ГЛОБАЛЬНА ЗГОРТКА

МОДЕЛЬ ДВОВИМІРНОГО ШАРУ НЕЙРОННОЇ МЕРЕЖІ НА ОСНОВІ ГЛОБАЛЬНОЇ ЗГОРТКИ

ЗМІСТ

АНОТАЦІЯ

Роботу виконано на 24 аркушах, вона містить перелік посилань на використані джерела з 13 найменувань. У роботі наведено 3 рисунки та 3 таблиці.

Актуальність теми. Впродовж останніх років у сфері глибинного навчання відбуваються революційні досягнення. Прикладами є нейронна мережа для генерації зображень високого розширення Stable Diffusion, велика мовна модель ChatGPT. Разом з цими досягненнями зростає і спектр застосунку штучних нейронних мереж у прикладних задачах. Такі як магазини самообслуговування Amazon Go та безпілотні автомобілі Tesla. Втім, результати досі не ідеальні, і на класичних датасетах все ще немає моделей, які б не робили помилок. Одним з таким класом задач є комп'ютерний зір.

Наше дослідження буде спрямоване саме на покращення нейромережевих моделей, що працюють з зображеннями, а точніше їхнього конкретного сімейства – згорткових нейронних мереж, що побудовані навколо математичної операції згортки. Знайти спосіб покращити якість роботи нейронних мереж з зображенням є дійсно актуальною задачею, оскільки згадані раніше Amazon Go та Tesla не обійшлись би без них, а покращення існуючих результатів може стати причиною появи нових успішних підприємств і продуктів, що використовують машинне навчання.

Мета і задачі дослідження. Метою і задачею дослідження є розробка моделі двовимірного шару нейронної мережі з використанням згорткових фільтрів розмірність яких еквівалента розмірності вхідних даних до шару.

Наукова новизна. У сучасних згорткових нейронних мережах не використовуються фільтри такої розмірності, тому результати цього дослідження можуть принести користь для розвитку нового покоління згорткових нейронних мереж.

Методи дослідження. Системний аналіз, методи машинного навчання.

Ключові слова: *Машинне навчання, глибинне навчання, нейронні мережі, згорткові нейронні мережі, класифікація зображень, комп'ютерний зір*

ВСТУП

З появою нейронних мереж типу Transformer [1], найкращі результати у задачах обробки тексту утримували нейронні мережі саме з цією архітектурою. Незабаром, ідею цієї архітектури було адаптовано та оптимізовано під задачі комп'ютерного зору, такі як класифікація зображень. Однією з таких архітектур є Vision Transformer [2], що з'явилась у 2020 році. Згорткові нейронні мережі, що базувались на математичній операції згортки були не в стані отримати такі високі результати. Тому, з того часу, більшість нових опублікованих досліджень було зосереджено саме навколо моделей архітектури типу Transformer.

Втім, жодна зі згаданих архітектур нейронних мереж не була в стані ефективно працювати з довгими даними. Під терміном «довгі» розуміються дані, що є послідовними, великими за обсягом, але при цьому є атомарними, тобто не підлягають розкладу на частини. Під неефективністю розуміється, що нейронна мережа або не в стані генерувати результати з задовільною точністю, або ж вимагає значних обчислювальних потужностей. Прикладом такої задачі є пошук ключових слів у аудіо записі, як у датасеті SC10 [3].

У цій задачі, інтерес викликає одна з state-of-the-art моделей: Structured State Space sequence model (S4) [4]. Для цього дослідження ця модель важлива, оскільки вона в змозі оброблювати глобальний контекст вхідних даних та утилізує операцію згортки. S4 має високі результати у різних задачах, включаючи обробку звуку та класифікацію зображень. Проте, зображення є, фактично, двовимірними або тривимірними матрицями (в залежності від відсутності або наявності кольорових каналів у ньому) з просторами висота, ширина та колір (при наявності) і для того, аби модель була в стані обробити це зображення, простір висоти та ширини поєднують в одне. Логічним є припущення, що при такому підході, структурна інформація про зображення буде дещо втрачена.

Тому, дана робота присвячена розробці шару нейронної мережі, що використовуватиме двовимірні згорткові фільтри з глобальним контекстом.

Іншими словами, задачею є створення такої архітектури нейронної мережі, що дозволить не порушувати початкову структуру зображень, а також оброблюватиме усі точки зображення одночасно.

