

**Секція 3. AI (Інтелектуальні системи.
Машинне навчання, bigdata).
Section 3. AI (Intelligent Systems.
Machine learning, big data.).**

УДК 004.93'12

Юрій Тимошин, Сергій Орленко

**ВПЛИВ ПЛЯМИ СПРИЙНЯТТЯ НА ШВИДКІСТЬ
ТА ТОЧНІСТЬ РОЗПІЗНАВАННЯ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ**

**INFLUENCE OF THE RECEPTIVE FIELD
ON THE SPEED AND ACCURACY OF RECOGNITION
OF THE CONVOLUTIONAL NEURAL NETWORK**

У статті розглянута проблема плями сприйняття та способу завдання її розмірів за допомогою зміни топології загорткової нейронної мережі. Наглядно були показані способи досягнення одного розміру плями сприйняття різними способами, виконаний ряд досліджень та розроблені програми для наборів даних MNIST та Sifar, що відображають досягнуту точність розпізнавання та затрачений час на навчання.

Ключові слова: згорткові нейронні мережі, пляма сприйняття, точність, час навчання.

Рис.: 6. Бібл.: 3.

The paper deals with the receptive field problems and methods of its definition with the help of changing topology of convolutional neural network. With examples shown ways how to achieve the same receptive field in different ways, also performed a number of studies and developed programs for datasets, MNIST and Cifar, which reflects the achieved recognition accuracy and time spent on training.

Key words: convolutional neural network, receptive field, accuracy, time of study.

Fig.: 6. Bibl.: 3.

Вступ. При побудові системи розпізнавання образів заснованої на штучній нейронній мережі (НМ) вибір типу мережі найчастіше вирішений наперед – це згорткова нейронна мережа (ЗНМ). Запропонована в 1988 році Яном Лекуном вона розроблена на подібні біологічного зору і зорової кори та складається з простих і складних клітин, викликала справжній фурор і відродила інтерес до НМ.

Назва мережа отримала через наявність операцій згортки. ЗНМ має деяку специфіку – вона полягає у перестановці загорткових шарів і шарів субдискретизації (пулінг), що і відсилає її до біологічного зору. Так у людини є прості клітини, які реагують на зміни градієнта, такі як кути і лінії в певній точці – аналог згортки; складні ж клітини інваріантні до зсувів і реагують при активації набору простих клітин – аналог пулінг [1]. На більш пізніх шарах ядра згортки поступово змінюються і формують певне семантичне значення. Оскільки структура мережі багат шарова – однонаправлена і закінчується вона декількома повнозв'язними шарами, втрачається просторова структура пікселів [2]. Хоча

саме процес вилучення абстрактних понять з вихідного зображення і забезпечує ЗНМ таку високу точність.

Постановка задачі. Як уже було згадано, виділення семантичних ознак на пізніх етапах є результатом ряду згорток і субдискретизацій, однак слід розуміти, що всі ці операції проводяться над певною початковою областю – плямою сприйняття (receptivefield) [3]. В цілому для ЗНМ вхідна картинка і є зображенням що сприймається, однак не все так просто. Кожна операція (або набір операцій) згортки перед пулінгом захоплює лише частину зображення і це визначає подальшу обробку.

Дане питання стосується архітектури ЗНМ, і оскільки її вибирає розробник, виходячи з кращих прикладів, розглянемо операції, що стосуються перетворення плями сприйняття. Також слід вказати про важливість зазначених нижче методів – всі вони зменшують число параметрів НМ, що покликане знизити обсяг обчислень і прискорити навчання/перенавчання ЗНМ. Завдання полягає у відстеженні отриманої точності і часу навчання при кожному підході.

Підходи до зміни плями сприйняття та число параметрів. До таких підходів, як не дивно, відноситься і пулінг. «Навіщо потрібен пулінг, якщо є фільтр побільше?» – таке питання може поставити будь-який уважний інженер. Субдискретизація, подібно до великого фільтру, покликана збільшити пляму сприйняття (хоча при цьому втрачається локалізація) і при цьому посилити виділення шуканої особливості. Так при пулінгу 2 на 2 пляма сприйняття збільшується в 2 рази.

