

Н.А. Гук, Д.С. Малишко

Дніпровський національний університет імені Олеся Гончара

ЗАСТОСУВАННЯ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ДО ЗАДАЧ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ

Роботу присвячено вибору архітектури згорткової нейронної мережі для розв'язання задач класифікації зображень. Побудовано математичну модель нейронної мережі. Для запобігання проблемі перенавчання застосовано регуляризацію Тихонова. Проведено серію обчислювальних експериментів, за якими створено рекомендації щодо проектування архітектури мережі.

Ключові слова: згорткова нейронна мережа, класифікація, перехресна ентропія, регуляризація.

N.A. Huk, D.S. Malyshko

Oles Honchar Dnipro National University

APPLICATION OF CONVOLUTIONAL NEURAL NETWORKS TO IMAGE CLASSIFICATION PROBLEMS

In this article, the properties and possibilities of convolutional neural networks for solving classification problems are investigated.

Based on the results of the analysis of literature sources, the state of the problem to date has been investigated, the main architectures of modern convolutional neural networks, and their parameters have been determined. The problem of retraining of neural networks is investigated and such a way of counteracting retraining as Tikhonov regularization, special cases of which are L1 and L2 regularizations, is clarified.

A mathematical model of a convolutional neural network has been built, which allows the use of a set of layers, such as convolutional, subsample, dropout, and fully connected layers. The use of a set of activation functions, error metrics, and regularization methods is envisaged. ReLU due to the high convergence of the training method and Softmax due to the possibility of interpreting the results as the probability of belonging of the input image to the desired class is considered as the main activation function. Cross entropy is considered as the main loss function, which is due to the choice of the Softmax activation function. The use of the dropout layer, L1, and L2 regularizations as effective methods of preventing retraining is proposed as regularization methods.

The given model of the convolutional neural network is programmatically implemented. Its work has been tested on the MNIST handwritten number set. Adam's adaptive batch gradient descent method was used to train the network. Testing was performed with different variations of the above network components. To stop training, the method of an early stop is used, which stops training the network in the absence of error reduction. This allowed the network's learning rate to be estimated to a minimum error value. Based on the results of computational experiments, recommendations for the choice of the architecture of the activation and error functions, setting the parameters of the network layers are formulated. Based on the results of the comparison of regularization procedures, the use of L2 regularization is proposed, which helps to prevent retraining of the network.

Keywords: convolutional neural network, classification, cross entropy, regularization.

Н.А. Гук, Д.С. Малышко

Дніпровський національний університет імені Олеся Гончара

ПРИМЕНЕНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ К ЗАДАЧАМ КЛАССИФИКАЦИИ ИЗОБРАЖЕНИЙ

Работа посвящена выбору архитектуры сверточной нейронной сети для решения задач классификации изображений. Построена математическая модель нейронной сети. Для предотвращения проблемы переобучения применена регуляризация Тихонова. Проведена серия вычислительных экспериментов, сформулированы рекомендации по проектированию архитектуры сети.

Ключевые слова: сверточная нейронная сеть, классификация, перекрестная энтропия, регуляризация.

Вступ. Сьогодні у багатьох галузях знань застосовуються сучасні технології штучних нейронних мереж, за допомогою яких успішно розв'язуються задачі класифікації та прогнозування. В залежності від постановки задачі використовуються багатошаровий перцептрон, мережа Кохонена, мережа Хопфілда, рекурентні мережі, згорткові мережі, архітектуру яких необхідно будувати та досліджувати для кожної окремої задачі. Важливим класом задач є задача класифікації зображень. До розв'язання задач класифікації можна застосовувати мережі різних типів, найпростішими із яких є перцептрон, а найпоширенішими – згорткові нейронні мережі.

Застосування перцептрону викликає певні труднощі, пов'язані зі значними розмірами зображень, що у свою чергу потребує подання на вхід нейронної мережі великих за обсягом вхідних векторів, наявності значної кількості нейронів у шарах мережі та синаптичних зв'язків між ними. Налаштування таких мереж ускладнюється, доволі часто спостерігається проблема перенавчання мережі. Крім того, для якісної класифікації бажано враховувати зв'язок між сусідніми пікселями зображення, оскільки їх сукупність описує певні ознаки об'єкту та може суттєво покращити результат. Однак застосування перцептрону не дозволяє врахувати двовимірну структуру зображень під час формування вхідних векторів навчальної вибірки.

