

# ПОКРАЩЕНИЙ МЕТОД АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА НА ОСНОВІ ЗГОРТКОВОЇ НЕЙРОМЕРЕЖІ SOP

<sup>1</sup>Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»;

<sup>2</sup>Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Норвезький університет науки та технології (NTNU)

*Аналіз акустичних сигналів у морському середовищі супроводжується численними викликами, такими як великий обсяг даних і швидка зміна умов навколишнього середовища. Під час пасивного прийому сигналів часто спостерігається низьке співвідношення сигнал/шум, що ускладнює їхню обробку. Для розв'язання цих задач потрібен ефективний і універсальний підхід, який можна забезпечити за допомогою нейронних мереж. Серед різних моделей нейромереж для аналізу акустичних сигналів одним із найефективніших є згорткові нейронні мережі. Тому є сенс у пошуку ефективного методу, оснований на CNN, який можна модифікувати для підвищення якості аналізу акустичних сигналів у водному середовищі.*

*Поліпшено метод, що базується на застосуванні нейромережі CNN SOP. Як вхідні характеристики для класифікації використано характеристики CQT. Оригінальні два шари згортки замінено багатомасштабною згорткою з різними розмірами ядер, що дозволило вилучити як загальні, так і локальні характеристики цільового об'єкта. У вилучених характеристиках відбувається зменшення розмірності шляхом обробки середнього пулінгу. Результат подається на вхід до шару пулінгу другого порядку (SOP), що дозволяє ефективно знаходити та зберегти інформацію про кореляції часових даних у формі узагальненого представлення. Виходом шару SOP є вектори, які нормалізуються поелементно квадратним коренем та l2-нормалізацією. Нормалізовані дані подаються на повнозв'язний шар з пакетною нормалізацією та функцією активації ReLU. Після цього дані передаються на повнозв'язний шар з функцією активації Softmax, яка виконує класифікацію. Для перевірки роботи нейромережі вибрано три датасети — два датасети реальних об'єктів водного середовища та один штучний датасет. Кожний датасет додатково зашумлено фоновими шумами таким чином, щоб підсумкові семпли мали низьке співвідношення сигнал/шум (SNR). У всіх випадках запропонований покращений метод показав вищу точність класифікації у порівнянні з оригінальним методом.*

**Ключові слова:** згорткові нейромережі, пулінг другого порядку, аналіз акустичних сигналів, багатомасштабна згортка.

## Вступ

Згорткові нейронні мережі (CNN) мають значні переваги перед іншими архітектурами завдяки своїй здатності автоматично виділяти найважливіші особливості даних, що зменшує потребу в ручному створенні ознак. Завдяки згортковим шарам CNN ефективно розпізнають просторові та частотні шаблони, а також мають меншу кількість параметрів, що дозволяє знизити обчислювальні витрати. У задачах аналізу акустичних сигналів у водному середовищі CNN особливо корисні для виділення характерних частотних компонентів, ігноруючи фоновий шум, що забезпечує високу точність класифікації навіть у складних умовах.

У роботі [1] запропоновано метод класифікації підводних сигналів на основі розрідженого анізотропного чирплетного перетворення (АСТ) і часово-частотної мережі (TFFNet). АСТ забезпечує високу роздільну здатність STFT, а TFFNet використовує багатомасштабні згортки для поліпшен-

