

Зінченко О.В., Звенігородський О.С., Кисіль Т.М.
Державний університет телекомунікацій, Київ

ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ ДЛЯ ВИРІШЕННЯ ЗАДАЧ КОМП'ЮТЕРНОГО ЗОРУ

***Анотація.** В даній статті зроблено огляд основних методів вирішення задач комп'ютерного зору класифікації, сегментації та обробки зображень, які реалізовані в CV-системах. Системи комп'ютерного зору запрограмовані на виконання вузькоспеціалізованих задач, що здатні виявляти об'єкти при ідентифікації, зчитувати серійні номери, виконувати пошук поверхневих дефектів. При застосуванні методів глибокого навчання в CV-системах значно підвищується їх швидкість обробки на дата-сетах великих розмірів та точність класифікації/сегментації зображень. Системи штучного зору спроможні визначати окремі пікселі за відповідними ознаками при обробці, надають якісний результат по розпізнаванню образів, відновлення зображень, вписування частини зображення. Хоча деякі алгоритми комп'ютерного зору були розроблені для імітації зорове сприйняття, та більша кількість запропонованих методів здатні повноцінно обробляти зображення і визначення на них характерних властивостей.*

Область застосування CV-систем надалі буде розширюватися, оскільки потреба в системах штучного інтелекту стрімко зростає. Метою даної статті — наведення структурованого огляд технологій комп'ютерного зору за їх перевагами та недоліками. В роботі узагальнено різновиди CV-систем зі штучним інтелектом за спектром їх застосувань, виділено основні проблематичні напрямки їх дослідження, такі як розпізнавання, ідентифікація та виявлення.

В статті проведено огляд згорткових нейронних мережі (CNN), які успішно застосовуються до аналізу візуальних зображень в глибокому навчанні. Архітектури CNN у деяких випадках перевершують результативність штучних нейронних мереж в задачах класифікації за своєю продуктивністю. На даний час, згорткові нейронні мережі – основний інструмент для класифікації та розпізнавання об'єктів, облич на фотографіях, розпізнавання відео- та аудіо- матеріалів. В даній роботі наведено порівняльний аналіз відомих CNN-моделей: LeNet 5, AlexNet, VGGNet, GoogLeNet, ResNet та ефективність в CV-системах. Запропоновано підходи до моделювання архітектур згорткових нейронних мереж, які дозволять, в подальшому, вирішити проблему класифікації в задачах для комп'ютерного зору, тим самим підвищуючи їх продуктивність, точність та якість обробки.

***Ключові слова:** штучний інтелект, комп'ютерний зір, задача класифікації, глибоке навчання, нейронні мережі, згорткові нейронні мережі, розпізнавання образів.*

Zinchenko O.V., Zvenigorodsky O.S., Kysil T.M.
State University of Telecommunications, Kyiv

CONVOLUTIONAL NEURAL NETWORKS FOR SOLVING COMPUTER VISION PROBLEMS

***Abstract.** This article provides an overview of the main methods of solving computer vision problems of classification, segmentation and image processing, which are implemented in CV systems. Computer vision systems are programmed to perform highly specialized tasks, capable of detecting objects during identification, reading serial numbers, and searching for surface defects. When applying deep learning methods in CV systems, their processing speed on large data sets and the accuracy of image classification/segmentation are significantly increased. Artificial vision systems are able to identify individual pixels according to the relevant features during processing, provide a high-quality result in pattern recognition, image restoration, and fitting part of the image.*

© Зінченко О.В., Звенігородський О.С., Кисіль Т.М.

2022

Although some computer vision algorithms were developed to simulate visual perception, a larger number of proposed methods are able to fully process images and determine their characteristic properties.

The scope of application of CV systems will continue to expand, as the need for artificial intelligence systems is growing rapidly. The purpose of this article is to provide a structured review of computer vision technologies based on their advantages and disadvantages. The work summarizes the types of CV-systems with artificial intelligence according to the spectrum of their applications, highlights the main problematic areas of their research, such as recognition, identification and detection.

