Інформаційні технології

УДК 004.8

Яковчук Олег Костянтинович
бакалавр комп'ютерних наук
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
Яковчук Олег Константинович
бакалавр компьютерных наук
Национального технического университета Украины
«Киевский политехнический институт имени Игоря Сикорского»
Yakovchuk Oleh
Bachelor of Computer Science of the
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"

АНАЛІЗ ОСНОВНИХ ПРОБЛЕМ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ПРИ ВИРІШЕННІ ЗАДАЧІ З РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

АНАЛИЗ ОСНОВНЫХ ПРОБЛЕМ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ПРИ РЕШЕНИИ ЗАДАЧИ ПО РАСПОЗНАВАНИЮ ИЗОБРАЖЕНИЙ

ANALYSIS OF THE BASIC PROBLEMS OF CONVOLUTIONAL NEURAL NETWORKS IN IMAGE RECOGNITION PROBLEM SOLUTION

Анотація. *У роботі досліджено виявлені проблеми у сучасних згорткових нейронних мережах при розпізнаванні зображень. Проведено аналіз базових принципів та архітектурних складових згорткових мереж. Обґрунтовано необхідність створення нового типу роутингу між шарами*

мережі замість існуючого методу агрегування.

Ключові слова: *глибинне навчання, розпізнавання зображень, класифікація зображень, згорткова нейронна мережа, агрегувальний шар.*

Аннотация. *В работе исследованы выявленные проблемы в современных свёрточных нейронных сетях при распознавании изображений. Проведён анализ базовых принципов и архитектурных составляющих свёрточных сетей. Обоснована необходимость создания нового типа роутинга между слоями сети вместо существующего метода субдискретизации.*

Ключевые слова: *глубокое обучение, распознавание изображений, классификация изображений, свёрточная нейронная сеть, слой субдискретизации.*

Summary. *The paper investigates the revealed problems in modern convolutional neural networks in the image recognition. It was carried out an analysis of basic principles and architectural components of convolutional networks. It was substantiated the necessity of creating a new type of routing between network layers instead of the existing pooling method.*

Key words: *deep learning, image recognition, image classification, convolutional neural network, pooling layer.*

Вступ. На сьогодні одним з найбільш популярних варіантів використання нейронних мереж є вирішення задачі з розпізнавання об'єктів на графічних зображеннях. На сьогодні найкращі результати в області розпізнавання зображень показує згорткова нейронна мережа (далі – ЗНМ), яка є логічним розвитком ідей таких архітектур нейронних мереж (далі – НМ), як когнітрон і неокогнітрон. Успішні результати роботи обумовлені можливістю врахування двовимірної топології зображення, на відміну від

багатошарового персептрона [1]. Проте у роботі ЗНМ також існують проблеми, викликані використанням агрегувальних шарів, які відкидують інформацію про просторові характеристики об'єктів.

Просторова інваріантність ЗНМ. Базуючись на архітектурі спільних вагових наборів, згорткові НМ ще називають інваріантними відносно зсуву або просторово інваріантними нейронними мережами. Це означає, що у розпізнаванні певних об'єктів на зображенні зовсім не важливо в якому саме місці знаходиться цей об'єкт, адже при проходженні по мережі ця інформація буде втрачена.

Кожен канал згорткового шару має спільні набори ваг, тож окремий об'єкт на зображенні та його зміщена копія будуть знайдені декількома нейронами згорткового шару, а далі передані до агрегувального шару, де вони обидві будуть стиснуті до одного й того самого нейрона, тобто інформація, з якого саме нейрона нижчого рівня вони прийшли, буде втрачена. Таким чином, пройшовши по всіх згорткових шарах, на виході буде отриманий один і той самий результат, тобто вихідне значення буде інваріантним відносно таких трансформаційних змін у вхідному зображенні.

Логіка згорткової НМ побудована саме на цьому принципі агрегування для зниження дискретизації. Без цього кроку вихідні значення згорткового шару являються еквіваріантними до змін зсуву (рис. 1).

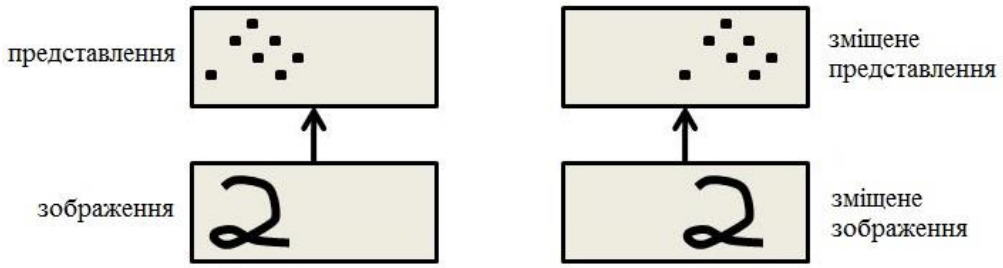

Рис. 1. Еквіваріантність ЗНМ без використання агрегування

І хоча ЗНМ успішно працюють та виконують свої задачі, проте логіка інваріантності здається хибною при розпізнаванні зображень. Тобто неправильно очікувати, що при зміні положення чого-небудь на зображенні, результат буде один і той самий. Правильно буде очікувати, що результат зміститься так само, відображаючи зміну положення об'єкта на зображенні.