ня агрегації ознак. На наборах даних UWA і Whale FM метод досягнув вищої точності класифікації, ніж STFT і традиційні CNN, з mAP до 90,6 % для більших зображень. Робота [2] описує метод класифікації кавітаційних шумів гвинтів з використанням DEMON-алгоритму і CNN. DEMON аналізує амплітудні варіації для поліпшення ідентифікації частот, а спектрограми служать вхідними даними для CNN. Архітектура включає шари нормалізації, згортки, максимального об'єднання і повністю з'єднані шари. Мережа досягла 90 % точності на реальних записах шумів. У роботі [3] запропоновано систему розпізнавання об'єктів на основі RBM-автокодера і нейронної мережі BP. Вхідними даними служать спектр потужності і демодуляції. RBM забезпечує вилучення ознак і вирішення проблеми обмеженого вибіркового обсягу, а BP виконує класифікацію. На датасеті ShipsEar середня точність досягла 92 %, перевершуючи DBM. У роботі [4] описано фреймворк глибокого навчання для класифікації сигналів на основі DeepShip, що містить записи 4 типів суден. Використано MFCC і лог-мел-спектрограми як ознаки. CNN з п'ятьма згортковими шарами досягла F-рахунку 78,92 %, значно перевершивши BP і традиційні методи машинного навчання. У роботі [5] запропоновано метод класифікації сигналів з використанням CNN у поєднанні з DWT. Перехід у вейвлет-область і нова архітектура забезпечують високу точність, стійкість до шумів і відсутність перенавчання. Результати показали перевагу методу над класичними CNN і BP. Особливої уваги заслуговує робота [6]. Ця робота представлена згортковою нейромережею CNN SOP. Модель CNN SOP включає об'єднання другого порядку (SOP) для фіксації часових кореляцій між характеристиками CNN, що дозволяє моделювати залежності вищого порядку, які мають вирішальне значення для підводних акустичних сигналів. На відміну від традиційних методів об'єднання, SOP зберігає відмінності між частотними піддіапазами, зберігаючи важливі частотно-специфічні характеристики для підвищення точності класифікації. Застосовуючи до результатів об'єднання підписаний квадратний корінь і нормалізацію L2, мережа досягає стабілізованого навчання і підвищеної надійності в різних умовах сигналу. Використання спектрограм з перетворення з постійним Q (CQT) забезпечує гарну частотну роздільну здатність на низьких частотах і часову роздільну здатність на високих частотах, що відповідає динаміці підводного сигналу.

*Метою роботи є підвищення точності класифікації акустичних сигналів водного середовища шляхом поліпшення архітектури згорткової моделі CNN SOP додаванням багатомасштабної згортки, середнього пулінгу та додаткового повнозв'язного шару.*

## Результати

Для виділення ознак у підводних нестационарних сигналах застосовується (CQT), що дозволяє перетворювати сигнал з часової області в часово-частотну область з високою частотною роздільною здатністю на низьких частотах, а це надає більше деталей порівняно з короткочасним перетворенням Фур'є (STFT). У цьому дослідженні CQT використовується для отримання часово-частотних ознак, які подаються на вхід запропонованої згорткової моделі. У цій роботі запропоновано поєднання багатомасштабної згортки та методу пулінгу другого порядку (SOP) для захоплення часових кореляцій в ознаках, що поліпшує можливості розрізнення різних цілей за допомогою вхідних даних CQT.

Для дискретного сигналу в часовому домені, постійне Q-перетворення (CQT) визначається таким чином:

$$X^{CQ}(k, n) = \sum_{j=n-\frac{N_k}{2}}^{n+\frac{N_k}{2}} x(j) a_k^* \left( j - n + \frac{N_k}{2} \right), \quad (1)$$

де  $k = 1, 2, \dots, K$  — кількість частотних бінів у CQT;  $a_k^*(n)$  — комплексно спряжена функція. Параметр  $N_k$  визначає змінну довжину вікна.

Центральна частота  $k$ -го біна розраховується за такою формулою:

$$f_k = f_1 2^{\frac{k-1}{B}}, \quad (2)$$

де  $f_1$  — позначає центральну частоту найнижчого біна, а  $B$  вказує на кількість бінів в октаві, що регулює співвідношення між часовою і частотною роздільною здатністю CQT.

Загальне число частотних бінів  $K$  у CQT можна обчислити за формулою

$$K = B \left( \log_2 \frac{f_{max}}{f_1} + 1 \right), \quad (3)$$

де  $f_{max}$  — центральна частота найвищого біна.