The article reviews convolutional neural networks (CNNs), which are successfully applied to the analysis of visual images in deep learning. CNN architectures in some cases outperform artificial neural networks in classification tasks by their performance. Currently, convolutional neural networks are the main tool for classification and recognition of objects, faces in photographs, recognition of video and audio materials. This paper provides a comparative analysis of well-known CNN models: LeNet 5, AlexNet, VGGNet, GoogLeNet, ResNet and their effectiveness in CV systems. Approaches to the modeling of architectures of convolutional neural networks are proposed, which will allow, in the future, to solve the problem of classification in tasks for computer vision, thereby increasing their performance, accuracy and quality of processing.

Keywords: artificial intelligence, computer vision, classification task, deep learning, neural networks, convolutional neural networks, pattern recognition.

1. Постановка проблеми

Комп'ютерний зір (Computer Vision, CV) – галузь штучного інтелекту, пов'язана з аналізом, класифікацією і розпізнаванням зображень, відео та аудіо даних. В основі CV-систем зазвичай реалізовані алгоритми на базі машинного навчання *Deep Learning*, за їх допомогою вдається розпізнавати одні об'єкти від інших, відрізняти патерни та закономірності. Завдяки методам *Deep Learning*, CV-системи імітують властивості людського зору, розпізнають образи на зображеннях, розмічають інформацію на зображеннях, аудіо та відеоданих. Комп'ютерний зір – наукова галузь яка об'єднує методики обробки, аналізу та розпізнавання зображень, а також великих масивів даних (послідовність зображень, зображення з камер, тривимірні дані тощо). Як окрема галузь науки комп'ютерний зір займається вивченням методів і технологій для створення багатофункціональних CV-систем із штучним інтелектом (рис. 1), які здатні оптимізувати, класифікувати, кластеризувати задані об'єкти в процесі автоматичного планування.

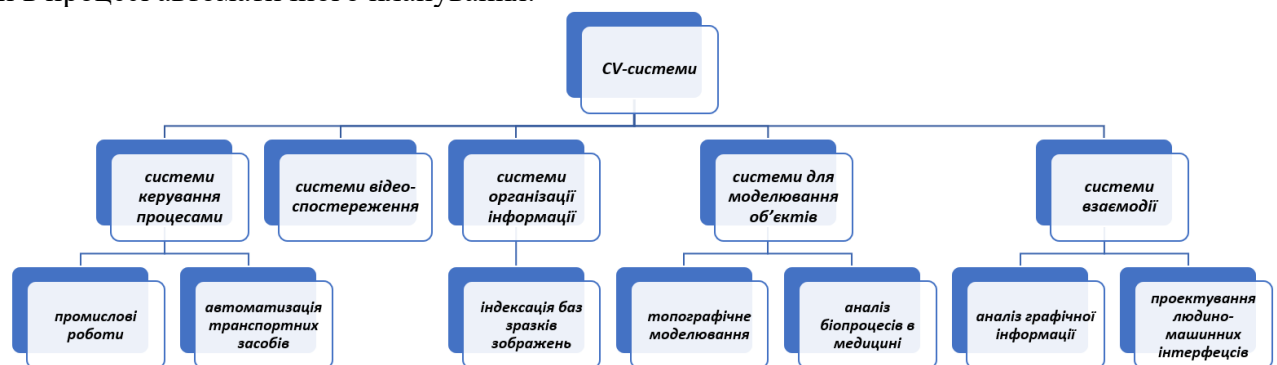


Рис. 1. Різновиди відомих CV-систем [9]

Спектр застосувань даної технології дедалі розширюється, а в напрямку комп'ютерного досягнуто значного прогресу, все ж залишається багато невирішених задач при дослідженнях. Насамперед, існуючим алгоритмам бракує загальності, а збільшення швидкодії обробки даних призводить до зниження якості. Тому актуальними напрямками досліджень є: підвищення швидкості обробки існуючих алгоритмів; застосування алгоритмів розпізнавання в уже реалізованих автоматичних системах; вирішення задач в режимі реального часу в умовах

обмежених ресурсів; розпізнавання зображень нейронними мережами з підвищеною точністю, якістю та розширеним спектром.

При дослідженні комп'ютерного зору, обробки зображень і машинного зору можна виділити три основні проблематичні напрямки:

- *Розпізнавання*: один/декілька попередньо заданих об'єктів/класів можуть бути розпізнані разом із їх двовимірним положенням на зображенні/тривимірним положенням на сцені.
- *Ідентифікація*: розпізнавання та порівняння індивідуальних біометричних даних об'єкта (колір очей, графологія сітчатки ока, відбитки пальців, геометрія руки, риси обличчя, тощо).
- *Виявлення*: відеодані перевіряються на наявність визначених умов (виявлення можливих змін клітин/тканин та їх властивостей при медичних обстеженнях, перешкод в системах керування робототехнікою, тощо) для отримання правильної інтерпретації.