Припустимо, у нас є зображення розміру 8x8 і ми хочемо отримати з нього якісь особливі властивості. І тут постає вибір – застосувати великий фільтр або фільтр поменше і операцію пулінгу. Припустимо, зображення було піддано згортці з ядром 3x3 і для отриманої карти ознак 6x6 ми застосували субдискретизацію вікном 2x2. В результаті отримали нову карту ознак розміром 3x3, яка містить найосновніші значення особливих властивостей і може використовуватися і далі. Наприклад, взяти ще одну згортку з ядром три на три, вийде послідовність (3x3 | pooling | 3x3), а результат буде одиничним значенням (Рис. 1. а). Тепер уявімо, що ми використовували фільтр розміром 8x8 (Рис. 1. б) – отримали таке ж єдине значення як і після двох попередніх перетворень.

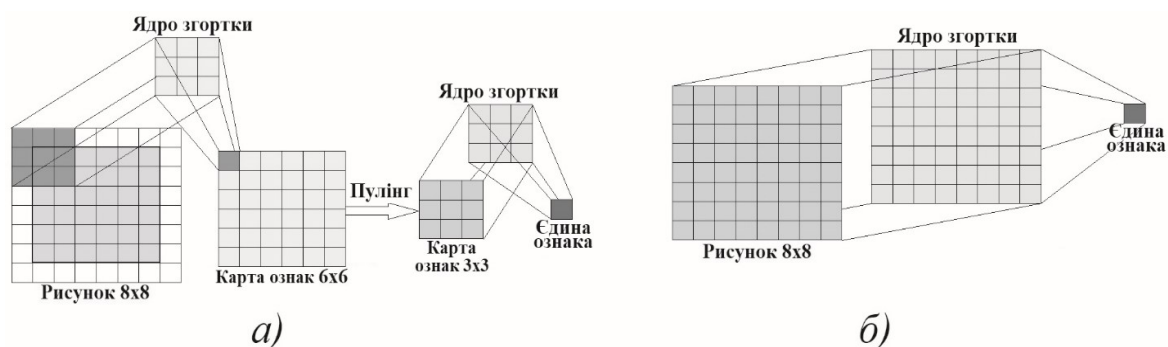


Рис. 1. Використання пулінга замість великого фільтра

Виходить, що при першому підході можна отримати узагальнений результат, який зручно використовувати і далі, а при другому – лише єдине значення.

Тепер розглянемо кількість параметрів. У першому підході при першій згортці використовувалися $3 \times 3 + 1$ параметрів, а для отримання одиничного значення ще $3 \times 3 + 1$ параметрів – разом 20, в той час при другому: $8 \times 8 + 1 = 65$.

Однак, при очевидному зменшенні параметрів спостерігається зворотна картина з обчисленнями при прямому проходженні через мережу – для даного рішення кількість арифметичних операцій для другого підходу в 5.2 рази менше ніж для першого.

Наступний підхід до вибору плями сприйняття вельми простий. Він полягає в послідовному використанні двох або більше невеликих фільтрів замість одного великого при збереженні розмірів початкової області. Розглянемо на простих прикладах.

Візьмемо фільтр розміру 5×5 і застосуємо його до зображення 7×7 пікселів – результатом буде карта ознак розмірністю 3×3 (Рис. 2. а). Досягти збереження плями сприйняття при заміні початкового ядра згортки можна за допомогою застосування поспіль двох фільтрів розмірністю 3×3 (Рис. 2. б).

Кількість параметрів при одному фільтрі дорівнює: $5 \times 5 + 1 = 26$, при двох маленьких: $3 \times 3 + 1 + 3 \times 3 + 1 = 20$; кількість арифметичних операцій для обчислення однієї результуючої карти ознак 50 і 180 відповідно.