Загалом, розробка нової архітектури шару згортки, що матиме кращі результати, дозволить створювати більш якісні продукти на основі комп’ютерного зору. Такими продуктами можуть бути: автопілоти автомобілів, автоматизовані системи безпеки, автоматизовані магазини, тощо.

ОГЛЯД ІСНУЮЧИХ РІШЕНЬ

SGConv. В більшості сучасних архітектур (ResNet [5], MobileNet [6]), згорткові фільтри мають достатньо невелику висоту і ширину ($N = M \leq 7$). Одна з очевидних причин такої розмірності – зменшення кількості параметрів для оптимізації. Це дозволяє тренувати моделі швидше та потребує менш потужних машин для тренування та експлуатації моделей. Втім використання фільтру такої розмірності означає, що згортка буде *локальною*, або, іншими словами, охоплюватиме лише частину контексту за раз.

Втім, деякі нещодавні дослідження були напрямлені на експерименти з *глобальною* згорткою, в яких розмірність згорткового фільтру еквівалентна розмірності вхідних даних. Одною з таких моделей є *Structured State Space sequence model* (S4) [4], яка є першою ефективною моделлю такого типу. З назви можна зрозуміти, що це є моделлю простору станів (state space model), відомий з теорії керування спосіб репрезентації поведінки динамічної системи. Незважаючи на високі результати в тестах, дана модель є доволі складною для розуміння і вимагає освіченості не лише у машинному навчанні, а й у теорії диференційних рівнянь та дотичних дисциплін.

Тому, іншою групою вчених було опубліковано нову архітектуру моделі глобальної згортки – ***SGConv*** [7] (*Структурована Глобальна Згортка*), яка намагається *«спростити складнощі моделі S4»* і не базується на теорії диференційних рівнянь. Хоча і модель позиціонують як спрощення S4, вона має кращі результати у більшості експериментів. У попередній роботі [8] було встановлено принципи на яких базується дана модель:

1) Кількість параметрів масштабується від довжини вхідних параметрів сублінійно. Суть підходу полягає в тому, що згортковий фільтр складається з послідовності під-фільтрів, які мають однакову кількість параметрів, втім розмірність кожного наступного під-фільтра послідовності в два рази більша за розмірність попереднього. Це досягається завдяки лінійній інтерполяції

параметрів під-фільтру. Такий підхід дозволяє забезпечити логарифмічну залежність між кількістю параметрів шару та довжиною вхідної послідовності.

2) При об'єднанні під-фільтрів у один фільтр відбувається процес зважения кожного з них. Вводиться параметр затухання величини параметрів, який збільшується для кожного наступного під-фільтру. Завдяки затуханню сусідні елементи у послідовності мають більший вплив на результат, ніж віддаленні. Це дозволяє підвищити можливості моделі до узагальнення і, відповідно, покращити остаточні результати.

Базуючись на цих двох принципах формулу для побудови глобального згорткового фільтру можна представити наступним чином:

$$Global\ Kernel = [k_1, k_2, \dots, k_N];\ \ k_i = \alpha^i \cdot Upsample_{2^{i-1}}(w_i) \quad (1)$$

Де $[\dots]$ – операція конкатенації, $Upsample_s$ – операція збільшення розмірності в s разів шляхом лінійної інтерполяції, w_i – параметри i-го під-фільтру, α – коефіцієнт затухання (типове значення $\frac{1}{2}$).

За цією формулою можна побачити, як фільтр є результатом конкатенації послідовності фільтрів, де кожен наступний під-фільтр збільшує свою розмірність вдвічі від попереднього, а також має меншу величину, внаслідок коефіцієнту затухання. Дана ідея побудови фільтру використовується у одновимірних задачах, такі як обробка сигналів та текстів, тому розмірність такого фільтру складає LxC, де L – довжина вхідної послідовності, а C – кількість характеристик або каналів. Під-фільтри, в свою чергу, мають розмірність L_1xC, де L_1 – задана константа. Конкатенація під-фільтрів відбувається вдовж першого виміру.

АРХІТЕКТУРА ШАРУ ДВОВИМІРНОЇ ГЛОБАЛЬНОЇ ЗГОРТКИ

Генерація одновимірного згорткового фільтру для задач комп'ютерного зору. У попередній роботі [8] було описано процедуру створення двовимірного глобального згорткового фільтру. Даний фільтр є ключовим компонентом архітектури, яка пропонується у цій роботі. Більш детально опишемо принцип побудови шару, який було запропоновано у попередній публікації.