Потреба у аугментації датасетів. Із принципу інваріантності ЗНМ слідує ігнорування позиційної інформації зображення. Тобто це напряму створює наступну проблему у розпізнаванні: наприклад, проводячи тренування нейронної мережі для розпізнавання рукописної літери "Я" на звичайному зображенні з рівним розташуванням літери (рис. 2 – а), мережа успішно навчиться її розпізнавати. Але коли на тестовому зображенні буде та ж сама літера, проте повернута на 20° (рис. 2 – б), мережа вже не зможе розпізнати на цьому зображенні трохи повернуту літеру, просто через те, що вона була навчена розпізнавати тільки рівну літеру.

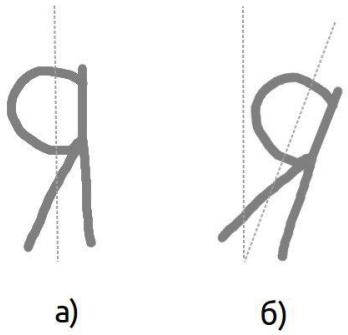

Рис. 2. Ігнорування позиційної інформації зображення на прикладі рукописної букви: а) – буква "Я"; б) – повернута буква "Я"

Для вирішення цієї проблеми необхідно було б створювати величезні датасети, які окрім базових зображень, містили б ще й усі можливі похідні від таких зображень, щоб нейронна мережа навчалася знаходити об'єкти з певними змінами. Такий підхід значно збільшує час та складність створення навчальних даних, тому в реальних умовах він не є ефективним.

На практиці використовується штучна генерація необхідних даних.

Беруться базові шаблони та "ідеальні" приклади, і застосовуючи певні спотворення, на їх основі генерується повний необхідний для навчання датасет. Цей підхід називається аугментація даних, тобто їх штучне роздуття.

Для аугментації можуть використовуватися наступні спотворення:

- геометричні трансформації (афінні, проекційні) (рис. 3);
- зміна яскравості, контрастності, фону зображення;
- відблиски, шуми, розмиття і т.д.

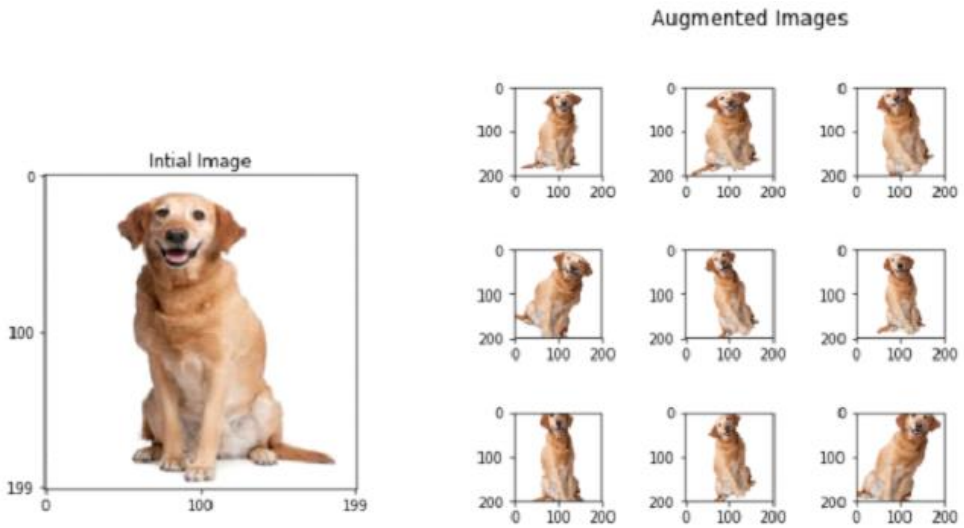

Рис. 3. Приклад аугментації базового зображення геометричними трансформаціями [1]

Також були спроби вирішення цієї проблеми шляхом вдосконалення архітектури ЗНМ, наприклад додавання циклічного згорткового шару для досягання саме поворотної інваріантності (Rotation Invariance Neural Network, 2017), проте необхідних результатів вони не дали [1].