Для вирішення задач комп'ютерного зору використовуються алгоритми та методи, серед яких виділяють наступні напрямки: застосування нечіткої логіки [11]; нейронні мережі, засновані на гнучкому порівнянні графів; прихованих Марковських моделях; перетвореннях Кархунена-Лоева для аналізу МРТ-зображень; згорткові нейронні мережі CNN (Convolutional Neural Networks) засновані на алгоритмах класифікації зображень Deep Learning [4] і т. д. В даній статті буде проаналізовано архітектури згорткових нейронних мереж та їх застосування в задачах комп'ютерного зору.

2. Аналіз останніх досліджень і публікацій.

Розробкою та дослідженням методів та моделей комп'ютерного зору займаються зарубіжні [1-9] та вітчизняні [10,11] науковці, які застосовують методи та моделі глибокого навчання. Науковцями започатковано напрямок Deep Learning, що базується на реалізації нейронних мереж, так званий (Deep Neural Network — DNN).

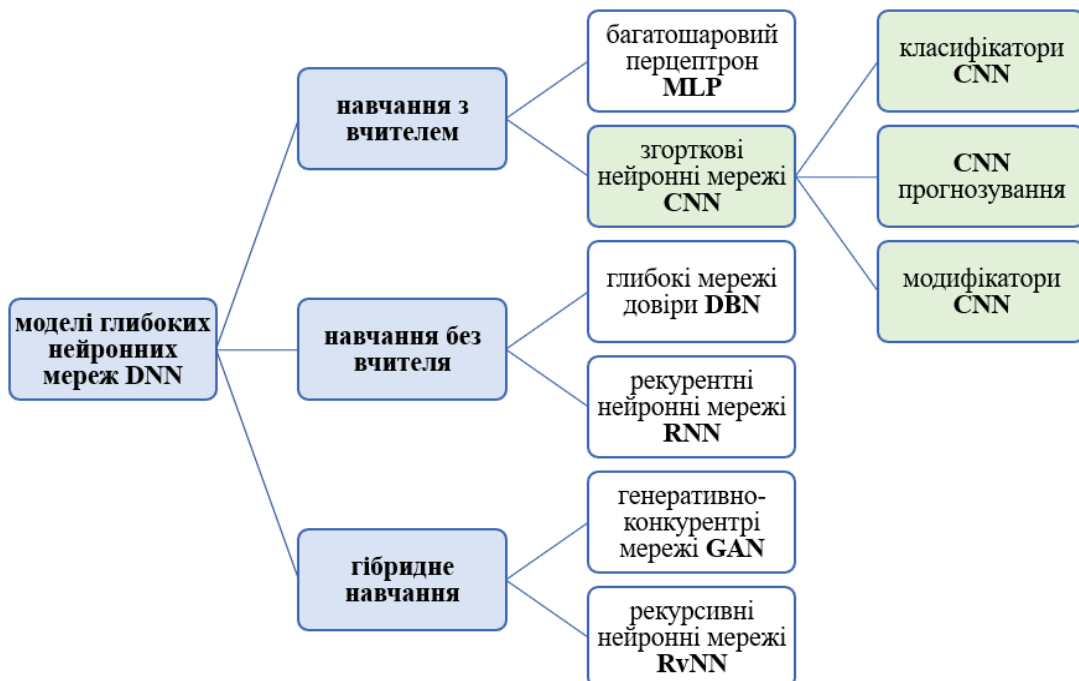


Рис. 2. Класифікація моделей нейронних глибоких мереж DNN для вирішення задач комп'ютерного зору [11]

Сутність навчання DNN полягає в підборі оптимального методу, що заснований на коректних математичних перетвореннях при обробці вхідних даних і отримання вихідного

результату, незалежно від лінійної/нелінійної кореляції. Навчання може проходити як з вчителем (supervised learning), так і без (unsupervised learning), а також гібридне, яке є поєднанням попередніх двох методів. Узагальнену класифікацію моделей DNN за способом навчання наведено на рис. 2.