У цьому підпункті, буде описано принцип ініціалізації згорткового шару з врахуванням факту, що замість часових рядів, вхідними даними виступатимуть зображення. Це є важливим пунктом, оскільки в оригінальній моделі [7], ініціалізація відбувається таким чином, що більшість параметрів для оптимізації зосереджено на краю фільтру. Це є логічним рішенням для часових рядів, втім не є цілком доречним для зображень.

Для зображень більшість параметрів для оптимізації має знаходитись у центрі та сублінійно зменшуватись до країв фільтру. Такий підхід базується на припущенні, що сусідні пікселі несуть більше інформації, ніж віддалені. Іншими словами, фільтр глобальної згортки складається з послідовності під-фільтрів, кожен з яких має однакову кількість параметрів, але розмірність кожного фільтру збільшується вдвічі від центру до країв за рахунок лінійної інтерполяції. Запропонований нами фільтр зображено на Рисунку 1.

Рисунок 1. Глобальний згортковий фільтр загальної довжини 26. Кожен під-фільтр зображено різним кольором. Кількість параметрів у одному під-фільтрі – 2. Сумарна кількість параметрів – 10 (5 під-фільтрів по 2 параметри).

Необхідно встановити, зі скількох під-фільтрів складається глобальний згортковий фільтр довжини L, якщо кожен під-фільтр має k параметрів. Запишемо наступне рівняння:

$$k + 2 \cdot 2k + 2 \cdot 4k + 2 \cdot 8k + \cdots = L \quad (2)$$

Перший множник 2 у доданках означає, що під-фільтри додаються зліва та справа від центрального одночасно, а значення nk відповідає розміру під-фільтру після його масштабування. Фактично, необхідно знайти кількість доданків у лівій частині рівняння (2), без врахування центрального під-фільтру. Виконаємо деякі перетворення:

$$\tilde{L} = L - k$$

$$4k + 2 \cdot 4k + 2 \cdot 8k + \cdots = \tilde{L} \quad (3)$$

Користуючись формулою суми геометричної прогресії:

$$\tilde{L} = \frac{4k(1 - 2^n)}{1 - 2}$$

$$-\frac{\tilde{L}}{4k} = 1 - 2^n$$

$$2^n = 1 + \frac{\tilde{L}}{4k}$$

$$n = \log_2\left(1 + \frac{\tilde{L}}{4k}\right) = \log_2\left(1 + \frac{L - k}{4k}\right) \quad (4)$$

Тепер за формулою (3) можна розрахувати, скільки під-фільтрів має бути створено при побудові моделі. Також варто відмітити, що результатом формули може бути ненатуральне число. В таких випадках результат буде округлено до більшого цілого числа, результуючий згортковий фільтр матиме розмірність більше за L, тому «зайві» параметри фільтри на краях буде відкинуто для того, щоб розмірність фільтру була еквівалентна L.

Генерація двовимірного згорткового фільтру. Користуючись підходом описаним у попередньому пункті, можна згенерувати два окремих одновимірних згорткових фільтри розмірності LxC, де L – ширина або висота зображення, а C – кількість каналів (у випадку RGB зображення – 3 канали).

Ці два фільтри, можна об’єднати у один фільтр більшої розмірності за допомогою операції *прямого добутку (outer product)*. Розглянемо на прикладі векторів-стовпчиків:

$$a \otimes b^T \rightarrow \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \otimes [b_1 \quad b_2] = \begin{bmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \\ a_3 b_1 & a_3 b_2 \end{bmatrix}$$

Бачимо, що результатом тензорного добутку двох тензорів є тензор більш високого рангу. За допомогою цієї операції можна розширити розмірність двох одновимірних фільтрів до двовимірного випадку. Результатом буде фільтр розмірності $LxLxC$ (за умови, якщо висота та ширина однакові).

Тобто, процедура ініціалізації шару двовимірної глобальної згортки включатиме генерацію двох одновимірних глобальних згорткових фільтрів окремо для висоти, окремо для ширини та об’єднання їх в один за допомогою операції прямого добутку. Даний підхід є доволі простим, але має певну особливість, яка потребує уваги: операція множення вплине на значення параметрів, їх знак та розподіл. Ініціалізація вагів нейронної мережі може мати значний вплив на фінальні результати. Параметри фільтрів ініціалізуються випадковим чином, фактично, вони є випадковими величинами взятими з певного розподілу. Найтиповіші розподіли, які використовуються для нейронних мереж це рівномірний та нормальний. Втім, при множенні двох випадкових величин, функція розподілу може змінитись. Тому у цій роботі також буде звернено увагу на вибір функції ініціалізації параметрів і перевірено декілька варіантів.