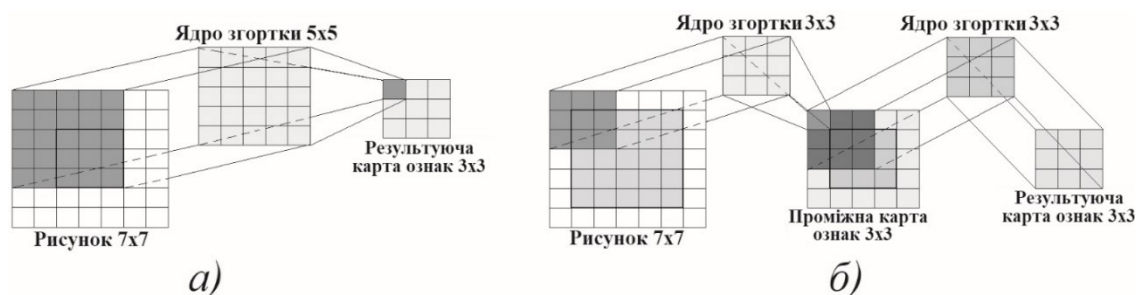


Рис. 2. Заміна фільтра 5×5 на комбінацію двох фільтрів 3×3

Подібну заміну можна провести і при початковому розмірі фільтра 7×7 за допомогою послідовного застосування вже трьох фільтрів розмірністю 3×3 . І для такої заміни у одного фільтра буде 50 параметрів та 98 арифметичних операцій, а у комбінації трьох фільтрів – 30 параметрів, але 630 операцій.

Також зазначимо можливість заміни ядра 5×5 послідовним використанням чотирьох односпрямованих векторів: $5 \times 5 \approx 1 \times 3 \rightarrow 3 \times 1 \rightarrow 1 \times 3 \rightarrow 3 \times 1$. Ілюстрація даної заміни показана на рис. 3.

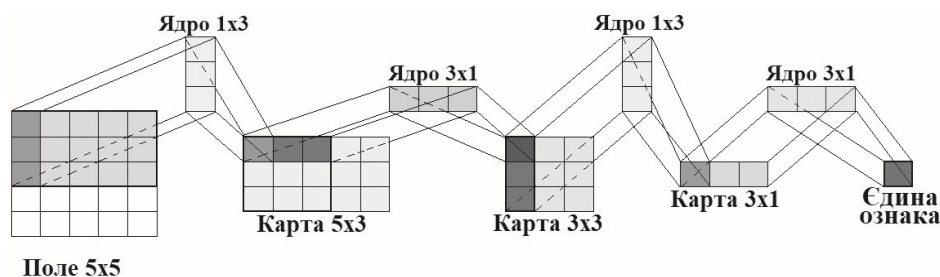


Рис. 3. Заміна фільтра 5×5 на набір одно напрямлених фільтрів

Властивості фільтра 5×5 залишилися незмінними: 26 параметрів і 50 операцій, для набору фільтрів кількість параметрів складає: $(3+1) \times 4 = 16$; кількість арифметичних операцій для обчислення останньої ознаки дорівнює 150.

Експериментально-дослідне дослідження задачі. Для перевірки ефективності наведених вище методів маніпуляції зоною сприйняття проведемо ряд експериментів. Почнемо з початкової моделі: в її якості була обрана мережа, яку легко можна було б переналаштувати для двох навчальних вибірок. Так, до вхідного зображення послідовно застосовувалися: згортка 3×3 і пулінг, глибина карт активації – 16, згортка 3×3 і пулінг, глибина – 32 і наостанок згортка 3×3 з глибиною 64; для всіх згорток функція активації ReLU, а вікно пулінгу – 2×2 . Далі дані по мережі передавалися трьома повнозв'язним шаром (функції активації перших двох шарів – тангенс гіперболічний, а третьої – softmax) по 256, 64 і 10 нейронів відповідно. ЗНМ навчалася за допомогою методу ADAM (adaptive moment estimation) з початковим темпом навчання 0.001. Так як навчальні вибірки були різних розмірів – змінювалося число параметрів, що надходять на повнозв'язні шари мережі, схожа картина спостерігалася при зміні вікна сприйняття, але оскільки ЗНМ була реалізована за допомогою бібліотеки PyTorch мови Python – мережу було легко і зручно перебудовувати.

Почнемо з збільшення першої згортки до розміру 5×5 . А потім замінимо її послідовним використанням двох згорток 3×3 шляхом додавання ще одного шару – Рис. 4 та Рис. 5. Аналогічну заміну проведемо для фільтра 7×7 двома 5×5 , паралельно перевіряючи ефективність пулінгу: прибравши його після другого згорткового шару – Рис. 6.

