

С.І. ШАПОВАЛОВА

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна,
e-mail: lanashape@gmail.com.

Ю.В. МОСКАЛЕНКО

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна,
e-mail: yuramuv@gmail.com.

О.М. БАРАНІЧЕНКО

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна,
e-mail: baranichenko.work@gmail.com.

ЗБІЛЬШЕННЯ РЕЦЕПТИВНОГО ПОЛЯ НЕЙРОНІВ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ

Анотація. Проведено аналіз архітектур згорткових нейронних мереж для класифікації 1D- та 2D-сигналів. Визначено, що у випадку вхідного сигналу великої розмірності достатню точність класифікації можна забезпечити лише застосуванням великої кількості шарів, що не можна виконати за умови обмежених обчислювальних ресурсів. Однак, у разі обмеження кількості шарів, починаючи з деякого критичного значення розмірності, точність знижується. Запропоновано спосіб модифікації згорткової нейронної мережі з відносно невеликою кількістю шарів для розв'язання цієї проблеми. Експериментально доведено його ефективність.

Ключові слова: згорткові нейронні мережі, ResNet, EfficientNet, WaveNet, рецептивне поле.

ВСТУП

У статті поставлено задачу збільшення точності класифікації 1D- та 2D-сигналів згортковими нейронними мережами. У загальному випадку у разі збільшення розмірності сигналу, який подають на нейронну мережу, точність класифікації зростає. Однак для згорткових нейронних мереж з обмеженою кількістю шарів, починаючи з деякого критичного значення розмірності, цей принцип не справджується. Погіршення точності розпізнавання пов'язане з тим, що рецептивне поле (receptive field) нейронів останнього шару не «покриває» всього вхідного сигналу.

Запропоновано спосіб модифікації згорткових нейронних мереж з відносно невеликою кількістю шарів, який забезпечує достатню точність класифікації на вхідних сигналах великої розмірності. Це досягнуто за рахунок додаткових блоків агрегації поточного сигналу між стандартними шарами нейронної мережі. У статті представлено модифікацію архітектури згорткової нейронної мережі для розв'язання цієї проблеми.

Проведено обчислювальні експерименти та порівняльний аналіз розв'язання тестових задач класифікації 1D- та 2D-сигналів для мереж ResNet18, EfficientNet-b0, WaveNet та їхніх модифікацій. Експериментально доведено ефективність запропонованого способу.

ПОСТАНОВКА ПРОБЛЕМИ

Оброблення сигналів, зокрема зображень, у сучасних системах зазвичай здійснюють з використанням нейронних мереж. Для кожної такої задачі визначено математичний апарат, у якому враховано її специфіку. Класифікація зображень є поширеною прикладною задачею, яку розв'язують у таких галузях як медицина, астрономія, фізика, автоматизація керування тощо. Однак не завжди наявні алгоритми забезпечують достатню точність у конкретному

випадку. Наприклад, згорткові нейронні мережі (convolutional neural network, CNN), які добре зарекомендували себе під час аналізу зображень, не у всіх випадках забезпечують достатню точність у разі класифікації зображень великої розмірності.

Достатньої точності розв'язання цих задач можна досягти шляхом використання мереж глибокого навчання з великою кількістю шарів. Однак під час роботи такої мережі (в inference-режимі) потрібно застосувати значні комп'ютерні ресурси, які не завжди є доступними кінцевому користувачу. Тому підвищення точності класифікації сигналів великого розміру згортковою нейронною мережею з фіксованою кількістю шарів є актуальною задачею та має практичну значущість.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ

Сучасною базовою архітектурою мереж з великою кількістю шарів вважають ResNet [1] та її модифікації. Особливістю цієї мережі є наявність skip-з'єднань, які дають змогу навчати глибокі мережі.

Найпопулярніші модифікації архітектури ResNet є такими:

- ResNext [2], суть якої полягає у використанні групових згорткок;
- SeResNet [3], в якій використано механізми уваги;
- SeResNext [3], в якій синтезовано перші два підходи.

Концепції цих архітектур дають можливість підвищити ефективність за рахунок додавання шарів.