У цьому дослідженні пропонується застосувати CQT для отримання часово-частотних ознак. Ці ознаки обчислюються за формулою

$$X = \{X^1, X^2, \dots, X^N\}, \quad (4)$$

де 
$$X^i = 20 \log_{10} \|X^{CQ}(i)\|; \quad (5)$$

CQT — це ознака для фрейму і позначається як  $X^i$ , а  $N$  вказує на загальну кількість кадрів у сигналі. Комплексний вектор  $X^{CQ}(i) \in R$  містить значення CQT для  $K$  частотних бінів, що описують кадр фрейм  $i$ .

Оригінальна архітектура складається з вхідного шару, двох згорткових шарів, шару SOP, шару нормалізації та повнозв'язного шару з активаційною функцією Softmax.

Алгоритм роботи нейромережі:

– CNN генерує карти ознак, де кожна карта відповідає певній частотній смузі у CQT-вхідному зображенні. Для кожної карти ознак формується траєкторія часових значень у кожному з каналів.

– SOP використовує скалярний добуток (векторний добуток) між двома часовими траєкторіями для визначення кореляції між ними. Для кожного частотного біна обчислюється кореляційна матриця, яка є симетричною, позитивно напіввизначеною та відображає кореляції між усіма каналами CNN для цього біна.

– Кожний частотний бін (висота карти ознак) зберігає окремий результат SOP, тобто для кожної частоти створюється своя кореляційна матриця. Це дозволяє мережі вловлювати кореляції, зберігаючи частотні характеристики для ліпшого розрізнення класів.

– Отримані SOP-матриці перетворюються в один вектор, який потім нормалізується. Спочатку для кожного елемента обчислюється квадратний корінь, що знижує чутливість до великих значень. Далі застосовується l2-нормалізація, яка стандартизує вектор до фіксованої довжини.

– Після цього отриманий SOP-вектор передається через повнозв'язний шар до шару Softmax для визначення ймовірностей належності до класів.

До нейромережі додано шар багатомасштабної згортки, що використовує три паралельні фільтри різних розмірів ( $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ), об'єднуючи їхні результати через конкатенацію та нормалізацію. Це дозволяє виділяти ознаки на різних масштабах. Шар MultiScaleConv застосовує три згорткові фільтри з різними розмірами ядер ( $3 \times 3$ ,  $5 \times 5$  і  $7 \times 7$ ) паралельно, після чого над результатом виконується конкатенція та пакетна нормалізація. Кожна згортка застосовує ядро до вхідного тензора для вилучення ознак. Розмір ядра визначає область вхідних даних, яку обробляє фільтр, тоді як крок (stride) і заповнення (padding) визначають, як ядро переміщується по вхідних даних.

$$x_{ij} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} w_{mn} \cdot x_{i+m, j+n} + b, \quad (6)$$

де  $x_{ij}$  — вхідне значення на позиції  $(i, j)$ ;  $x_{i+m, j+n}$  — значення ядра (фільтра) згортки для фільтра розміром  $i+m, j+n$ ;  $w_{mn}$  — значення ядра (фільтра) згортки для фільтра розміром  $k \times k$ , де  $k \in 3, 5, 7$ ;  $b$  — зміщення (bias) для відповідного фільтра розміром  $k \times k$ , де  $k$  — розмір ядра згортки.

Наступний етап — конкатенція, що дозволяє об'єднати вилучені характеристики в один тензор, подається за формулою

$$x = [x1, x2, x3], \quad (7)$$

де  $x1, x2, x3$  — виходи згортки багатомасштабної згортки.

Об'єднані результати роботи згортки обробляються пакетною нормалізацією, за формулою

$$\hat{x}_{ij} = \frac{x_{ij} - \mu}{\sqrt{\sigma^2 + \epsilon}}, \quad (8)$$

де  $\mu$  — середнє значення,  $\sigma^2$  — дисперсія, а  $\epsilon$  — мала константа для стабільності.