Порівнювати між собою різні моделі глибоких нейронних мереж, які досліджуються науковцями, доволі проблематично, але, при моделюванні, важливо оцінювати ефективність застосування моделей на різних наборах даних або підмножини конкретного набору. Останнім часом, для вирішення задач комп'ютерного зору, широко використовуються такі моделі глибокого навчання, які прийнято називати *згортковими нейронними мережами*. Найбільш успішні у застосування моделі, які використовуються для виявлення, класифікації та аналізу зображень, є: *AlexNet, ResNets, EfficientNets, YOLO, R-CNN, LambdaNetworks, VGG*.

3. Мета і задачі дослідження.

В даному дослідженні необхідно провести порівняльний аналіз моделей згорткових нейронних мереж та їх характеристик, які суттєво впливають на реалізацію типових задач комп'ютерного зору. Перед нами поставлена мета у виборі топології згорткових нейронних мереж, в яких, при дослідженнях, виділяють наступні етапи:

- визначення задачі, яка вирішується *CNN* (класифікація/прогнозування/модифікація);
- виявлення обмежень у реалізованій задачі (швидкість/точність навчання);
- вибір вхідних даних (типові: зображення, звук; розмір: 100x100, 40x40, 30x30; формат: *RGB*, в градаціях сірого *Grayscale*) та вихідних даних (кількість реалізованих класів);
- визначення архітектури згорткової мережі (кількість згорткових, підвибіркових, пов'язаних прошарів; кількість карт ознак та властивостей, розмір ядер, моделі навчання за відповідними функціями активації та функціями втрат).

Основною задачею дослідження являється огляд методів класифікації на основі відомих згорткових нейронних мереж, які орієнтовані на обробку та розпізнавання образів зображень, сегментації текстових даних, а також звукових потоків.

4. Результати дослідження.

Найкращі результати в задачах класифікації показали *Convolutional Neural Network*, які постали логічним розвитком після запронованих топологій когнітрона та неокогнітрона Фукушіми. Згорткові нейронні мережі (ЗНМ) забезпечують часткову стійкість до змін масштабу, зсувів, поворотів, зміни ракурсу та інших зміщень. На даний момент згорткові нейронні мережа та її модифікації, на думку вчених, вважаються найкращими за точністю та швидкістю навчання за алгоритмами знаходження об'єктів у просторі.

В своїй основі, згорткові нейронні мережі поєднують три архітектурні ідеї, при забезпечення інваріантності: зміни до масштабу, поворот зсуву та просторові зміщення. Завдяки такій архітектурі, *CNN* забезпечують:

- локальні рецепторні поля (локальний двовимірний взаємозв'язок нейронів);
- загальні синаптичні коефіцієнти (детектування певних рис в будь-якому просторовому зображенні, зменшуючи, при цьому, загальну кількість вагових коефіцієнтів);
- ієрархію в організації просторових підмножин на вибіркових даних.

ЗНМ містять різних види прошарків: згорткові (convolutional) прошарки, субдискретизуючі (subsampling, підмножин) прошарки та прошарки штучної нейронної мережі (персептрону), які наведено на рис. 3.

Перші два типи прошарків (convolutional, subsampling), по чергово змінюються, формуючи вхідний вектор ознак на повнозв'язаних прошарках (fully-connected layer, F-layers)

на виході [6, с.] в нейронній мережі. В ЗНМ всі види прошарків можуть в довільному порядку по чергово змінюватись. Реалізована в CNN операція згортки використовується для обробки зображень і може бути описана наступною формулою [11]:

$$C_{i,j} = \sum_{u=0}^{m_x-1} \sum_{v=0}^{m_y-1} A_{i+u} B_{u,v} \quad (1)$$

де $C_{i,j}$ – розраховане значення пікселя на зображенні, яке обробляється;

$B_{u,v}$ – значення елемента ядра згортки (u, v);

A_{i+u} – значення пікселя вхідного зображення, в якому m_x-1, m_y-1 – розмір ядра згортки.

Згортка (*convolution*) виконується над парою матриць A з розмірністю $n_x \times n_y$ та B з розмірністю $m_x \times m_y$. Тоді кожен піксель обчислюється як скалярний добуток матриці B і деякої підмножини A та формується результативна матриця $C = A \cdot B$ розмірністю $(n_x - m_x + 1) \cdot (n_y - m_y + 1)$ [11].