На відміну від них мережа EfficientNet [4] уможливорює масштабування як за глибиною (тобто за рахунок збільшення кількості шарів), так і за шириною (шляхом збільшення кількості вихідних каналів). Наступниками EfficientNet стали мережі EfficientNetV2 [5] та EfficientNetX [6], які дали змогу зменшити розмірність моделі і збільшити її точність. Це було досягнуто завдяки пошуку оптимальної архітектури на основі обчислювальних експериментів.

Зазначені мережі мають модифікації залежно від кількості шарів, наприклад, ResNet18, EfficientNet-b0. Ці модифікації добре зарекомендували себе як мережі з відносно невеликою кількістю шарів і, як наслідок, з достатньо швидким навчанням.

Ще однією сучасною модифікацією ResNet є мережа TRResNet [7]. Вона значно прискорює опрацювання зображень за рахунок зміни вузьких місць (bottlenecks) в архітектурі ResNet.

Мережу WaveNet [8] розроблено для генерації звукових сигналів. Її концепція полягає у застосуванні проріджених згорткок (dilated-convolution). Цю мережу можна застосовувати для розв'язання багатьох інших задач, зокрема задач класифікації, за умови заміни останніх шарів такими, що відповідають поточному класу задачі.

У статті [9] визначено та експериментально апробовано методи модифікацій базових архітектур згорткової нейронної мережі, а також їхніх комбінацій, які забезпечують максимальне підвищення точності класифікації та потребують незначних додаткових обчислювальних ресурсів.

Проблема полягає в тому, що реалізації мереж з кількістю шарів, що задовольняють вимоги щодо обмежених обчислювальних ресурсів, а саме мереж ResNet18, EfficientNet-b0, WaveNet, не забезпечують достатню точність класифікації на сигналах великої розмірності. У прикладних задачах намагаються збільшити розмірність вхідного сигналу мережі, оскільки це призводить до зростання точності класифікації.

Погіршення точності розпізнавання пов'язано з тим, що рецептивне поле нейронів останнього шару не «покриває» всього вхідного сигналу. Тому для розв'язання цієї проблеми потрібно модифікувати архітектуру згорткової нейронної мережі у такий спосіб, щоб можна було використовувати нейронні мережі з відносно невеликою кількістю шарів, забезпечуючи при цьому достатню точність класифікації на сигналах великої розмірності.

Метою роботи є розроблення способу збільшення рецептивного поля нейронів згорткових нейромереж для підвищення точності класифікації сигналу великої розмірності за умови обмежених обчислювальних ресурсів.

Для досягнення цієї мети потрібно розв'язати такі задачі:

- розробити додатковий блок згорткової нейромережі для збільшення рецептивного поля нейронів наступних шарів;
- модифікувати базові архітектури згорткових нейронних мереж, пристосованих для розв'язання задач класифікації;
- виконати обчислювальні експерименти з розв'язання задач класифікації за базовими та модифікованими архітектурами нейронних мереж.

БЛОК АГРЕГАЦІЇ СИГНАЛІВ РІЗНОЇ РОЗМІРНОСТІ

В основу концепції покладено ідею збільшення рецептивного поля нейрона. Звичайно цього досягають за рахунок збільшення кількості шарів нейронної мережі, пониження розмірності вхідного сигналу або використання проріджених згорток. У цій роботі запропоновано реалізувати в мережі глибокого навчання додатковий блок. Функцією цього блоку є пониження розмірності вхідного сигналу з подальшим застосуванням згортки. Оскільки після цих операцій вхідна та вихідна розмірності не збігаються, у блоці потрібно передбачити відновлення розмірності до початкової.

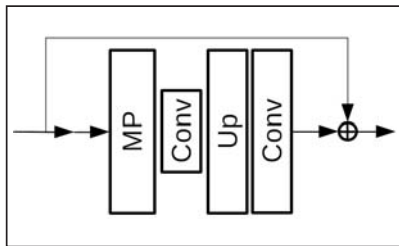


Рис. 1. Структура блоку агрегації сигналів різної розмірності (AggrBlock)

На рис. 1 зображено структуру додаткового блоку, який названо блоком агрегації сигналів різної розмірності — AggrBlock.

Запропонований блок містить такі компоненти:

- MP — шар MaxPooling, який понижує розмірність сигналу вдвічі.