Для врахування зазначених особливостей останнім часом розвитку набуло застосування згорткових нейронних мереж [6], архітектура яких підтримує обробку локальних ознак зображення, формування шарів мережі у вигляді карт ознак, локальне усереднення, що сприяє скороченню числа параметрів за рахунок спільного використання синаптичних ваг нейронами однієї карти, стійкості мережі до зміщення та деформації зображення.

Проте не існує чітких критеріїв для вибору архітектури мережі, а лише рекомендації, які засновані на власному досвіді та проведених обчислювальних експериментах.

Аналіз літературних джерел. Аналіз основних архітектур глибоких нейронних мереж, що застосовуються до задач класифікації, виявляє тенденції їх

розвитку та основні параметри, вибір та налаштування яких надає мережам якісні переваги.

Першою згортковою мережею прийнято вважати LeNet5 [6], архітектура якої передбачає використання шарів згортки, підвибірki за середнім елементом та застосування двошарового перцептрону в якості класифікатора. Її архітектура є фундаментальною для побудови сучасних мереж класифікації зображень, оскільки наявність матриць згорток дозволяє класифікувати ознаки об'єкту, а не окремі пікселі зображення.

Прорив у використанні графічних процесорів для тренувань мереж привів до створення архітектури AlexNet [4], яка є глибокою мережею та складається з 9 шарів, серед яких вперше використано шар dropout, функцію активації ReLU та підвибірki за принципом максимального елемента.

Застосування наведених архітектур до розв'язання практичних задач класифікації зображень потребує вибору та налаштування функції активації, регуляризуючих процедур, послідовності розташування та кількості шарів у архітектурі. Ґрунтовний вибір параметрів суттєво покращує роботу мереж та надає їм нові властивості.

Незалежно від обраної архітектури мережі в процесі навчання часто виникає проблема збільшення помилки мережі на тестовій вибірці зі збільшенням числа епох навчання. Зазначене явище називають проблемою перенавчання нейронних мереж. Для усунення проблеми пропонується скорочувати кількість параметрів, які потребують налаштування, удосконалювати алгоритми навчання. Однак найчастіше для усунення проблеми перенавчання застосовуються процедури регуляризації. Вони можуть бути проведені як на рівні представлення вхідних даних, так і на рівні моделі мережі.

Процедурою, що виконує регуляризацію на рівні вхідних даних є аугментація даних (data augmentation). Її застосування передбачає створення додаткових даних для навчання мережі з тих, що вже наявні. Аугментація даних направлена на зменшення впливу локальних особливостей зображень і підвищення здатності мережі до узагальнення. Так в роботі [11] застосовується вилучення частини зображення, в роботі [9] пропонується створення нових зображень з частин наявних, а автори [10] використовують змішування інформації з двох зображень навчальної вибірki. Однак можливість застосування таких методів сильно залежить від виду даних в навчальній вибірці і аугментація не може бути універсальним засобом для усунення проблеми перенавчання.

Регуляризація на рівні моделі здійснюється шляхом накладення обмежень на параметри мережі. До таких методів можна віднести класичний метод регуляризації А.М. Тихонова [1]. Стосовно нейронних мереж обмеження накладається на значення ваг синаптичних зв'язків. В роботі [5] показано, що наявність обмежень послаблює вплив нерелевантних компонент вектора ваг, а також зменшує вплив статичного шуму.

Метод проріджування (Dropout) [8, 2] ґрунтується на принципах навчання нейронної мережі і передбачає виключення деякої кількості нейронів шару,

вибраних випадковим чином, під час навчання. Кількість нейронів для включення вказується у відсотках нейронів шару. Таким чином, генеруються нові мережі, а процес навчання зводиться до навчання ансамблю з декількох мереж з наступним усередненням отриманих результатів. Стосовно згорткових шарів описаний метод полягає у введенні розподілу ймовірностей для ваг матриць згортки.

Ще один з підходів до підвищення узагальнюючої здатності висуває вимогу до властивостей вагової матриці шарів [3], а саме її ортогональності, лінійна незалежність вагових векторів нейронів забезпечує ідентифікацію різних особливостей зображення і підвищує здатність до узагальнення.

У даній роботі розглянуто застосування згорткової нейронної мережі для задачі класифікації зображень, експериментально досліджено вплив особливостей архітектури мережі на результат класифікації та доцільність застосування регуляризації для усунення проблеми перенавчання мережі.