"Проблема Пікассо". ЗНМ працюють таким чином, що шукають присутність необхідних ознак на зображенні і на основі цих даних роблять передбачення, чи існує певний об'єкт на цьому зображенні. Отже ЗНМ враховує лише наявність ознак, а не їх розташування, саме тому їй набагато важче розрізняти правильні об'єкти на зображенні, та об'єкти з потрібними

ознаками, проте з хибним розташуванням цих ознак. Для прикладу можна привести обличчя людини з різним розташуванням його характерних ознак, наприклад рота, носа, очей (рис. 4).

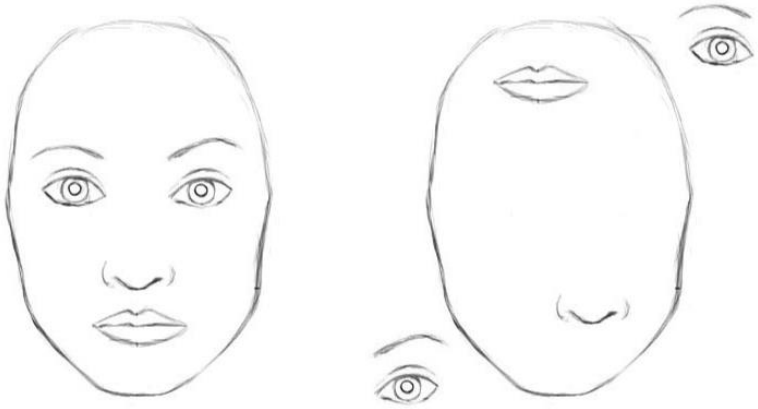

Рис. 4. Приклад зображень, обидва з яких на основі знайдених ознак будуть класифіковані ЗНМ як обличчя людини.
Так звана "проблема Пікассо" [2]

Варто зазначити, що складна багатошарова мережа все-таки може навчитися видавати вірний результат на даному прикладі, проте з такою логікою роботи їй доводиться обходити цю хибність класифікації через багато проміжних шарів, що вимагає великої глибини мережі та більшої кількості навчальних ваг [2]. Тобто така логіка пошуку об'єктів не є базово вірною з біологічної точки зору.

Тестування типу "білого ящика". Аналіз роботи нейронних мереж проводять, в тому числі, за допомогою тестування типу "білого ящика", коли знаючи всю інформацію про структуру нейронної мережі намагаються порушити її роботу або обманути, отримавши хибний результат.

Ґрунтуючись на тому, що для навчання ЗНМ використовується зворотнє розповсюдження похибки з градієнтним спуском, для обману нейронної мережі було використано змагальну мережу. Вона допомагає знайти градієнти, зміна яких найбільше впливає на отримані ймовірності при класифікації зображень [3]. Таким чином знайдені величини градієнтів

необхідно додати до вхідного зображення, щоб нейронна мережа класифікувала його хибно, при тому, що саме зображення візуально залишається таким самим (рис. 5).

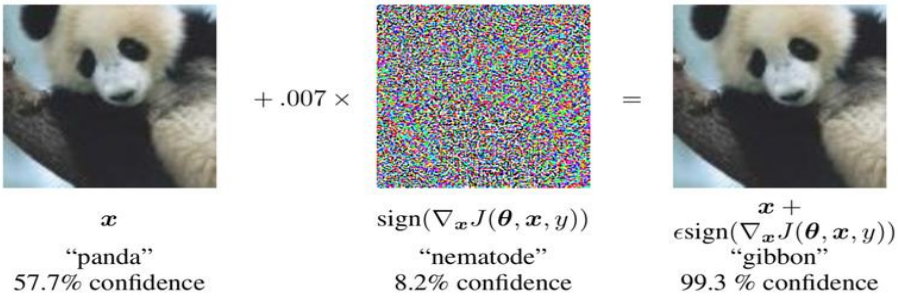

Рис. 5. Результати обману ЗНМ шляхом зміни градієнтів зображення [3]

Нещодавно було описано ще один метод обману нейронної мережі, при якому не отримується доступ до градієнтів моделі - атака одного пікселя (One pixel attack, 2017) [4]. Використовуючи метод диференціальної еволюції, проводяться ітерації генерування "дочірніх" зразків відносно "батьківських" та вибірка найкращого. На основі цих даних можна знайти один піксель на зображенні, зміна якого має найбільший вплив на вихідний результат мережі (рис. 6).