У зв’язку з вищеописаним впливом операції прямого добутку на розподіл на параметри, також буде застосовано нормалізацію для згенерованого згорткового фільтру GK:

$$\widetilde{GK_c} = \frac{GK_c}{\|GK_c\|}$$

Де c – канал згорткового фільтру, $\|GK_c\|$– норма матриці.

Дану нормалізацію буде застосовано лише при генерації фільтру для стабілізації початку навчання, під час тренування нейронної мережі її застосовано не буде.

Лінійне з'єднання між згорткою, нормалізація та активація. Окрім використання фільтру глобальної згортки додатково додамо паралельний лінійний шар до моделі. Його функція нагадуватиме принцип надлишкового з'єднання моделей ResNet, але окрім цього матиме власні ваги, що можуть бути налаштовані.

Якщо розмірність вхідних даних X до шару складатиме $[H, W, C]$, то буде створено згортковий фільтр GK розмірності $[H, W, C]$ та лінійний шар L розмірності $[C, 1]$. Тоді результат Y буде обчислено за формулою:

$$Y = GK * X + XL$$

Де $*$ – операція згортки.

Подальшим кроком буде використання певного шару нормалізації. Буде реалізовано підтримку для Batch Normalization [9]:

$$\hat{x} = \frac{x - E[x]}{\sqrt{Var[x] + \epsilon}}$$

Де, x – вхідні дані до шару, $E[x]$ – середнє значення, $Var[x]$ – дисперсія, ϵ – константа (зазвичай дорівнює $1e - 05$)

Нормалізовані дані буде передано до функції активації. У даній архітектурі було обрано функцію GELU (Gaussian Error Linear Unit) як основну:

$$GELU(x) = x * \Phi(x)$$

Де $\Phi(x)$ – функція розподілу ймовірностей стандартного нормального розподілу.

За аналогією до одновимірної згортки у SGConv також буде включено підтримку декількох «голів». Кожна «голова» фактично представлятиме один шар двовимірної глобальної згортки. Їх об’єднання в один шар дозволить значно підвищити кількість каналів у тензорах і мати більше інформації для нейронної мережі. Це необхідно через особливості згортки по глибині, яка виключає можливість встановлення кількості вихідних каналів вручну.

Візуалізація архітектури шару двовимірної глобальної згортки. Зобразимо графічно модель шару двовимірної глобальної згортки, що відображатиме компоненти даного шару а також порядок проходження вхідної інформації через ці компоненти, на Рисунку 2:

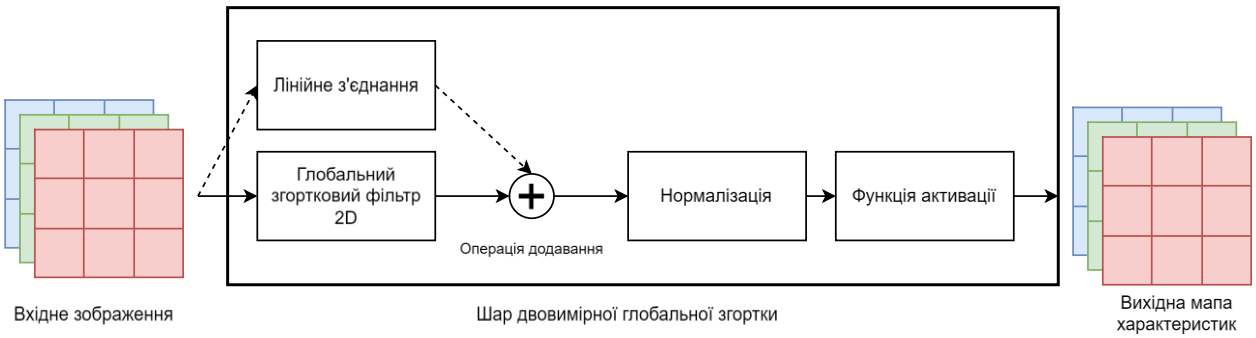

Рисунок 2. Архітектура шару двовимірної глобальної згортки

Але оскільки в шарі буде також реалізована підтримка декількох «голів», то повна архітектура шару виглядатиме так, як на Рисунку 3:

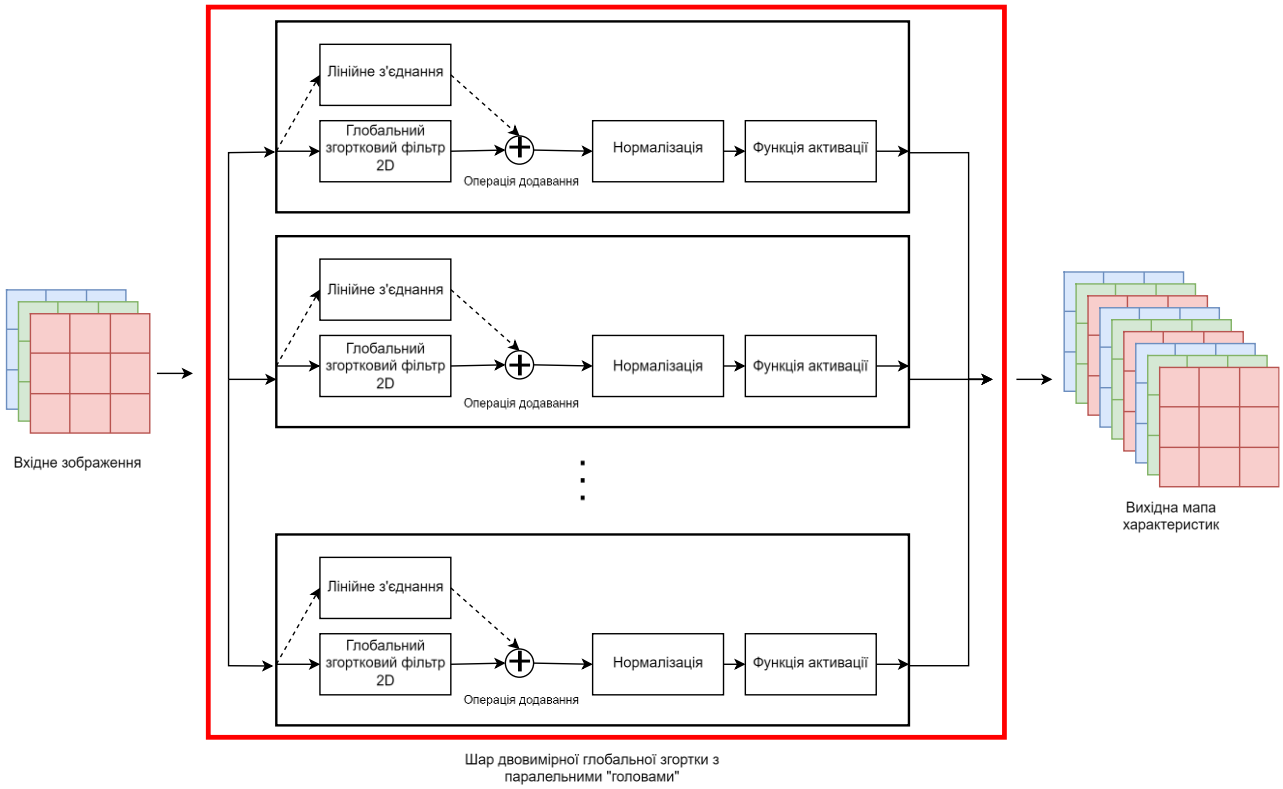

Рисунок 3. Архітектура шару двовимірної глобальної згортки з підтримкою декількох «голів»

На даному рисунку можна побачити, що завдяки декільком паралельним «голів», вихідна мапа характеристик має більше каналів, ніж вхідне зображення. Це можна використати, наприклад, у самому першому шарі глобальної згортки: оскільки вхідне зображення матиме лише 1 або 3 канали, корисно буде використати декілька паралельних голів і збільшити кількість каналів.

ЕКСПЕРИМЕНТИ ТА РЕЗУЛЬТАТИ

Проміжне тестування на MNIST. Мета першого експерименту – встановити, що шар двовимірної згортки реалізовано коректно і модель, що використовує лише запропоновані шари двовимірної згортки, здатна навчитись вирішувати найпростішу задачу по класифікації зображень на датасеті MNIST [10].