У проведених дослідженнях також протестовано варіант заміни цього шару згорткою з кроком (stride) 2, але це рішення показало гірші результати;

- Conv — шари згорток відповідно до типу задачі 1D або 2D з розміром ядра 3, а також шари нормалізації та активації;
- Up — шар підвищення поточної розмірності сигналу до початкової;
- $\oplus$ — оператор додавання вхідного сигналу до вихідного.

Для того, щоб, по-перше, зберегти можливість використання попередньо навчених мереж, та, по-друге, не збільшувати кількість параметрів мережі, результат блоку не конкатенують з вхідним сигналом, а додають до нього. Інакше кажучи, можна зафіксувати ваги всієї мережі крім шару класифікатора та блоків агрегації і провести її навчання.

АРХІТЕКТУРА МОДИФІКОВАНОЇ МЕРЕЖІ

Запропоновано використовувати блоки агрегації сигналів різної розмірності AggrBlock між наявними стандартними блоками згорткової мережі. Стандартним блоком вважають сукупність шарів оброблення сигналу однієї

розмірності. Блоки агрегації суттєво збільшують рецептивне поле нейронів останніх шарів для сигналів великої розмірності, а отже, і можливості мережі.

Мережі типу ResNet можна схематично зобразити як вхідний шар і 4 стандартні блоки Layer_i , $i = 1..4$. Кожен стандартний блок містить певну кількість шарів згорток, активації та нормалізації. На рис. 2 представлено каркасну схему мережі ResNet із запропонованим блоком агрегації.

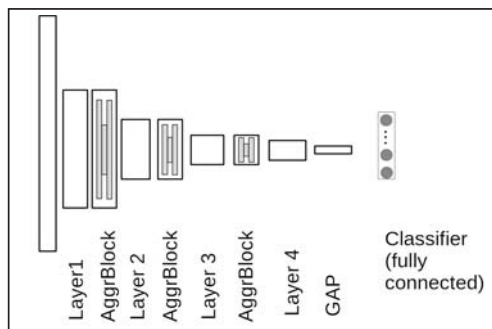


Рис. 2. Каркасна схема мережі ResNet

Кожен наступний блок обробляє вхідний сигнал меншої розмірності, що схематично зображено на рис. 2. Глобальний пулінг (GAP) зменшує розмірність сигналу до 1D шляхом агрегування. Останній компонент мережі є повноз'єднаним шаром, який безпосередньо здійснює класифікацію.

Аналогічну схему мають мережі архітектури EfficientNet.

ОБЧИСЛЮВАЛЬНІ ЕКСПЕРИМЕНТИ

Обчислювальні експерименти проведено для двох задач класифікації сигналів: числових рядів (1D-сигнал) та зображень (2D-сигнал).

Мережі для проведення експериментів. Метою обчислювальних експериментів був порівняльний аналіз результатів класифікації для восьми мереж, перші чотири з яких мають архітектури, що не містять запропонованого блоку, а останні чотири є такими самими архітектурами (з позначкою ++), з доданим блоком агрегації:

- 1) ResNet18;
- 2) EfficientNet-b0;
- 3) WaveNet;
- 4) WaveNetGRU;
- 5) ResNet18++;
- 6) EfficientNet-b0++;
- 7) WaveNet++;
- 8) WaveNetGRU++.

Каркасну схему мережі типу ResNet наведено на рис. 2. Базову реалізацію WaveNet запозичено з платформи Kaggle [10]. Оскільки тестовою задачею є аналіз послідовності сигналів, цю архітектуру доповнено рекурентними шарами, а саме блоком Bidirectional Gated Recurrent Unit (BiGRU). Цей блок обробляє вхідний сигнал паралельно з основною архітектурою. Вихід блоку конкатенується з виходом основної архітектури, після чого подається на кінцевий повноз'єднаний шар. Отриману мережу позначено WaveNetGRU.