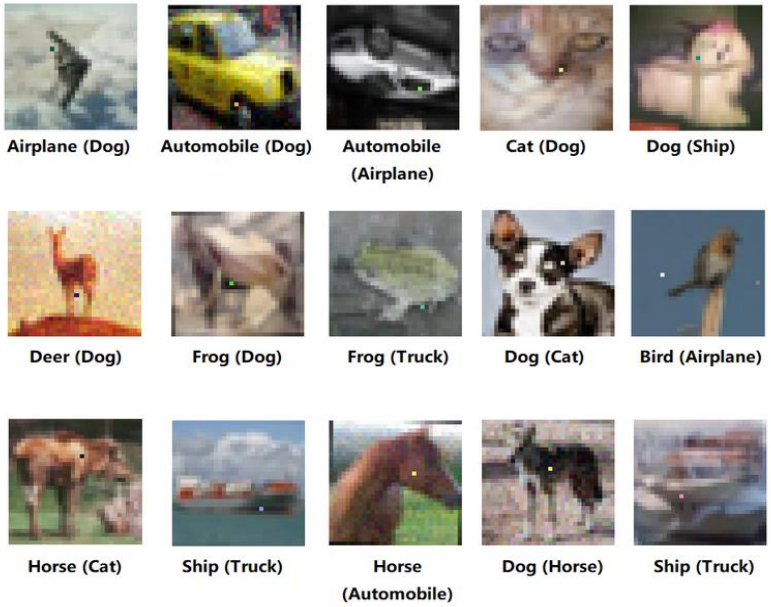

Рис. 6. Результати атаки одного пікселя на ЗНМ для класифікації зображень з датасету CIFAR10. У дужках невірна відповідь мережі на зображення з одним зміненим пікселем [4]

Дослідження показали, що 68% зображень з датасету CIFAR10 та 42% зображень з датасету ImageNet можуть бути обдурені принаймні на 1 клас зміною лише одного пікселя з ймовірністю 97% [4].

Базові принципи ЗНМ. Сучасні ЗНМ, які показують найкращі результати у розпізнаванні зображень, базуються на наступних принципах (знаком питання позначено найбільш неоднозначну концепцію структури):

- ✓ Використовується багато шарів згортки навчених детекторів ознак;
- ✓ Детектори ознак є локальними і кожен тип детектору поширений на всю мережу;
- ✓ Набори детекторів ознак стають все більшими у верхніх шарах;
- ✓ Шари виділення ознак чергуються з шарами підвибірки, які агрегують виходи сусідніх детекторів ознак одного типу.

Аргументи проти використання методу агрегування:

1. Агрегування не відповідає психології розуміння форми об'єктів. Воно не пояснює, чому ми присвоюємо об'єктам певні внутрішні границі координат і чому це має такий значний вплив.
2. Агрегування вирішує не ту задачу. Ми прагнемо досягти еквіваріантності у розпізнаванні зображень, а не інваріантності. Потрібно групувати та взаємопов'язувати, а не відкидати зайве.
3. Агрегування програє, тому що не використовує структуру нижчого рівня. Воно взагалі не звертає уваги на початкове лінійне різномаїття, яке містить у собі важливу дисперсію для аналізу.

Можна зробити висновок, що основні ідеї згортки нейронів працюють вірно та логічно, проте саме чергування згорткових шарів з агрегувальними шарами є ключовим неправильним принципом який забирає з процесу навчання важливу інформацію про знайдені ознаки.

Висновки. На основі досліджень та тестування типу "білого ящика" визначено існуючі проблеми при розпізнаванні зображень згортковими нейронними мережами. До них належать: необхідність попереднього

штучного розширення датасету для розпізнавання невеликих трансформаційних змін об'єкта; можливість обману мережі додаванням обрахованого градієнту до вхідного зображення; можливість обману мережі зміною одного пікселя у вхідному зображенні.

ЗНМ є інваріантними до зсуву ознак вхідного зображення, проте насправді правильною логікою вважається зміна характеристик результату при зміні початкових положень різних об'єктів зображення.

Тому актуальною є потреба у створенні нового типу роутингу нейронів нижчого рівня до високорівневих шарів, який би враховував просторові властивості кожної знайденої ознаки, а також їх взаємопов'язаність при формуванні всього знайденого об'єкта.

Література

1. Deshpande A. Understanding Convolutional Neural Networks [Електронний ресурс] / Adit Deshpande. – 2016. – Режим доступу до ресурсу: https://adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks
2. Tarrasse M. What is wrong with Convolutional neural networks? [Електронний ресурс] / Mahmoud Tarrasse / Medium. – 2017. – Режим доступу до ресурсу: https://towardsdatascience.com/what-is-wrong-with-convolutional-neural-networks-75c2ba8fbd6f
3. Gupta H. Four ways to easily fool your neural network [Електронний ресурс] / Harshvardhan Gupta / Medium. – 2018. – Режим доступу до ресурсу: https://buzzrobot.com/4-ways-to-easily-fool-your-deep-neural-net-dca49463bd0.
4. Kouichi S. One pixel attack for fooling deep neural networks [Електронний ресурс] / S. Kouichi, J. Su. / Cornell University Library. – 2017. – Режим доступу до ресурсу: https://arxiv.org/abs/1710.08864