Постановка задачі. Задачу класифікації зображень сформулюємо у такий спосіб: нехай існує скінченна множина класів Y та множина об'єктів D . Для скінченної підмножини об'єктів M , $M \subseteq D$ відомо, до якого класу відносяться об'єкти. Зазначена підмножина використовується у якості навчаючих прикладів. Необхідно побудувати алгоритм, який здійснює відображення $D \rightarrow Y$ та класифікує довільний об'єкт $d \in D$. Для забезпечення належної якості розв'язання задачі необхідно налаштувати параметри алгоритму та дослідити його властивості.

Математична модель згорткової мережі. Для розв'язання задачі класифікації зображень застосуємо згорткову нейронну мережу [6]. Основними елементами архітектури згорткової нейромережі є два модулі, перший відповідає за згортку, тобто обробку зображення певним чином, а другий за класифікацію даних, що отримані з першого модулю. Основними шарами частини мережі, що обробляє зображення, є згортковий шар та шар підвибірки. Згортковий шар містить у собі нейрони-матриці, на вхід до яких поступово надходять ділянки зображення. Кожен з таких нейронів тренується так, щоб набути здатність до класифікації певної ознаки об'єкту. Для кожної вхідної ділянки зображення нейрон надає відгук, який можна інтерпретувати як міру подібності вхідних даних до шаблону. Такі відгуки формують двовимірний масив, який називається картою ознак. Шар підвибірки приймає ділянку карти ознак та за визначеним правилом формує єдиний відгук для масиву вхідних даних. Наприклад, нейрони підвибіркового шару мережі LeNet5 приймали ділянку 2×2 та розраховували для неї одне середнє значення, що суттєво зменшує розмірність даних. Нейрони підвибіркового шару дозволяють сформувати карту найістотніших ознак. Модулем класифікації є повнозв'язні шари прямого розповсюдження по типу перцептронів.

Модель нейронів повнозв'язного шару має вигляд:

$$\hat{y} = F_{act} \left(G(x_1, x_2, \dots, x_s, w_1, w_2, \dots, w_s) \right), \quad (1)$$

де x_i – елемент вхідних даних; w_i – вага синаптичного зв'язку; G – функція зваженої суми; $G(x_1, x_2, \dots, x_K, w_1, w_2, \dots, w_K) = \sum_{i=1}^S (x_i \cdot w_i)$; $i = \overline{1, S}$; S – кількість синаптичних зв'язків; $F_{act}(x)$ – функція активації нейрону.

У якості функції активації обирається довільна нелінійна функція однієї змінної. В роботі застосовується функція активації ReLU:

$$f(x) = \max(0, x). \quad (2)$$

Особливою функцією активації є функція Softmax, яка враховує суму виходів всіх нейронів шару та набуває вигляду:

$$\hat{y}_i = \frac{e^{x_i}}{\sum_{k=1}^N e^{x_k}}, \quad (3)$$

де N – кількість нейронів у шарі; $i, k = 1, N$.

Завдяки властивостям функції Softmax виходи мережі зручно інтерпретувати як ймовірність належності вхідних даних до певних класів.

Для організації процедури тренування мережі необхідно ввести функцію похибки. Враховуючи вибір функції активації Softmax у якості функції похибки шару доцільно використати перехресну ентропію.

$$H(y, \hat{y}) = -\frac{1}{T} \sum_t \sum_t [y_i \ln \hat{y}_i + (1 - y_i) \ln(1 - \hat{y}_i)], \quad (4)$$

де y_i – відповідь i -ого нейрону, яка очікується, \hat{y}_i – дійсна відповідь i -ого нейрону, $t = 1, T$, T – кількість елементів тренування (для стохастичних методів $T = 1$), $i = \overline{1, N}$, N – кількість нейронів у шарі.

Модель нейрону згорткового шару має вигляд:

$$y_{ij} = b + \sum_{s=1}^M \sum_{t=1}^M w_{st} x_{(i-M/2+s)(j-M/2+t)}, \quad (5)$$

де (i, j) – координати елементу матриці, з яким суміщено центр ядра згортки; y_{ij} – відгук ядра згорткового шару; b – зміщення ядра згорткового шару; M – розмірність матриці згортки; w_{st} – вага синаптичного зв'язку елементу ядра згортки; x – елемент вхідної матриці.