Додатково додано середній пулінг, який зменшує кількість даних, фокусуючись на важливіших ознаках, що спрощує процес класифікації, формула якого така:

$$y_{ij} = \frac{1}{k^2} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x_{i+m, j+n}, \quad (9)$$

де  $x_{ij}$  — вхідне значення на позиції  $(i, j)$ ,  $y_{ij}$  — вихід після пулінгу,  $k$  — розмір вікна (наприклад,  $k = 2$ ).

Часо-частотне представлення акустичного шуму демонструє значні часові залежності, які можуть ефективно розрізнити об'єкти або джерела. Використано шар SOP, що дозволяє тонше вловлювати часові зв'язки, присутні у вхідних даних з (CQT), за формулою

$$SOP(s^j, s^k) = \sum_{i=1}^w s_i^j s_i^k = s^{j^T} s^k, \quad (10)$$

де  $SOP(s^j, s^k)$  — представляє часові кореляції між траєкторіями ознак  $s^j$  та  $s^k$  з  $j$ -ї та  $k$ -ї карт ознак.

Шар SOP захоплює взаємодії фільтрів згортки протягом часу. Для карт ознак використовується матриця часових ознак  $S \in R^{c \times w}$ , і оператор SOP в матричній формі:

$$SOP(S) = SS^T, \quad (11)$$

де  $\epsilon$   $SOP(S) = R^{c^2}$  симетричною позитивно визначеною матрицею, яка захоплює часові кореляції для однієї частотної смуги.

На відміну від інших стратегій пулінгу, які підсумовують кореляції по всьому зображенню, зберігаються результати SOP для всіх частотних смуг, щоб підтримувати частотні відмінності для класифікації. Кінцева ознака SOP складається з  $h$  операторів SOP, що відповідають висоті карт ознак.

Нормалізація значно поліпшує продуктивність глибоких мереж. Авторами застосовано квадратний корінь поелементно та нормалізацію до операторів SOP. Оператори SOP перетворюються у вектор  $p \in R^l$ , де  $l = c \times c \times h$ . Цей підсумковий вектор  $x$  обробляється із застосуванням квадратного кореня поелементно

$$x_{sr} = (\sqrt{x_1}, \sqrt{x_2}, \dots, \sqrt{x_n}), \quad (12)$$

де  $x_1, x_2, \dots, x_n$  — елементи вектора.

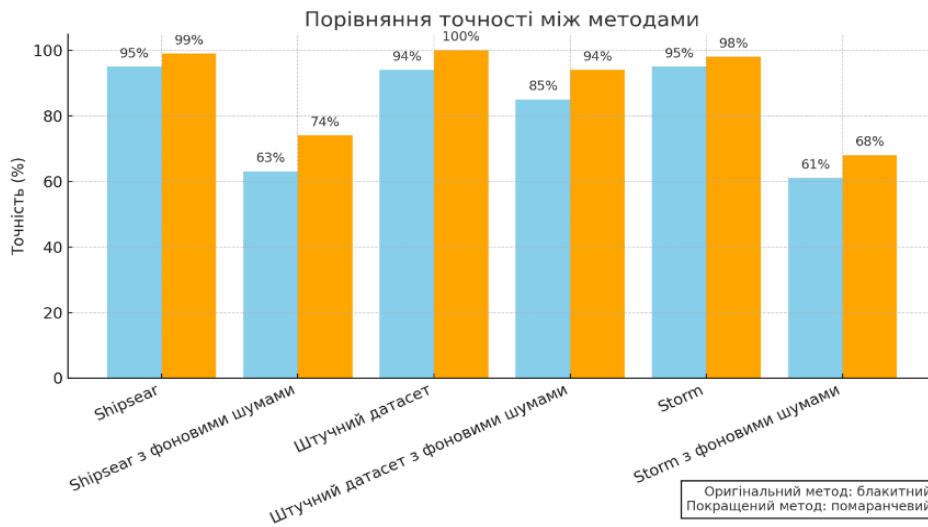
Також дані оброблено  $l_2$ -нормалізацією

$$x_{norm} = \frac{x}{\|x\|_2}, \quad (13)$$

Отриманий вектор об'єднання другого порядку  $z$  потім вводиться в шар softmax для класифікації після щільного шару.