Опишемо детальніше протокол експерименту. Спершу, особливості використання датасету:

- Взято датасет MNIST. Тестова вибірка розміром 10 тисяч зображень використовується лише для тестування і не зазнала жодних змін.
- Тренувальну вибірку розміром 50 тис. зображень стратифіковано розбито на навчальну і валідаційну у пропорції 4:1 по класу зображення.
- Внаслідок стратифікованого розбиття навчальна та валідаційна вибірки є збалансованими.

Тепер перейдемо до деталей процедури тренування моделей:

- Для навчання моделі на вхід подавались зображення з навчальної вибірки групами по 64 зображення за ітерацію.
- Після однієї навчальної епохи, валідаційна метрика використовувалась для обчислення валідаційних метрик.
- Було встановлено ліміт в 15 навчальних епох. Критерій ранньої зупинки навчання – відсутність покращень у валідаційних метриках протягом 3 епох.
- Для усіх моделей використовується оптимізатор Adam з параметром інтенсивності навчання - $1e-03$.
- Після завершення процесу навчання розраховується метрика точності на тестовій вибірці.

Для тестування шару двовимірної глобальної згортки було створено наступну модель:

- Перший шар глобальної згортки має 256 паралельних «голів», що дозволяє отримати 256 каналів у вихідних мапах характеристик.
- Послідовність з 5 шарів глобальної згортки.

- Шар, що рахує середнє значення кожного каналу вздовж координат зображення.
- Лінійний шар, що отримує 256 параметрів на вхід та виводить 10 значень, які відображають ймовірність належності вхідного зображення до одного з десяти можливих класів.
- Тип нормалізації у шарах глобальної згортки – Batch Normalization, тип функції активації – GELU.

Отже, загальна модель складається з 6 шарів глобальної згортки та одного лінійного шару. Кількість параметрів у моделі – 50186.

Також, для порівняння було натреновано модель ResNet-50. Отриманні результати представлено у Таблиці 1:

Таблиця 1. Результати на датасеті MNIST

Модель	**Тестова точність**	**Кількість епох**	**Середній час навчання на одній епосі**
Глобальна згортка	98.81%	14	4 хвилини
ResNet-50	98.62%	7	45 секунд

За отриманими результатами можна побачити, що досліджувана модель отримала дещо вищу точність на тестовій вибірці. Це свідчить, що модель здатна навчатись і дозволяє проводити подальші експерименти. Втім, до таблиці також було додано інформацію про час навчання моделі. Проаналізуємо його.

Кількість епох, яку витратила модель на навчання, можна обґрунтувати відмінностями архітектур ResNet і глобальної згортки. Хоча моделі глобальної згортки і знадобилось вдвічі більше епох для навчання, ніж іншій, дане число все одно допустиме в рамках протоколу. Втім, проблемою є середній час, витрачений на одну епоху. Не дивлячись на те, що модель ResNet-50 має більше параметрів і шарів, модель з шарами глобальної згортки вимагала набагато більше часу на одну епоху. Причиною цього може бути необхідність інтерполяції

значень фільтру після кожного оновлення. Ця проблема буде більш детально розглянута у підпункті, присвяченому часовим затратам моделі.

Не зважаючи на явні проблеми з часом навчання, модель з використанням глобальної згортки змогла отримати кращі результати, ніж ResNet-50 на тестовій вибірці датасету MNIST. Протокол навчання, дані та їх розбиття були абсолютно ідентичними для обох моделей.

Швидкість роботи глобальної згортки. У даному пункті адресуємо питання часу згортки даного фільтру. У першій версії моделі при її навчанні на датасеті MNIST, модель з конфігурацією згаданою у попередньому підпункті вимагала 12 хв. 10 с. на одну навчальну епоху.

Як видно, остаточна версія моделі з першого експерименту досягла втричі вищої швидкості проходження однієї епохи. Це було досягнуто завдяки використанню швидких перетворень Фур'є та теоремі згортки, яка встановлює, що перетворення Фур'є згортки двох сигналів дорівнює поелементному добутку перетворень Фур'є цих сигналів. Правильно реалізувавши логіку такої згортки, вдалось отримати аналогічні результати з великим приростом у швидкості.