Зауважимо, що для вхідної матриці розміру $a \times b$ $M/2 \leq i \leq a - M/2$; $M/2 \leq j \leq b - M/2$. Тоді розмірність вихідної карти ознак дорівнює $(a - M) \times (b - M)$.

Підвибірковий шар виконує перетворення у такий спосіб:

$$y_{ij} = b + 1/M^2 \sum_{s=1}^M \sum_{t=1}^M x_{(i-M/2+s)(j-M/2+t)}. \quad (6)$$

Методи навчання мережі та запобігання перенавчанню. Навчання нейронної мережі має на меті мінімізацію функції похибки (4) шляхом знаходження ваг синаптичних зв'язків між нейронами шарів.

У якості методу навчання застосовується метод градієнтного спуску, заснований на принципі ітераційного наближення до найближчого мінімуму функціоналу похибки.

Враховуючи особливості формування функціоналу похибки у задачах класифікації зображень, а саме неповноту тренувальної вибірки, що призводить до виділення несуттєвих рис класу об'єктів, можливе перенавчання нейронної мережі. Це такий стан мережі, коли успішна класифікація відбувається виключно на даних, які використовувалися для тренування. Для даних, що не входили до тренувальної вибірки, точність класифікації є низькою, а значення помилки мережі зростає.

Для подолання перенавчання у роботі застосовується метод проріджування (Dropout) та регуляризація Тихонова [1]. Регуляризація Тихонова обмежує надмірне зростання компонент вектору ваг синаптичних зв'язків мережі шляхом додавання складової до функціоналу похибки (4).

У машинному навчанні широко застосовують L1 та L2 регуляризацію, для яких додаткові складові є нормованими модулем $\frac{\lambda}{2T} \sum_{i=1}^{S_{net}} |w_i|$ та квадратич-

ною функцією $\frac{\lambda}{2T} \sum_{i=1}^{S_{net}} w_i^2$ відповідно, де S_{net} – загальна кількість синапсів

мережі. Зауважимо, що регулюючий доданок не враховує ваги зсувів (biases) нейронів.

Тоді функція похибки набуває вигляду:

$$H_{L_1}(y, \hat{y}) = -\frac{1}{T} \sum_t \sum_i [y_i \ln \hat{y}_i + (1 - y_i) \ln (1 - \hat{y}_i)] + \frac{\lambda}{2T} \sum_{i=1}^{S_{net}} |w_i| \quad (7)$$

$$H_{L_2}(y, \hat{y}) = -\frac{1}{T} \sum_t \sum_i [y_i \ln \hat{y}_i + (1 - y_i) \ln (1 - \hat{y}_i)] + \frac{\lambda}{2T} \sum_{i=1}^{S_{net}} w_i^2 \quad (8)$$

Проектування, навчання та тестування нейронної мережі. Для опису архітектури мережі введемо такі позначення:

«Cnv([a, b], r, f)» – згортковий шар, де [a, b] – розмір ядра згортки, r – кількість ядер згортки, f – функція активації;

«Sbs([a, b])» – підвибірковий шар, де [a, b] – розмір ядра підвибірки, функція активації – $f(x_1, x_2, \dots, x_n) = \max(x_1, x_2, \dots, x_n)$;

«Fcn(a, f)» – повнозв'язний шар, де a – кількість нейронів шару, f – функція активації шару;

«Flt()» – шар випрямлення, який перетворює матриці карт ознак, або зображення у вектор;

«>» – оператор, що описує послідовність з'єднання шарів у мережі.

Для проведення обчислювальних експериментів використовувалась базова архітектура згорткової мережі, опис якої згідно до введених означень має вигляд: «Cnv([3, 3], 32, «ReLU») > Sbs([2, 2]) > Flt() > Fcn(128, «ReLU») > Fcn(10, «Softmax»)». Для навчання та тестування роботи мережі було використано базу даних рукописних цифр MNIST, яка складається з 70000 зображень у відтинках сірого 28x28 пікселів. 60000 екземплярів було обрано для навчання мережі, а 10000 об'єктів для валідації. Тренування було проведено за адаптивним алгоритмом пакетного градієнтного спуску «Адам»[7].