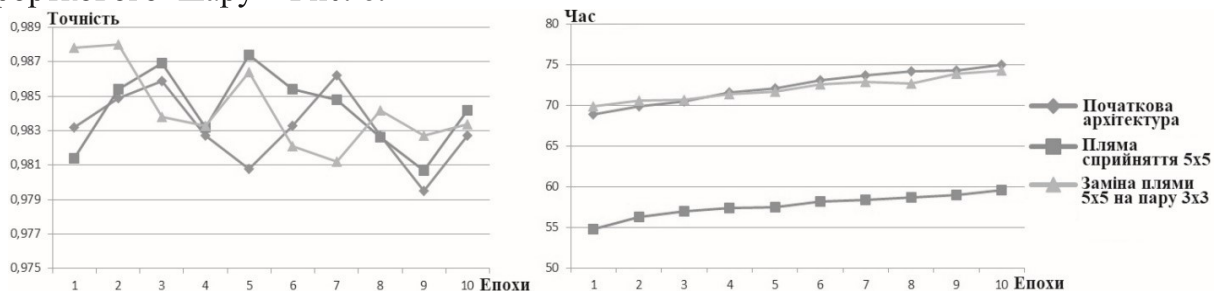


Рис. 4. Дослідження плями сприйняття 5×5 для набору MNIST

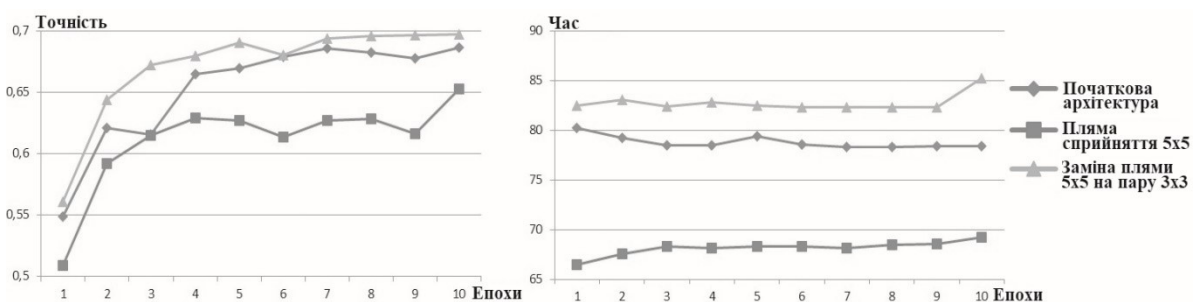


Рис. 5. Дослідження плями сприйняття 5×5 для набору CIFAR10

На обох графіках простежується чітка закономірність – просте збільшення плями сприйняття призводить до збільшення швидкості навчання ЗНМ завдяки зменшенню числа арифметичних операцій, проте на прикладі навчальної вибірки CIFAR10 видно, що при такому підході точність розпізнавання падає. Зворотна ситуація відбувається при збільшенні плями сприйняття шляхом заміщення великого фільтра двома маленькими шляхом додавання ще одного згорткового шару. Так при загальному зменшенні числа параметрів в порівнянні

з великим фільтром все ж вдається виділити більше ознак, що сприяє підвищенню точності розпізнавання, однак час витрачений на навчання зростає в зв'язку зі збільшенням обчислень.

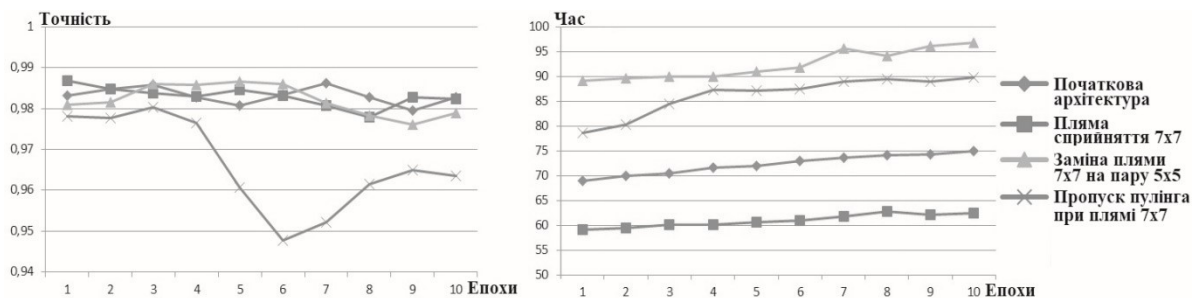


Рис. 6. Дослідження плями сприйняття 7x7 набір MNIST і пропуск шару пулінгу

Наведені вище результати підтверджуються і для вікна розміром 7x7, хоча не так очевидно зменшення точності. Пропуск субдискретизації призводить до істотного збільшення арифметичних операцій і знижує здатність ЗНМ виділяти семантичні ознаки, що і призводить до зниження точності.