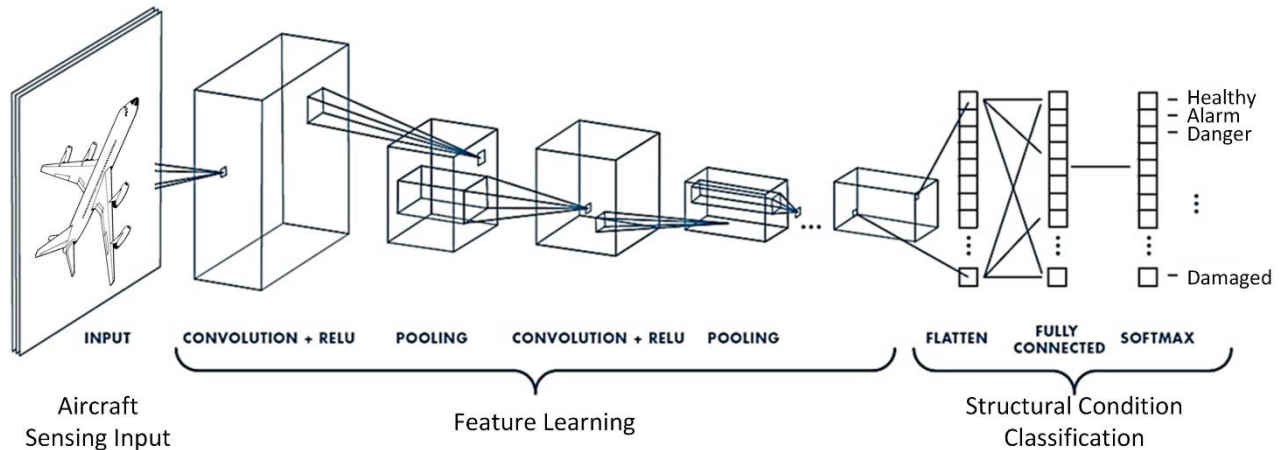


Рис. 3. Загальна архітектура згорткової нейронної мережі CNN [9]

Алгоритм функціонування для операції згортки можна описати наступним чином: на карті з розмірністю ядра B переглядається із заданим кроком вхідне зображення A ; на кожному кроці поелементно підсумовується результат і формується в матрицю C (рис. 4). В залежності від методу обробки, результат вихідної матриці може бути меншим (*valid*)/більшим (*full*) від вихідного зображення, або може співпадати розмір (*same*) з розпізнаним зображенням.

У структурі згорткової нейронної мережі основними видами прошарків являються:

- згорткові прошарки (*convolutional layer*);
- пулінгові прошарки (*pooling layer*);
- повнозв'язані прошарки (*fully-connected layer*).

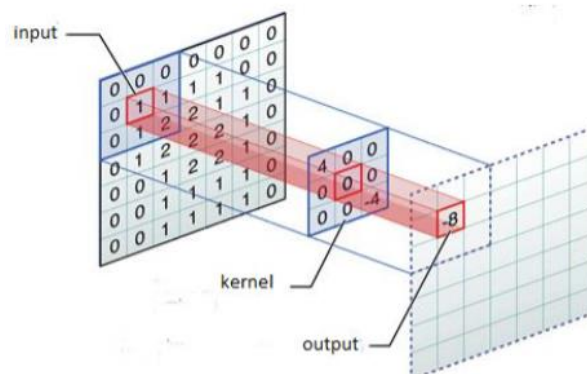


Рис. 4. Операція згортки та отримання значень згорткової карти [8]

У згортковому прошарку нейронної мережі застосовується операція згортки до виходів з попереднього шару, де вагові коефіцієнти ядра (пула) згортки являються навчальними параметрами. Додатково використовується ваговий коефіцієнт, як константний *bias*-зсув. При модифікації нейронної мережі, важливо враховувати наступні параметри:

1. В одному згортковому прошарку допускається декілька згорток. Наприклад, якщо на вході подано зображення розмірністю $w \times h$, а в прошарку задано n згорток з пулом розмірності $k_x \times k_y$, то на виході буде сформоване зображення розмірністю $n \cdot (w - k_x + 1) \cdot (h - k_y + 1)$.

2. Пули згортки можуть бути тривимірними. Наприклад, для узагальнення інформації R-, G-, B-каналів вихідного зображення, на першому прошарку можна використовувати згортку розмірністю $d \times w \times h$, тоді на виході даного прошарку буде сформоване одне зображення, замість трьох.