Створено датасет на основі датасету ShipsEar [7]. Ці сигнали складаються із записів, що включають 4 типи кораблів і поділяються на 4 класи. Файли сегментовано на фрагменти розміром 2 секунди кожен. Також створено датасет з високим рівнем шуму, де середнє співвідношення сигнал-шум становило  $-10$ . На базі датасету Storm [8] створено аналогічний датасет. Він містить 5 класів кораблів. Штучний датасет є комбінацією декількох частот із декількома гармоніками різної амплітуди. Додатково створено штучний датасет з додаванням фонових шумів. Для штучного датасету створено 4 класи. Усі датасети розділені на набори для навчання, валідації та тестування у пропорції 70 %, 20 % та 10 %.

Оригінальний та поліпшений методи протестовано на вказаних вище трьох оригінальних датасетах та датасетах з додаванням фонових шумів. Результати показані на рисунку.



Результати тестування оригінальної та поліпшеної нейромережі, протестовані на різних датасетах

### Висновки

В результаті проведеного дослідження поліпшено метод класифікації акустичних сигналів водного середовища та отримано вищі результати класифікації у порівнянні з оригінальним методом. Запропонований метод ефективніше вилучає потрібні характеристики у вхідних даних, застосовуючи багатомасштабну згортку та вилучаючи характеристики різних масштабів. Застосування середнього пулінгу забезпечило зменшення розмірності вхідних даних. Додатково доданий шар з функцією активації ReLU та пакетною нормалізацією забезпечив поліпшення узагальнювальної можливості та гнучкості нейромережі. Удосконалена нейромережа показала вищі результати у порівнянні з оригінальним методом, зокрема — підвищення точності класифікації. В подальшому поліпшений метод планується застосувати для класифікації результатів нейромережі для придушення фонових шумів, що може підвищити точність класифікації датасетів з низьким співвідношенням сигнал-шум.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Y. Miao, Y. V. Zakharov, H. Sun, J. Li, and J. Wang, "Underwater Acoustic Signal Classification Based on Sparse Time-Frequency Representation and Deep Learning," *IEEE Journal of Oceanic Engineering*, vol. 46, no. 3, pp. 952-962, 2021. <https://doi.org/10.1109/JOE.2020.3039037>.
- [2] N. Bach, L. Vu, and V. Nguyen, "Classification of Surface Vehicle Propeller Cavitation Noise Using Spectrogram Processing in Combination with Convolution Neural Network," *Sensors*, vol. 21, p. 3353, 2021. <https://doi.org/10.3390/s21103353>.
- [3] L. Xinwei, Y. Feng, and M. Zhang, "An underwater acoustic target recognition method based on combined feature with automatic coding and reconstruction," *IEEE Access* 9, pp. 63841-63854, 2021. <https://doi.org/10.1109/ACCESS.2021.3075344>.
- [4] M. Ahmad, M. A. Ansari, R. Anwar, B. Shahzad, and A. Ikram, "Deep Learning Based Classification of Underwater Acoustic Signals," *Procedia Computer Science*, vol. 235, pp. 1115-1124, 2024. <https://doi.org/10.1016/j.procs.2024.04.106>.
- [5] K.-I. Kim, M.-I. Pak, B.-P. Chon, and C.-H. Ri, "A Method for Underwater Acoustic Signal Classification Using Convolutional Neural Network Combined with Discrete Wavelet Transform," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 19, 2021. <https://doi.org/10.1142/S0219691320500927>.
- [6] X. Cao, R. Togneri, X. Zhang, and Y. Yu, "Convolutional Neural Network With Second-Order Pooling for Underwater Target Classification," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 3058-3066, 2019. <https://doi.org/10.1109/JSEN.2018.2886368>.
- [7] D. Santos-Domínguez, S. Torres-Guijarro, A. Cardenal-Lopez, and A. Pena-Gimenez, "ShipsEar: An underwater vessel noise database," *Applied Acoustics*, vol. 113, pp. 64-69, 2016. <https://doi.org/10.1016/j.apacoust.2016.06.008>.
- [8] *Вимірювальні системи та програмне забезпечення для морських охоронних систем і дослідницьких полігонів*, звіт про НДР (заключ.) НТУУ «КПІ»; кер. роб. Є. Мачуський. Київ, 2012, 104 с. + відеокожух + CD-ROM. Д/б №2429-п.