На рис. 1 наведено залежності точності класифікації та значення похибки від кількості епох навчання. Можна бачити, що базова модель згорткової нейронної мережі досягає точності у 97% і похибки 0,09 за десять епох навчання. При збільшенні кількості епох навчання для тієї ж моделі можна бачити (рис. 2), що 100% точність досягається лише на тренувальних зображеннях. Залежність похибки на валідаційній вибірці починає зростати вже після 15 епохи навчання, її поведінка має осцилюючий характер, що вказує на перенавчання мережі, мережа завчає тренувальні зображення і реагує лише на них.

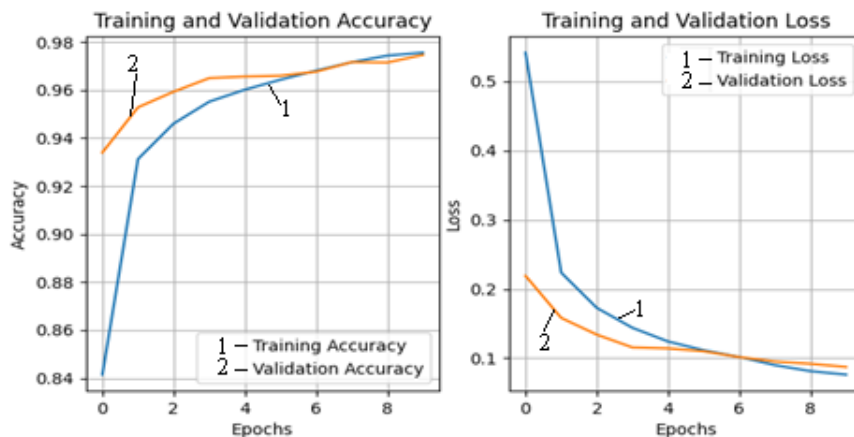


Рис. 1. Графік тренувань базової згорткової мережі (10 епох)

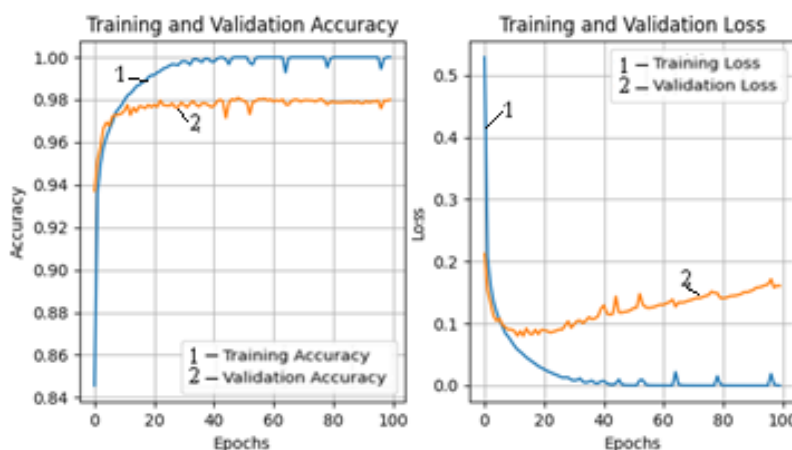


Рис. 2. Графік тренувань базової згорткової мережі (100 епох)

Оскільки для розглянутої мережі встановлено, що 10-15 епох навчання є достатнім для досягнення прийнятної точності, у роботу алгоритму навчання було введено механізм «ранньої зупинки», який визначає оптимальну кільк-

кість епох навчання та зупиняє процес навчання, якщо функція помилки не зменшує своє значення протягом 5 епох.

Було проведено серію обчислювальних експериментів для мереж з різною архітектурою та параметрами. На рис. 3 наведено залежності точності класифікації та значення похибки в залежності від кількості епох навчання для мережі, яка відрізняється від базової наявністю додаткової пари згорткового $Cnv([3, 3], 32, «ReLU»)$ та підвибіркового шарів $Sbs([2, 2])$.

З аналізу наведених залежностей можна бачити, що додавання додаткових шарів уповільнює процес навчання, точність після 10 епохи навчання сягає 95% та незначно покращується до 97% після 40 епох навчання. Проте розглядаючи детально шари мережі і зіставляючи їм формат вхідних даних (матриця 28×28) стає очевидно, що для кожної з 32 карт другий шар згортки перетворює карту ознак на матрицю 2×2 , а другий шар підвибірки на вихід надає лише один піксель з карти, що явно ускладнює роботу класифікатора і уповільнює навчання.