3. Формування доповненого зображення. Застосування згортки приводить до зменшення зображення, також пікселі, які знаходяться на межі зображення, зменшують кількість та розмірність згорток. В зв'язку з цим у згорткових прошарках застосовують доповнене зображення (*padding*). Виходи з попереднього прошарку доповнюються пікселями так, щоб після згортки зберігся початковий розмір зображення. Доповнені згортки прийнято називають *однаковими* (*same convolution*), згортки без доповнення - *правильними* (*valid convolution*).

4. Зсув (*stride*) згорткового прошарку. Зазвичай згортка застосовується для кожного пікселя, іноді використовується зсув, що $\neq 1$, тоді скалярний добуток враховується не з усіма можливими положеннями ядра, а тільки з тими положеннями, які кратні деякому зсуву s . Наприклад, якщо вхідне зображення з розмірністю $w \times h$, ядром згортки розмірністю $k_x \times k_y$ із застосуванням використовується зсув s , то вихідне зображення буде мати розмірність, яка визначається як:

$$\left\lfloor \frac{w - k_x}{s} \right\rfloor + 1 \cdot \left\lfloor \frac{h - k_y}{s} \right\rfloor + 1 \quad (2)$$

Пулінговий прошарок наділений властивістю знижувати розмірність зображення. Вихідне зображення ділиться на блоки розміром $w \times h$, кожен блок обчислюється за деякою функцією. Найчастіше в моделях використовується *функція максимуму* (*max pooling*) або (*зваженого середнього*) (*weighted average pooling*). Застосування пулінгового прошарку надає можливість:

- зменшення зображення наступних згорток, для того щоб з підвищеною точністю розпізнавати більші області вихідного зображення;
- збільшення інваріантності на виходах нейронної мережі по відношенню до поданих вхідних даних та їх властивостей;
- підвищення швидкості обробки даних.

Повнозв'язаний прошарок (*Insertion module*) нейронної мережі, запропонований у праці науковців [2], який використовується в мережі *GoogLeNet*. Автори запропонували, що кожному елементу попереднього прошарку відповідає певна ділянка вихідного зображення. Кожна згортка за такими елементами збільшує область вихідного зображення, до тих пір поки елементи на останніх прошарках не будуть відповідати повністю вхідному зображенню. Але, якщо, при обробці, згортки стануть розміром 1×1 , то не буде знайдено елементів за ознаками на вихідному зображенні. Для вирішення цієї проблеми, автори реалізували прошарок *insertion module*. Наведена конфігурація згорткової мережі призводить до різкого збільшення прошарків, що не дозволяє побудувати модель глибокої нейронної мережі. Для цього автори запропонували використання модифікованого *insertion module* прошарку з додатковим зменшенням розмірності: до кожного фільтру на зображенні додається прошарок згортки 1×1 , який перетворює всі згорткові прошарки в один прошарок. Даний метод дозволяє зберегти всі властивості при розпізнаванні зображень на незначній кількості згорткових прошарків.

В різних наукових працях зустрічаються інші види згорток в архітектурах згорткових нейронних мережах, а саме:

- *Розширена згортка* (*Dilated convolution*) завдяки якій експонентно розширюється рецептивне поле без втрати якості зображення та зберігаються витратні ресурси мережі.

- *Часткова згортка (Partial convolution)* дозволяє опрацювати вхідне зображення з бінарною маскою – додатковими ознаками при розпізнаванні. Наприклад, якщо маска вказує на викривленні пікселі, то постає завдання часткового вписування/відновлення зображення у заданих областях.

- *Стрбована згортка (Gated convolution)* дозволяє зберегти додаткові ознаки вхідного зображення на всіх прошарках згортки за відповідною маскою. У цьому випадку замість того, щоб працювати з жорсткою маскою, яка оновлюється за деякими правилами, стрбована згортка вчиться автоматично виявляє маску з заданих даних. Дана згортка навчена динамічному відбору ознак на зображенні та в кожній логічній області маски, що значно покращуючи якість розпізнаних вихідних даних.

