі данні нашого цього зображення із створеного об’єкта та отримуємо із цих даних значення RGBA.

Щоб перетворити колір з колірного простору моделі RGBa, ми використаємо формулу зваженої суми, вона застосовується до лінійних компонентів кольору R_{linear}, G_{linear}, B_{linear} для обчислення лінійної яскравості Y_{linear}, це і будуть потрібні нам бітові значення в відтінках сірого.

$$Y_{linear} = 0.299 * R_{linear} + 0.587 * G_{linear} +,0.114 * B_{linear} \qquad (2.2)$$

```
const handleCanvas = async (canvas) => {
        if (canvas && cropperImageUri) {
                const image = new CanvasImage(canvas)
                canvas.width = 28
                canvas.height = 28
                const context = canvas.getContext('2d')
                image.src = `${cropperImageUri.base64}`
                image.addEventListener('load', () => {
```

```
                        context.drawImage(image, 0, 0, 28, 28)
                        })
        let imageData = await context.getImageData(0, 0, 28, 28)
        const RGBAValues = Object.values(imageData.data)
        const grayStyleArray = []
for (let i = 0; i < RGBAValues.length; i += 4) {
        let avg = Number(((0,229*RGBAValues[i] + 0,587*RGBAValues[i + 1] + 0,114*RGBAValues[i
+ 2]) / 255).toFixed(2))
        imageData.data[i] = avg
        imageData.data[i + 1] = avg
        imageData.data[i + 2] = avg
        grayStyleArray.push(avg)
        }
        context.putImageData(imageData, 0, 0)
        setGrayStyleImageData(grayStyleArray)
        }
}
```

Далі імпортуємо модель створену на Kaggle та ваги за допомогою метода loadLayersModel. Метод convertToTensor формує тензор використовуючи отримані раніше бітові значення в відтінках сірого. В методі startPrediction ми передаємо тензор на вхід, а на виході отримуємо масив ймовірностей від 0 до 1 для кожного значення букви. В методі populateData ми обробляємо цей масив та фільтруємо значення ймовірностей за найбільшими значеннями.

```
const modelJson = require('../assets/model/model.json')
const modelWeights = require('../assets/model/weights.bin')
const TENSORFLOW_CHANNEL = 1
export const getModel = async () => {
        await tf.ready()
        return await tf.loadLayersModel(bundleResourceIO(modelJson, modelWeights))
}

export const convertToTensor = async (props) => {
        return tensor(props, [1, BITMAP_DIMENSION, BITMAP_DIMENSION,
TENSORFLOW_CHANNEL])
}
export const startPrediction = async (model, tensor) => {
        const output = await model.predict(tensor)
        return output.dataSync()
}
const LABELS = ['A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W',
'X', 'Y']
export const populateData = (typedArray) => {
        const predictions = Array.from(typedArray)
        return predictions.map((item, index) => {
                if (item > 0.7) {
                return {
```

```
                        label: LABELS[Number(index)],
                        prediction: item
                        }
                } else return false
        }).filter(Boolean)
}
```

ВИСНОВКИ

Був проведений огляд предметної області нейронної мережі та її основних математичних складових.

Було обрано згорткову нейронну мережу, яка була змодельована та навчена на ресурсі Kaggle Notebook. Вона була реалізована з використанням бібліотеки Keras мовою Python. Також на ресурсі Kaggle Notebook були проведені експерименти, та було обрано найбільш точне налаштування нейронної мережі. Була досягнута точність на тестовому наборі - 99.98%.

Результатом роботи є розроблений та протестований мобільний додаток визначення мовних жестів на основі штучного інтелекту, та описані основні математичні положення нейронної мережі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. A. Gordo, J. Almazan, J. Revaud, and D. Larlus. Deep image ´ retrieval: Learning global representations for image search. In European Conference on Computer Vision, 2016.

2. Gradient-Based Learning Applied to Document Recognition: URL: http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf

3. David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. nature, 323(6088):533–536, 1986.

4. PMBOK - Work Breakdown Structure URL: https://www.workbreakdownstructure.com/work-breakdown-structure-accordingtopmbok.php

5. Ciresan, D. C., Meier, U., and Schmidhuber, J. Multi-column deep neural networks for image classification.

6. Martin Arjovsky and Leon Bottou. Wasserstein generative adversarial networks. In Proceedings of the 34 th International Conference on Machine Learning, Sydney, Australia, 2017.

7. Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. Proc. ICLR 2019.

8. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.

9. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. nature, 521(7553):436–444, 2015.

10. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1):1929–1958, 2014.