Рекомендована кафедрою автоматизації та інтелектуальних інформаційних технологій ВНТУ

Стаття надійшла до редакції 14.12.2024

**Олексій Артур Олегович** — аспірант кафедри інженерії програмного забезпечення в енергетиці, e-mail: aoleksii@gmail.com.

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ;

**Верлань Андрій Анатолійович** — д-р. техн. наук, професор, професор кафедри інженерії програмного забезпечення в енергетиці, e-mail: verlandr.apeps@gmail.com .

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Норвезький університет науки та технології (NTNU)

**A. O. Oleksii<sup>1</sup>**  
**A. A. Verlan<sup>2</sup>**

## **Improved Method for Analyzing Acoustic Signals of Water Environment Based on Convolutional Neural Network SOP**

<sup>1</sup>National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”;

<sup>2</sup>National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”,  
Norwegian University of Science and Technology

*The analysis of acoustic signals in marine environments poses numerous challenges, including handling large volumes of data and adapting to rapidly changing environmental conditions. Passive signal reception is often characterized by a low signal-to-noise ratio (SNR), complicating signal processing. Addressing these challenges requires an efficient and versatile approach, which neural networks can provide. Among the various neural network models for acoustic signal analysis, Convolutional Neural Networks (CNNs) are recognized as some of the most effective due to their ability to handle complex data structures and extract meaningful features. Therefore, it is reasonable to explore an efficient CNN-based method that can be modified to enhance the quality of acoustic signal analysis in aquatic environments. An improved approach has been developed, utilizing the CNN SOP network. CQT features were employed as input characteristics for classification. The original two convolutional layers were replaced with a multi-scale convolution using kernels of different sizes, enabling the extraction of both global and local features of the target object, thereby enhancing the model's ability to process diverse signal properties. The extracted features were processed using second-order pooling and then passed to the second-order pooling (SOP) layer. This layer enhances the identification of temporal correlations. The SOP layer outputs feature vectors, which are normalized element-wise using square root and l2 normalization. The normalized data are then fed into a fully connected layer with batch normalization and the ReLU activation function. Subsequently, the data are passed to another fully connected layer with a Softmax activation function, which performs the final classification. To evaluate the performance of the neural network, three datasets were used: two based on real-world underwater objects and one artificial dataset. Each dataset was further augmented with background noise to produce samples with a low signal-to-noise ratio (SNR). In all cases, the proposed improved method demonstrated higher classification accuracy compared to the original method, showcasing its effectiveness in handling noisy data and improving signal analysis.*

**Keywords:** Convolutional neural networks, second-order pooling, acoustic signal analysis, multiscale convolution

**Oleksii Artur O.** — Post-Graduate Student of the Chair of Software Engineering for Power Industry, e-mail: aoleksii@gmail.com ;

**Verlan Andrii A.** — Dr. Sc. (Eng), Professor, Professor of the Chair of Software Engineering for Power Industry, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Associate Professor of the IVB, Norwegian University of Science and Technology, e-mail: verlandr.apeps@gmail.com