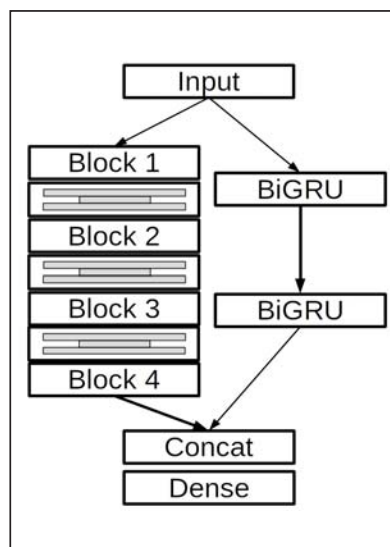


Рис. 3. Архітектура мереж WaveNetGRU++

На основі архітектур WaveNet і WaveNetGRU розроблено мережі WaveNet++ та WaveNetGRU++, які містять запропонований блок агрегації.

Схему WaveNetGRU++ представлено на рис. 3.

Мережу WaveNet++ сформовано аналогічно. Різниця полягає у відсутності BiGRU шарів.

Для обчислювальних експериментів на 1D-сигналах як базову архітектуру було вибрано WaveNet, для 2D-сигналів (зображень) — ResNet18, EfficientNet-b0, які є найпростішими модифікаціями мереж ResNet та EfficientNet.

Задача класифікації 1D-сигналів. Для цього типу класифікації тестовою задачею вибрано задачу визначення кількості відкритих іонних каналів на основі електрофізіологічних сигналів [11]. Задачу надано на платформі Kaggle Ліверпульським університетом.

Представлено приклади сигналу, кожен з яких складається з 500000 дискретних значень. Тренувальний датасет містить 10 прикладів даних, а тестовий — чотири. Кількість вихідних класів становить 11. Кількість відкритих іонних каналів визначали для кожного дискретного значення вхідного сигналу.

Оскільки вхідний сигнал має велику розмірність, для навчання він був поділений на проміжки по 4000 дискретних значень.

Метрикою оцінювання результатів класифікації визначено macro F1 score, яку оцінювали як

$$F_1 = 2 * \frac{precision * recall}{precision + recall}, \quad (1)$$

де *precision* — точність, яка визначається співвідношенням кількості правильно спрогнозованих прикладів, що належать даному класу, серед всіх прикладів, які класифікатор відніс до цього класу, *recall* — повнота, тобто частка правильно спрогнозованих класифікатором прикладів відносно всіх прикладів.

У свою чергу *precision* обчислюють за формулою:

$$precision = TP / (TP + FP), \quad (2)$$

де *TP* — True positive, *FP* — False positive.

Повноту *recall* обчислюють за формулою

$$recall = TP / (TP + FN), \quad (3)$$

де *FN* — False negative.

Було проведено початкову серію обчислювальних експериментів для визначення оптимальної функції втрат. Оскільки метрика точності змагання подібна до метрики Dice, до тестування додано метрики, характерні для задач сегментації. Проведено експерименти та проаналізовано результати розв'язання поточної задачі класифікації за такими функціями втрат:

- CCE (categorical cross entropy);
- Focal Loss [12];
- Tversky Loss [13];
- Dice Loss [14].

Результати проаналізовано як за кожною функцією окремо, так і за їхніми сукупностями. В результаті вибрано таку гібридну функцію втрат:

$$loss = CCE + FocalLoss + TverskiyLoss, \quad (4)$$

де *CCE* — categorical cross entropy loss; *FocalLoss* — функція втрат для важких прикладів; *TverskiyLoss* — функція втрат для незбалансованих прикладів.

Навчання проведено з використанням крос-валідації з розбиттям датасету на 5 частин. Для збільшення точності фінального результату (на відміну від тренування) в робочому режимі мережа передбачувала приклади по 4000, 8000, 16000 дискретних значень вхідного сигналу. Після передбачування здійснено узагальнення результату.

Таблиця 1. Результати розв’язання задачі “Identify the number of channels open at each time point”

#	Нейронна мережа	Метрика Macro F1 score		
		Local Cross-validation Score	Kaggle public Score	Kaggle private score
1	WaveNet	0.93866	0.94254	0.94105
2	WaveNetGRU	0.93921	0.94447	0.94512
3	WaveNet++	0.94085	0.94281	0.94406
4	WaveNetGRU++	0.94225	0.94481	0.94562

Запропоновані мережі брали участь у змаганні Kaggle. Результати їхньої участі наведено в табл. 1:

— Local Cross-validation Score — результат, отриманий на валідаційних частинах вибірок,

— Kaggle public Score — результат 30% тестової вибірки, який оголошували негайно,

— Kaggle private score — результат 70% тестової вибірки, який оголошували після завершення змагання і за яким проходило фінальне оцінювання результатів.

