УДК 004.896

Поляков А.С.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

Корнага Я.І.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ВИДІЛЕННЯ ЗВУКУ ПТАХА В ЖИВІЙ ПРИРОДІ ЗА ДОПОМОГОЮ НЕЙРОННОЇ МЕРЕЖІ

У статті висвітлюються практичні аспекти використання нейронної мережі для виділення звуку птахів у живій природі за допомогою нейронної мережі. Зокрема, навчання нейронної мережі за допомогою потрібних даних. У статті буде розглянуто, як подається інформація для навчання нейронної мережі, а також як нейронна мережа навчається. Також ми розглянемо попередні актуальні системи виділення звуку птаха у живій природі за допомогою нейронної мережі. Наостанок у статті будуть показані результати експерименту.

Ключові слова: *нейронна мережа, спектрограма, навчання, функція, алгоритм.*

Постановка проблеми. Важливою проблемою в екології є вивчення взаємодії між організмами та їх оточенням, що полягає в тому, щоб стежити за тваринними популяціями, зважаючи на постійну загрозу зміни клімату. Використання акустики для моніторингу та класифікації птахів у природних умовах останнім часом викликало великий інтерес.

Класифікація видів птахів на основі записаних звукових даних є, наприклад, корисною під час моніторингу поведінки розмноження, біорізноманіття та динаміки популяції. Птахи є особливо корисним екологічним індикатором, оскільки вони швидко реагують на зміни в їхньому середовищі. Класифікація птахів може здійснюватися вручну спеціалістами домену; однак зі зростанням кількості даних це швидко стає нудним і трудомістким процесом. Тому необхідні автоматичні інструменти, які можуть допомогти у цьому процесі.

Аналіз останніх досліджень і публікацій. Bird Classification Challenges

Кілька видів змагань щодо класифікації птахів, які були проведені протягом останніх кількох років, є тісно пов'язаними, але мали різні описи завдань. Інтерес та участь у цих змаганнях були високими, що вказує на те, що це важливі проблеми та що їх необхідно вирішити.

Змагання, як правило, передбачають, які види присутні в наборі записів із прихованими мітками, які називаються тестовим набором, а також для подання прогнозованого виду для кожної точки тестування для оцінки з нанесенням ілюстрації на землю.

Опис завдання може відрізнятися від прогнозування лише наявністю або відсутністю птахів у записі для прогнозування всіх активно співаючих видів птахів. Тобто проблеми мають різні ступені складності. Ми розглянемо деякі змагання.

MLSP 2013. The IEEE International Workshop on Machine Learning for Signal Processing (MLSP) оголосила про змагання категорії ідентифікації видів птахів у 2013 році [1]. Завдання було визначити всі акустично активні види птахів у кожному аудіозаписі тестового набору з 19 різними видами птахів. Тобто завдання розглядаються як одноразові проблеми з кількома мітками.

Набір даних складався з 645 десятисекундних аудіозаписів, які були розділені на тренувальний набір (50%) та тестовий набір (50%). Мітки птахів для кожного запису в наборі тренувань були оприлюднені, але етикетки для кожного запису в тестовому наборі були збережені в секреті.

Команда-переможець використовувала random forest (RF) classifier, де функції було витягнуто з вводу за допомогою відповідності шаблону. Шаблони обчислювали за допомогою спеціальної методики сегментації часової частоти, де кожний сегмент зберігався як шаблон і обчислювався лише з 81 аудіозапису, що були помічені одним класом звуку. Потім розраховували спектрограму кожного запису.

Функції були витягнуті для кожної спектрограми, обчислюючи нормалізовану крос-

кореляційну карту між спектрограмою та кожним шаблоном, подібність між шаблоном і спектрограмою була потім оцінена на максимальному значенні нормалізованої крос-кореляційної карти з використанням методу відповідності шаблону, що означає, що кожна спектрограма дає вектор функції з тією ж довжиною, що і число шаблонів, які було витягнуто з 81 запису міток сигналу. Метод також використовував функції, доступні як базова лінія у виклику, такі як гістограма сегментів, які були додані до вектора функцій і використовувались як вхідні дані для класифікатора.

Багато команд розробили спеціальні функції, проте одна команда, яка посіла четверте місце, використовувала необроблені дані спектрограми для навчання згорткової нейронної мережи. Після цього було зазначено, що подальше дослідження використання згорткових нейронних мереж у цій проблемній галузі є виправданим.

NIPS4B 2013. У Neural Information Processing Scaled for Bioacoustics (NIPS4B) опис завдання був подібний до опису завдання MLSP 2013. Учасникам було запропоновано ідентифікувати всіх активно співаючих птахів у кожному з тестових файлів. Однак число можливих видів становило 87 замість 19, а записи могли змінюватися в довжину (від 0,5 до 5,5 с).

Переможець цього змагання використовував подібний [9] підхід, як в останньому. Головна відмінність полягала в тому, що для кожного аудіофайлу витягуються додаткові функції. Окрім функцій, витягнутих шляхом оцінки відповідності шаблону за максимальним значенням нормалізованої крос-кореляційної карти, вектор функції ще більше доповнюється статистикою файлу та сегмента (наприклад, середнє значення, стандартне відхилення тощо).