Глибокі згорткові нейронні мережі перевершили за якістю рівень класифікації зображень. CNN отримують низько-, середньо- і високорівневі ознаки наскрізним багатошаровим способом, а збільшення кількості *stacked layers* підвищують рівні ознак, що значно впливає на якість та точність розпізнавання. Перерахуємо найвідоміші моделі згорткових нейронних мереж та охарактеризуємо властивості їх архітектур:

- LeNet 5 [5] мережа складається з п'яти прошарків, серед яких три згорткові прошарки (*карти ознак feature map*) розміром 6, 16 і 120) та два повнозв'язних прошарки. Кожен прошарок згортки має розмір пула 5x5 з кроком 1. Згорткові прошарки активуються функцією гіперболічного тангенса, прошарок середнього пулінга має розмір ядра 2x2. Вихідний прошарок останньої згортки має розмір (1, 1, 120) та передається в подальшому до повнозв'язного з'єднання. LeNet 5 досягає 5% точності валідації та досягає точності навчання 91,5% при застосованій категоріальній крос-ентропічній функції втрат.

- AlexNet [1] складається з двох частин, які слабо взаємодіють між собою, що дозволяє використовувати їх паралельно для різних GPU та мінімальним обміном даними. Нейронна мережа має 60 млн. параметрів та 650 тис. нейронів, складається з п'яти згорткових шарів, прошарків згортки з максимальними пулами, трьох пов'язаних прошарків з використанням функції активації *softmax*; для прискорення навчання, використана додаткова функція активації згорткового прошарку; для зменшення перенавчання у повнозв'язкових прошарках застосовано метод *регуляризації*.

- VGGNet [3] архітектури нейронних мереж VGG-13, VGG-16 і VGG-19, підтримують 13/16/19 прошарків і може класифікувати зображення за 1000 категоріями об'єктів. Перевагою даної мережі - є простота топології мережі. Декілька прошарків згортки, об'єднані в єдиний прошарок, що зменшує висоту і ширину зображення; в мережі доступні для використання 64 фільтри, які можна збільшити до 128/256; в останніх прошарках використовуються 512 фільтрів. Модель VGG16 навчалася з використанням графічних процесорів Nvidia Titan Black протягом кількох тижнів.

- GoogLeNet (*Inception Network*) Складається в основному з *Inception*-модулів та містить 22 прошарки параметрами, які автоматично налаштовуються (+5 пулінгових прошарки). Мережа GoogLeNet не має високу точність навчання, але модель її ефективна за розмірністю та кількістю обчислень. В мережі AlexNet реалізовано 9 *Inception*-блоків, завдяки яким зменшуються практично в 10 разів кількість вхідні параметрів, тим самим підвищується швидкість обчислень, так як зменшується розмірність в процесі обробки. В результаті виявилось, що мережа класифікує зображення найкраще в порівнянні з попередніми моделями мереж CNN.

- ResNet мережа містить понад 150 прошарків [4]. Кожен блок ResNet має два рівні глибини (ResNet 18/34) або три рівні глибини (ResNet 50/101/152). Для 50-шарової ResNet кожен 3-шаровий блок замінюється в 34-шаровою мережею, в які можуть використовуватись два варіанти збільшення розмірностей. Завдяки такій архітектурі, навіть після збільшення глибини 152-шарова ResNet (11,3 млрд. FLOPs) мережа має простішу структуру ніж мережі типу VGG-16/19 (15,3/19,6 млрд. FLOPs).

Як результат, наведемо основні параметри ефективності функціонування вказаних моделей, які представлені в табл. 1.

Порівняльна характеристика параметрів згорткових нейронних мереж.

Назва моделі	Час навчання (s)	Швидкість навчання (s)	Точність навчання (%)	Функція втрат (%)
<i>LeNet 5</i>	122.72	0.001	98.5	0.43
<i>AlexNet</i>	162.74	0.010	84.6	15.4
<i>VGGNet</i>	>500.00	0.900	92.77	7.33
<i>GoogLeNet</i>	102.5	0.001	93.30	6.70
<i>ResNet</i>	150.5	0.001	96.43	3.57

*Джерело наведено авторами

5. Висновки і перспективи подальших досліджень.

В результаті проведеного дослідження можна зробити висновок, що в глибоких нейронних мережах в процесі згортки виникає проблема: із збільшенням глибини мережі точність обробки на початку збільшується, а в подальшому значно погіршується. Проте, зниження точності навчання ускладнює процес оптимізації.