За результатами змагань з визначення кількості іонних каналів модель WaveNetGRU++ отримала 20 місце серед 2618 учасників, дала змогу одержати віртуальну срібну медаль і увійти в топ 1% учасників.

Задача класифікації 2D-сигналів. Тестовою задачею вибрано задачу класифікації за рисунком (скетчем) [15] користувачів сервісу Google Quick Draw. Тестові дані зібрано за допомогою цього сервісу, де користувачам пропонували намалювати нескладні об’єкти. Задачу надано на платформі Kaggle компанією Google. Вхідні дані є траєкторією руху миші, які перетворюються у зображення. Тренувальний набір даних містить 112163 приклади. Задля збільшення швидкості тестування кількість прикладів зменшено у 10 разів, тобто вона становить 11216.

Тестування проведено на зображеннях розмірності 768x768 пікселів.

Функцією втрат нейронної мережі вибрано Categorical Cross Entropy.

Метрикою оцінювання результатів класифікації платформою Kaggle визначено Mean Average Precision @3 (MAP@3). Ця функція отримує 3 найбільш ймовірні класи, які передбачила нейронна мережа. Якщо цільовий клас міститься на першому місці, то — виходом метрики буде 1, на другому — 1/2, на третьому — 1/3, в іншому випадку — 0.

Навчання проведено з розбиттям набору даних на тренувальний та валідаційний у співвідношенні 80% : 20%.

Результати тестування для мереж ResNet18 та EfficientNet-b0 наведено в табл. 2.

Отже, запропонований спосіб модифікації архітектур згорткових нейронних мереж дав змогу підвищити точність класифікації як для 1D- сигналу, так і для 2D-сигналу.

Таблиця 2. Результати розв’язання задачі “Quick, Draw! Doodle Recognition Challenge”

#	Нейронна мережа	Метрика MAP@3 score
1	ResNet18	0.9112
2	ResNet18++	0.9141
3	EfficientNet-b0	0.9167
4	EfficientNet-b0++	0.9174

ВИСНОВКИ

1. Запропоновано блок агрегації сигналів різної розмірності, який забезпечує достатнє збільшення рецептивного поля нейронів останнього шару для поточного сигналу.

2. На основі запропонованого блоку створено модифікації базових архітектур ResNet18, EfficientNet-b0, WaveNet, а також WaveNetGRU — архітектури WaveNet з використанням рекурентних шарів.

3. На основі обчислювальних експериментів доведено ефективність запропонованого способу модифікації згорткової нейронної мережі для підвищення точності класифікації сигналу великої розмірності за умови обмежених обчислювальних ресурсів для мереж ResNet18, EfficientNet-b0, WaveNet. У випадку 1D-сигналу на мережі WaveNet точність за метрикою Macro F1 Score підвищено з 0.94406 до 0.94562; у випадку 2D-сигналу (зображення) точність за метрикою MAP@3 підвищено з 0.9112 до 0.914.