Втім, модель досі вимагає багато часу на обробку даних, тому було проведено тестування на вхідному зображенні розмірністю [224, 224, 3], одним шаром глобальної згортки з фільтром аналогічного розміру і 85 паралельними «головами» в шарі. За результати даного тестування вдалось становити відношення часових витрат кожної операції до загального часу обробки вхідних даних. З ними можна ознайомитись у Таблиці 2:

Таблиця 2. Відносний час операцій у згортковому шарі

Операція	**Відносний час, %**
Побудова фільтру	1.5 %
Нормалізація	1.4 %
Швидке перетворення Фур'є зображення і фільтру	15.6 %

Множення у частотному просторі	11.1 %
Обернене перетворення Фур'є	64.8 %
Інше	5.6 %

Як виявилось, необхідність інтерполяції параметрів не є проблемою, оскільки побудова фільтру (об'єднання послідовності під-фільтрів після їх інтерполяції) займає лише 1.5% від загального часу обробки даних і дане припущення було хибних. Найбільшою проблемою є час розрахунку результату згортки, який сумарно займає 91.5% часу. Ймовірно, що процес зворотнього розповсюдження похибки також не є швидким.

З цього можна зробити два висновки:

- Швидкі перетворення Фур'є дозволяють значно підвищити швидкість роботи нейронної мережі при роботі зі згортковими фільтрами великої розмірності.
- Навіть з цим приростом швидкості, використання великих згорткових фільтрів призводить до кардинального збільшення часу навчання нейронної мережі.

CIFAR-10 [11]**.** Наступним експериментом після MNIST було навчання на датасеті CIFAR-10. Головною відмінністю від попереднього датасету, в рамках даного дослідження, є поява 3 кольорових каналів у вхідному зображенні. Це є важливим аспектом для запропонованої моделі шару, оскільки він базується на згортці по глибині (*Depthwise convolution* [6]), яка дозволяє значно зменшити кількість параметрів шару, втім може вплинути на здатність моделі до навчання.

Протокол експерименту аналогічний першому за виключенням наступних пунктів:

- До тренувальних даних було застосовано аугментацію AugMix [12]
- Оптимізатор було замінено на AdamW [13] з параметром затухання вагів 0.03
- Максимальна кількість епох замінена на 100, а критерій ранньої зупинки збільшено до 10 епох без покращень.

- Коефіцієнт затухання з оригінальної формули (1) було зроблено параметром, що підлягає оптимізації, початкове значення $\frac{1}{2}$, оптимізується окремо для кожного каналу.
- Було вимкнено паралельне лінійне з'єднання між вхідними даними і результатом згортки.
- Було додано параметр зсуву, що додається до кожного каналу окремо після розрахунку згортки.

Дана конфігурація була встановлена емпіричним шляхом і показала найкращі результати, у порівняні з іншими перевіреними налаштуваннями шару глобальної згортки.

Для порівняння результатів також було натреновано модель ResNet-18. Результати представлені у Таблиці 3:

Таблиця 3. Результати на датасеті CIFAR-10

Модель	**Тестова точність**
Глобальна згортка	69.75%
ResNet-18	75.41%

Бачимо, що результати найкращої конфігурації моделі з шарами глобальної згортки не змогли досягнути результатів кількарічної моделі ResNet. Спробуємо проаналізувати, чому на даній задачі глобальна згортка показала набагато гірші результати.

Можна зробити два припущення, що є взаємопов'язаними. По-перше, модель, що базується на глобальній згортці є набагато меншою за кількістю параметрів, ніж ResNet-18 (64 тис. проти 11.1 мільйонів), і модель глобальної згортки виявилась замалою для більш успішної роботи на даному датасеті. Друга причина – використання згортки по глибині, дана техніка дозволяє значно зменшити кількість параметрів у моделі, але призводить до того, що інформація між різними каналами не змішується. Можливо, шар глобальної згортки потребує використання звичайної згортки, що дозволить обмінювати

інформацію між каналами, а також призведе до збільшення параметрів, що може дозволити збільшити обсяг інформації, яку здатен «вивчити» один шар.

ОБГОВОРЕННЯ

Отже, використовуючи запропоновану архітектуру шару двовимірної глобальної згортки було проведено тестування на двох датасетах: одноканальному (відтінки сірого) датасеті MNIST та трьохканальному (RGB) датасеті CIFAR-10. За результатами експериментів було встановлено, що глобальна згортка показала кращі результати у задачі з одним кольоровим каналом і мала значно нижчі результати за конкурента при роботі з трьома каналами.

Основне припущення, що випливає з даних результатів – даний шар потребує додаткових експериментів з використанням «звичайної» згортки, коли інформація між каналами обмінюється. З одного боку це може призвести до значного збільшення кількості параметрів, з іншого – шар матиме здатність краще аналізувати вхідні дані.