В процесі підвищення поточних характеристик згорткових нейронних мереж можна використовувати в процесі їх моделювання наступні методи: мінімізацію при зменшенні розмірності, нормалізацію батчів, збільшення та розширення вибірки, просунуті режими оптимізації, застосування різних типів функцій активації (*ReLU*, *SoftMax* і т.д.).

Крім того, необхідно враховувати, при моделюванні, архітектури згорткової нейронної мережі, які дозволять вирішити проблему класифікації в задачах для комп'ютерного зору. Для ефективного розв'язання тієї чи іншої задачі найкращим може бути поєднання декількох методів. В подальших дослідженнях планується вирішення наступних питань, а саме: застосування CV-систем в умовах обмежених ресурсів, підвищення швидкодії при збереженні задовільної точності в умовах реального часу.

Список використаних джерел

1. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks // Communications of the ACM, June 2017, pp.84–90, <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
2. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Going Deeper with Convolutions // Computer Vision Foundation, 2015, <https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf> 1-9 pages.
3. Gaudenz Boesch, VGG Very Deep Convolutional Networks (VGGNet) – What you need to know // Read more at: <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/> 2022.
4. Jon Krohn, Gnannd Beyleveld, Aglae Bassens // Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence, Pearson Education, Inc, 2020.
5. Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun, Deep Residual Learning for Image Recognition // <https://arxiv.org/pdf/1512.03385.pdf> 10 Dec 2015.
6. LeCun Y., Y. Bengio Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, 1995.
7. Linda G. Shapiro, George Stockman, Computer Vision // Pearson; 1st edition (February 2, 2001) 608 pages.

8. Théodore Bluche, São Paulo Deep Neural Networks – Applications in Handwriting Recognition Meetup - 9 Mar. 2017.
9. Zahra Elhamraoui, Introduction to convolutional neural network, <https://medium.com/analytics-vidhya/introduction-to-convolutional-neural-network-6942c189a723>, May 28, 2020.
10. Ковальчук А.М., Марчук Г.В., Марчук Д.К. Застосування згортової нейронної мережі для розпізнавання рукописних символів // зб. наук. пр. «Вчені записки ТНУ імені В.І. Вернадського». Серія: технічні науки, Том 30 (69) Ч. 1 № 4 2019, с. 68-73.
11. Тимчишин Р.М., Волков О.Є., Господарчук О.Ю., Богачук Ю.П., Сучасні підходи до розв'язання задач комп'ютерного зору // зб. наук. пр. «Control systems and computers», УСиМ, 2018, № 6, с. 46-73.

References:

1. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks // Communications of the ACM, June 2017, pp.84–90, <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
2. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich, Going Deeper with Convolutions // Computer Vision Foundation, 2015, <https://www.cs.unc.edu/~wliu/papers/GoogLeNet.pdf> 1-9 pages.
3. Gaudenz Boesch, VGG Very Deep Convolutional Networks (VGGNet) – What you need to know // Read more at: <https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/> 2022.
4. Jon Krohn, Gnanm Beyleveld, Aglae Bassens // Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence, Pearson Education, Inc, 2020.
5. Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun, Deep Residual Learning for Image Recognition // <https://arxiv.org/pdf/1512.03385.pdf> 10 Dec 2015.
6. LeCun Y., Y. Bengio Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M. A. (Eds), The Handbook of Brain Theory and Neural Networks, 1995.
7. Linda G. Shapiro, George Stockman, Computer Vision // Pearson; 1st edition (February 2, 2001) 608 pages.
8. Theodore Bluche, São Paulo Deep Neural Networks – Applications in Handwriting Recognition Meetup - 9 Mar. 2017.
9. Zahra Elhamraoui, Introduction to convolutional neural network, <https://medium.com/analytics-vidhya/introduction-to-convolutional-neural-network-6942c189a723>, May 28, 2020.
10. Kovalchuk A.M., Marchuk G.V., Marchuk D.K. Application of a convolutional neural network for recognition of handwritten symbols // coll. of science pr. "Scientific notes of TNU named after V.I. Vernadsky". Series: technical sciences, Volume 30 (69) Part 1 No. 4 2019, p. 68-73.
11. Tymchyshyn R.M., Volkov O.E., Gospodarchuk O.Yu., Bogachuk Y.P., Modern approaches to solving computer vision problems // coll. of science pr. "Control systems and computers", USyM, 2018, No. 6, p. 46-73.