СПИСОК ЛІТЕРАТУРИ

1. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (27–30 June 2016, Las Vegas, NV, USA). Las Vegas, 2016. P. 770–778. <https://doi.org/10.48550/arXiv.1512.03385>.
2. Xie S., Girshick R., Dollár P., Tu Z., He K. Aggregated residual transformations for deep neural networks. *Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (21–26 July 2017, Honolulu, HI, USA). Honolulu, 2017. P. 5987–5995. <https://doi.org/10.48550/arXiv.1611.05431>.
3. Hu J., Shen L., Sun G. Squeeze-and-excitation networks. *Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (18–23 June 2018, Salt Lake City, UT, USA). Salt Lake City, 2018. P. 7132–7141. <https://doi.org/10.48550/arxiv.1709.01507>.
4. Tan M., Le Q.V. EfficientNet: Rethinking model scaling for convolutional neural networks. *Proc. 36th International Conference on Machine Learning (ICML 2019)* (9–15 June 2019, Long Beach, CA, USA). Long Beach, 2019. P. 6105–6114. <https://doi.org/10.48550/arxiv.1905.11946>.
5. Tan M., Le Q.V. EfficientNetV2: Smaller models and faster training. *Proc. 38th International Conference on Machine Learning (ICML 2021)* (18–24 July 2021, virtual event). P. 10096–10106. <https://doi.org/10.48550/arxiv.2104.00298>.
6. Li S., Tan M., Pang R., Li A., Cheng L., Le Q., Jouppi N. Searching for fast model families on datacenter accelerators. *Proc. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)* (20–25 June 2021, Nashville, TN, USA). Nashville, 2021. P. 8081–8091. <https://doi.org/10.1109/CVPR46437.2021.00799>.
7. Ridnik T., Lawen H., Noy A., Ben E., Sharir B.G., Friedman I. TResNet: High performance GPU-dedicated architecture. *Proc. 2021 IEEE Winter Conference on Applications of Computer Vision (WACV)* (03–08 January 2021, Waikoloa, HI, USA). Waikoloa, 2021. P. 1399–1408. <https://doi.org/10.1109/WACV48630.2021.00144>.
8. Oord A., Dieleman S., Zen H., Simonyan K., Vinyals O., Graves A., Kalchbrenner N., Senior A., Kavukcuoglu K. Wavenet: A generative model for raw audio. 2016. P. 1–15. <https://doi.org/10.48550/arxiv.1609.03499>.
9. Shapovalova S., Moskalenko Y. Methods for increasing the classification accuracy based on modifications of the basic architecture of convolutional neural networks. *ScienceRise*. 2020. N 6 (71). P. 10–16. <https://doi.org/10.21303/2313-8416.2020.001550>.

10. Wavenet with SHIFTED-RFC Proba and CBR. *Kaggle/Code*. URL: <https://www.kaggle.com/nxrprime/wavenet-with-shifted-rfc-proba-and-cbr> (last accessed 08.10.2022).
11. University of Liverpool — Ion Switching. *Kaggle/Competitions*. URL: <https://www.kaggle.com/c/liverpool-ion-switching> (last accessed 08.10.2022).
12. Lin T.-Y., Goyal P., Girshick R., He K., Dollár P. Focal loss for dense object detection. *Proc. 2017 IEEE International Conference on Computer Vision (ICCV)* (22–29 October 2017, Venice, Italy). Venice, 2017. P. 2999–3007. <https://doi.org/10.1109/ICCV.2017.324>.
13. Salehi S.S.M., Erdogmus D., Gholipour A. Tversky loss function for image segmentation using 3D fully convolutional deep networks. *Proc. MLMI 2017*. In: *Lecture Notes in Computer Science*. Wang Q., Shi Y., Suk H.I., Suzuki K. (Eds). 2017. Vol. 10541. P. 379–387. https://doi.org/10.1007/978-3-319-67389-9_44.
14. Sørensen T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biologiske Skrifter/Kongelige Danske Videnskabernes Selskab*. 1948. Vol. V, N 4. P. 1–34.
15. Quick, Draw! Doodle Recognition Challenge. *Kaggle/Competitions*. URL: <https://www.kaggle.com/c/quickdraw-doodle-recognition> (last accessed 08.10.2022).

S. Shapovalova, Yu. Moskalenko, O. Baranichenko

INCREASING THE RECEPTIVE FIELD OF NEURONS OF CONVULSIONAL NEURAL NETWORKS

Abstract. The convolutional neural network architectures for the classification of 1D- and 2D-signals are analyzed. It has been determined that in the case of a large-dimensional input signal, sufficient classification accuracy can be ensured by only using a large number of layers, which cannot be done under limited computing resources. However, if the number of layers is limited, starting from some critical dimension value, the accuracy decreases. A method of modifying a convolutional neural network with a relatively small number of layers is proposed to solve this problem. Its effectiveness is proved experimentally.

Keywords: convolutional neural networks, ResNet, EfficientNet, WaveNet, receptive field.

Надійшла до редакції 25.10.2022