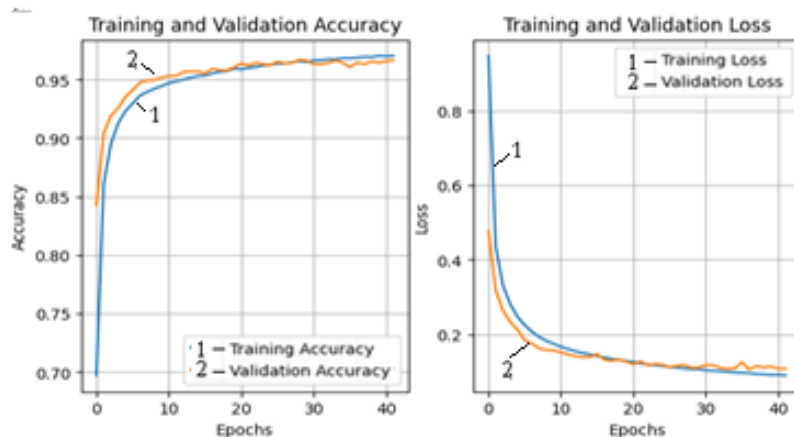


Рис. 3. $Cnv([3, 3], 32, «ReLU») > Sbs([2, 2]) > Cnv([3, 3], 32, «ReLU») > Sbs([2, 2]) > Flt() > Fcn(128, ReLU) > Fcn(10, Softmax)$

Також було виявлено, що додавання згорткового шару із матрицею 5×5 суттєво покращує роботу мережі, після 20 епох навчання точність класифікації на навчальній вибірці наближається до 100%, а на валідаційній вибірці сягає 98%. Відповідні залежності наведено на рис. 4.

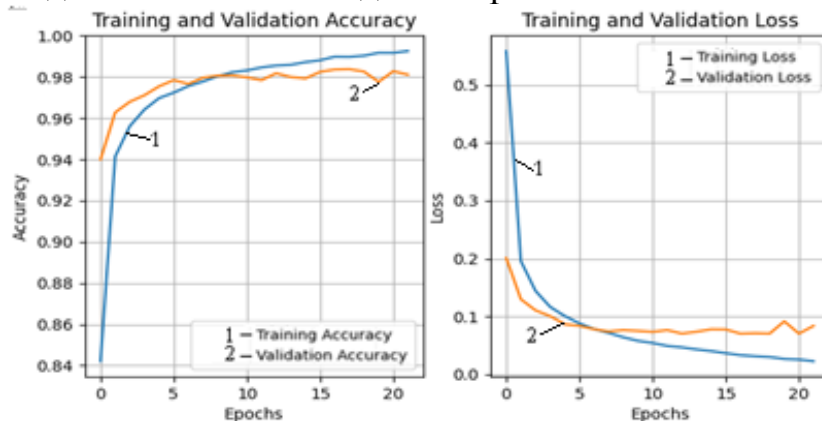


Рис. 4. $Cnv([5, 5], 6, «ReLU») > Cnv([3, 3], 50, «ReLU») > Flt() > Fcn(128, ReLU) > Fcn(10, Softmax)$

Для подолання проблеми перенавчання мережі було застосовано регуляризуючі процедури, а саме метод проріджування (dropout) [8] та L1, L2 регуляризацію. У табл. 1 наведено результати обчислювальних експериментів із застосуванням різних методів регуляризації. Контролювались кількість епох навчання, значення функції похибки та значення точності класифікації для валідаційної вибірки. З аналізу таблиці можна бачити, що до найкращих результатів призвело застосування L2 регуляризації; за меншу кількість епох навчання досягнуто точність класифікації 97,5% та значення похибки 0,1. Застосування L2 регуляризації має алгоритмічну складність, яка порівняна із стандартною метрикою похибки.

Таблиця 1

Застосування різних типів регуляризації до сталої мережі

Метод регуляризації (параметр методу)	Кількість епох до зупинки	Значення функції похибки на валідаційній вибірці	Значення точності класифікації для валідаційної вибірки
L2($\lambda=0.01$)	35	0,788	0,934
L2($\lambda=0.001$)	39	0,209	0,973
L2($\lambda=0,0001$)	19	0,109	0,975
L1($\lambda=0,01$)	65	0,273	0,959
L1($\lambda=0,001$)	42	0,161	0,970
L1($\lambda=0,0001$)	45	0,122	0,976
Dropout(10%)	35	0.07	0.972
Dropout(25%)	39	0.083	0.970
Dropout(40%)	36	0.115	0.965

З аналізу проведених експериментів обрано архітектуру мережі: $Cnv([5, 5], 6, \text{«ReLU»}) > Cnv([3, 3], 50, \text{«ReLU»}) > Flt() > Fcn(128, ReLU) > Fcn(10, Softmax)$ та рекомендовано застосовувати L2 регуляризацію під час навчання.