Тому подальшими кроками є імплементація звичайної згортки у запропонованій моделі шару. Ця імплементація вимагатиме додаткового тестування, але дасть можливість виконати дуже важливий експеримент: використовуючи існуючу архітектуру (ResNet, MobileNet, тощо) замінити у ній усі фільтри локальної згортки на глобальну. На даному етапі проведення такого експерименту є проблематичним через неможливість вільно налаштовувати кількість вихідних каналів у згортці по глибині. Ця зміна може потенційно покращити результати на датасеті CIFAR-10.

Також не варто виключати можливість використання неоптимальної конфігурації у експериментах. Необхідно провести більше експериментів з існуючою моделлю і переконатись чи є точність у 69.75% максимально можливою. Але головною проблемою є час навчання моделі і, на жаль, як було встановлено, причиною цього є розмір шару згорткового фільтру, а не особливість його побудови, що ускладнює процес дослідження.

ВИСНОВКИ

У даній роботі було досліджено архітектуру одновимірного згорткового шару, суть якого полягає у застосуванні глобальної згортки, що охоплює увесь вхідний контекст, а не лише його локальну область. Одновимірна глобальна згортка має високі результати у одновимірних задачах обробки сигналів, втім наразі не існує її імплементації для багатовимірних задач. Тому метою даної роботи є пошук моделі шару, який дозволить використати глобальну згортку у двовимірних задачах.

Отриманим результатом є модель шару двовимірної глобальної згортки, яка, базуючись на тих самих принципах, що і модель SGConv, дозволяє використовувати глобальну згортку у двовимірних задачах, таких як класифікація зображень.

Дана модель може бути використана при роботі з більшістю задач області комп'ютерного зору таких як: детекція об'єктів, сегментація об'єктів, тощо, оскільки представляє з себе такий самий згортковий шар, як і типова локальна згортка.

Втім, наразі дана модель потребує додаткових досліджень. Використовуючи нейронну мережу з шарами двовимірної глобальної згортки, вдалось отримати відносно високі результати при класифікації зображень з датасету MNIST, але на багатоканальному датасеті CIFAR-10 результати були значно нижчими у порівнянні з іншою нейронною мережею. Тому необхідно дослідити інші можливі конфігурації нейронної мережі з шарами глобальної згортки, а також модель шару глобальної згортки без використання згортки по глибині.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Mar. 17, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[2] A. Dosovitskiy *et al.*, “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.” arXiv, Jun. 03, 2021. doi: 10.48550/arXiv.2010.11929.

[3] P. Warden, “Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition.” arXiv, Apr. 09, 2018. doi: 10.48550/arXiv.1804.03209.

[4] A. Gu, K. Goel, and C. Re, “Efficiently Modeling Long Sequences with Structured State Spaces,” in *International Conference on Learning Representations*, 2022. [Online]. Available: https://openreview.net/forum?id=uYLFoz1vlAC

[5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[6] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[7] Y. Li, T. Cai, Y. Zhang, D. Chen, and D. Dey, “What Makes Convolutional Models Great on Long Sequence Modeling?,” presented at the International Conference on Learning Representations, 2023. [Online]. Available: https://openreview.net/pdf?id=TGJSPbRpJX-

[8] Третиник В.В., Шуліка О.О., Архітектура двовимірного згорткового шару нейронної мережі на основі глобальної згортки. Прикладна математика та комп'ютинг. ПМК-2023 : шістнадцята наук. конф. магістрантів та аспірантів, Київ, 28—30 лист. 2023 р. : зб. тез доп. / [редкол.: Дичка І. А. та ін.]. — К. : Просвіта, 2023. — 680 с.: 157-161 ISBN 978-617-7010-28-8

[9] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, Eds., in Proceedings of Machine Learning Research, vol. 37. Lille, France: PMLR, Jul. 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html

[10] Li Deng, “The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web],” *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012, doi: 10.1109/MSP.2012.2211477.

[11] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images,” *Univ. Tor.*, May 2012.

[12] D. Hendrycks*, N. Mu*, E. D. Cubuk, B. Zoph, J. Gilmer, and B. Lakshminarayanan, “AugMix: A Simple Method to Improve Robustness and Uncertainty under Data Shift,” in *International Conference on Learning*

Representations, 2020. [Online]. Available: https://openreview.net/forum?id=S1gmrxHFvB

[13] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *International Conference on Learning Representations*, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7