Висновки. Отриманий результат дозволяє стверджувати, що вибір плями сприйняття є важливим моментом при проектуванні ЗНМ, а вибір підходу завдання плями безпосередньо впливає на точність і час навчання мережі. Збільшення шарів, як правило, призводить до збільшення точності, для цієї ж мети слід намагатися підібрати оптимальний метод навчання, функції активації мережі і початковий коефіцієнт навчання. Всі ці підходи здатні вплинути на точність і час навчання ЗНМ, тому бажано завжди тестувати і порівнювати спроектовані топології ще й на практиці.

Список використаних джерел

1. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, +Winter 1989.
2. Krizhevsky, Alex, and Geoffrey Hinton. Learning multiple layers of features from tiny images [Електронний ресурс] – 2009. – С.60. – Режим доступу: <http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
3. K. Fukushima, Neocognitron for handwritten digit recognition. Neurocomputing – April 2003 – (Vol. 51, pp. 161-180).

ДОВІДКА ПРО АВТОРІВ

Тимошин Юрій Афанасійович – к.т.н., доцент кафедри технічної кібернетики НТУУ “КПІ ім. Ігоря Сікорського”.

Наукові інтереси: інформаційно-управляючі системи, обробка інформації, складні об'єкти управління, телекомунікаційні системи та бази даних.

Timoshin Yuri – Ph.D., Associate Professor, Department of Technical Cybernetics NTUU "Igor Sikorsky Kyiv Polytechnic Institute".

Scientific interests: information-operating system, information processing, complex facilities management, telecommunications systems and databases.

E-mail: y.timoshin@gmail.com

Орленко Сергій Петрович – студент кафедри технічної кібернетики НТУУ “КПІ ім. Ігоря Сікорського”

Наукові інтереси: теорія цифрової обробки сигналів та зображень, розпізнавання образів, штучний інтелект

Orlenko Sergey – student of the Department of Technical Cybernetics NTUU "Igor Sikorsky Kyiv Polytechnic Institute".

Scientific interests: theory of digital signal processing and image pattern recognition, artificial intelligence.

E-mail: orlenko_sergey@ukr.net

Yuri Timoshyn, Sergey Orlenko

INFLUENCE OF THE RECEPTIVE FIELD ON THE SPEED AND ACCURACY OF RECOGNITION OF THE CONVOLUTIONAL NEURAL NETWORK

Introduction. The convolution neural network was developed by Yann LeCun in 1988 in likeness of biological vision and visual cortex and consists of simple and complex cells. The network has some specificity - it consists in rearranging the convolutional layers and layers of sub-sampling (pooling), which sends it to the biological view.

Formulation of the problem. The separation of semantic signs occurs at later stages, but it should be understood that all these operations are carried out over a certain initial region - a receptive field. In general, there is an image that is perceived, but it all is not so simple. Each convolutional operation before the pooling captures only part of the image and this determines the further processing. This question relates to the architecture of the convolution neural network so we will consider the operations concerning the transformation of the receptive field. The task will be to track the obtained accuracy and time of training with each approach.

Approaches to change receptive field and number of parameters. We considered pooling from the point of increasing the absorption of signs from the larger field, considered the possibility of replacing the 5x5 filter with two 3x3 or by a sequence of $1 \times 3 \rightarrow 3 \times 1 \rightarrow 1 \times 3 \rightarrow 3 \times 1$, and replacing 7x7 with two 5x5. We have calculated the number of parameters and arithmetic operations for each approach, also created illustration for them.

Experimental-research study of the problem. We built and changed the convolution neural network according to the methods described above. As the data set we were used traditional MNIST and CIFAR10. In parallel, we recorded the resulting accuracy and elapsed training time and depicted them on the corresponding graphs.

Conclusions. Choice of receptive field is an important point in the design of CN, and the choice of approach to the problem of receptive field directly affects the accuracy and time of learning the network. Increasing the layers, as a rule, leads to increased accuracy, for the same purpose should try to choose the best method of training, network activation functions and the initial learning factor. So it is advisable to always test and compare projected topologies even in practice.

Key words: convolutional neural network, receptive field, convolution, topology, parameter, accuracy, time of study.