Узагальнюючий алгоритм налаштування мережі передбачає таку послідовність дій: за результатами аналізу класу задач, до якого застосовується нейронна мережа, необхідно обрати архітектуру мережі; зафіксувати базову мережу; шляхом виконання серії розрахунків визначити кількість епох навчання після виконання яких спостерігається перенавчання мережі; для визначення типу регуляризації необхідно застосувати доступні для даного класу мереж методи та обрати оптимальний; дослідити вплив використання різних функцій активації та параметрів шарів (розмір ядра згортки чи підвибірки); дослідити вплив кількості шарів та нейронів у шарах. Отриману фінальну архітектуру обрати у якості базової моделі та повторити кроки алгоритму.

Висновки. У роботі досліджено властивості та можливості згорткових нейронних мереж для розв'язання задач класифікації. За результатами аналізу літературних джерел досліджено стан проблеми, визначено основні архітектури сучасних згорткових нейронних мереж, та їх параметри. Розглянуто існуючі методи, застосування яких дозволяє протидіяти перенавчанню.

Побудовано математичну модель згорткової нейронної мережі, опис якої складається з опису згорткового та підвибіркового шарів, шару проріджуван-

ня та повнозв'язного шару. Обрано функції активації, наведено опис функції похибки з врахуванням регуляризуючих складових.

Програмно реалізовано модель згорткової нейронної мережі. За результатами проведених обчислювальних експериментів сформульовано рекомендації щодо вибору архітектури мережі, функцій активації та похибки, налаштовано параметри шарів мережі.

Бібліографічні посилання

1. **Тихонов, А. Н.** О регуляризации некорректно поставленных задач [Текст] / А. Н. Тихонов // Доклад. АН СССР – 1963.– №153(1) – С. 49–52.
2. **Baldi, P.** Understanding dropout [Text]/ P. Baldi P. J. Sadowski // Advances in Neural Information Processing Systems –2013.– №26 – С. 2814–2822.
3. **Gayer, A. V.** Convolutional neural network weights regularization via orthogonalization [Text] / Alexander V. Gayer and Alexander V. Sheshkus // International Conference on Machine Vision –2020. – №12 – С. 1143326-1 – 1143326-8
4. **Krizhevsky, A.** Imagenet classification with deep convolutional neural networks [Text] / I. Sutskever, and G. Hinton.// Advances in neural information processing systems, – 2012. – № 25(2) – С.1097–1105.
5. **Krogh, A.** A simple weight decay can improve generalization [Text] / A. Krogh, J. A. Hertz. // NIPS'91: International Conference on Neural Information Processing Systems –1991.– №4 , С. 950–957.
6. **Lecun, Y.** Gradient-based learning applied to document recognition [Text] / Y. Lecun, L. Bottou, Y. Bengio, P. Haffner // IEEE – 1998. – №. 86 (11), – С. 2278–2324.
7. **Singarimbun, R. N.** Adaptive Moment Estimation To Minimize Square Error In Backpropagation Algorithm [Text] / R. N. Singarimbun, E. B. Nababan and O.S. Sitompul, //International Conference of Computer Science and Information Technology – 2019. – С. 1-7.
8. **Srivastava, N.** Dropout: A simple way to prevent neural networks from overfitting [Text] / N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov // Journal of Machine Learning Research –2014.– №15 –С. 1929–1958.
9. **Takahashi, R.** Ricap: Random image cropping and patching data augmentation for deep cnns, [Text] / R. Takahashi, Takahashi, T. Matsubara, and K. Uehara // Proceedings of Machine Learning Research, – 2018. – № 95, – С. 786–798.
10. **Zhang, H.** Mixup: Beyond empirical risk minimization [Text]/ H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz // International Conference on Learning Representations – 2018. – не нашел страницы
11. **Zhong, Z.** Random erasing data augmentation [Text]/ L. Zheng, G. Kang, S. Li, and Y. Yang // AAAI Conference on Artificial Intelligence – 2020 – № 34(07) – С. – 13001-13008.

Надійшла до редколегії 16.10. 2020.