BirdCLEF 2016. Завдання BirdCLEF використовувало дуже великий набір даних із записом птахів. Набір даних BirdCLEF [2] є підмножиною бази даних і складається з 999 різних видів, зареєстрованих у Південній Америці. Набір даних складається приблизно з 33 200 записів, які були нормалізовані до 44,1 кГц 16-розрядний моноформатний формат аудіофайлів.

Завдання полягало у визначенні видів птахів, присутніх у кожному записі. Учасникам було запропоновано надати список найуспішніших птахів для записів прихованого тестового набору. Набір даних поділений на 1/3 даних тесту та 2/3 навчальних даних, а показник, який використовується для оцінки ефективності класифікації тестів, що поставляються командами-учасниками, є середньою точністю (MAP) над усіма записами в тестовому наборі.

Найкращий метод [3] використав метод згорткової нейронної мережі, де вхід до мережі був сегментами спектрограми, обчисленої зі звукових файлів. Звукові файли попередньо обробляються шляхом вилучення двох звукових класів із кожного аудіофайлу: шум та сигнал (пташиний спів), які поділяються на однаково довгі звукові сегменти приблизно на 3 секунди. Сегменти сигналу являють собою фактичний пташиний спів, кожен з яких має пов'язані з ним види птахів. Потім зразки, показані нейронній мережі, завантажуються та розширюються випадковим чином. Кожен сегмент сигналу додатково поєднується з іншим сегментом сигналу одного класу, вибраним у випадковому порядку, а також трьома випадковими сегментами шуму. Зразки потім додатково збільшуються випадковим зрушенням у часовій сфері та невеликим випадковим зсувом (5%) у частотній сфері.

З усього цього у програмі MLSP 2013 рішення, що виграли, були random forests, що навчалися з ймовірностей, отриманих на основі зіставлення шаблонів специфічних для видових спектрограм [10]. Переможець NIPS4B 2013 року використовував ці результати як вихідну точку, але ввів додатковий набір функцій, статистично виведених з аудіофайлів. Лассек [8] також використовував подібний метод, щоб виграти BirdCLEF 2015 завдання.

Проте під час BirdCLEF 2016 було показано, що згорткові нейронні мережі, навчені спектральних даних, обчислених із звукозаписів, можуть перевершити інші сучасні системи. Ця теза використовує роботу Спренгеля [3] як відправну точку та базову лінію, а також досліджує використання нового методу згорткової нейронної мережі, що називається глибинними залишковими нейронними мережами, а також нової методики збільшення даних, яка називається збільшенням дельта-даних частот з множинною шириною.

Постановка завдання. Потрібно навчити нейромережу виділяти відповідні спектри голосу птахів з попередньою підготовкою звуку.

Виклад основного матеріалу дослідження. Необроблені звукові дані не підходять для введення нейронної мережі, тому аудіосигнал зазвичай перетворюється на тимчасове спектральне подання.

Спектрограма дискретного звукового сигналу $\bar{x} = x_1, \ldots, x_n$ обчислюється в два або три кроки. По-перше, Short-Time Fourier Transform (коротке

тимчасове перетворення Фур'є) (STFT) застосовується до аудіосигналу. STFT розраховується стандартним способом, розділяючи сигнал [4] на різні фрейми, що перекриваються, а потім обчислюють Discrete Time Fourier Transform (DTFT) для кожного кадру, що призводить до матриці зі складними значеннями:

$$STFT\left\{\bar{x}\right\}(m,\omega) \equiv X_m(\omega) = \sum_{n=-\infty}^{\infty} x_n w(n - mR) e^{-j\omega n},$$

де x_n – вхідний сигнал у момент часу n, $w(n)$ – довжина M = 512 вікно хеннінга орієнтовно навколо n, а R = 128 являє собою розмір переходу між послідовними кадрами. Ми використовуємо librosa.stft метод бібліотеки librosa для обчислення STFT. По-друге, обчислюється квадратна амплітуда величини STFT, яку ми називаємо ampspectrogram, і, по-третє, натуральний логарифм амплітудної спектрограми обчислюється, яку ми називаємо logspectrogram.

$$ampspectrogram\left\{\bar{x}\right\}(\omega) \equiv \left|\bar{X}(\omega)\right| \equiv \left|\bar{X}(\omega)\right|^2$$

$$Logspectrogram\left\{\bar{x}\right\}(\omega) \equiv \log_e\left(\left|\bar{X}(\omega)\right|^2\right)$$

Виявлення типу звуку птаха

Аудіофайли спочатку переробляються у формат, який можна використовувати для навчання нейронної мережі. Аудіофайли нормалізуються до 16-бітних даних моноканальної хвилі, повторно відібрані з 44100 Гц до 22.050 Гц.

Виявлення пісні птаху

Записи пташиних пісень розділені на два різні звукові класи: сигнал (птах вокал) та шум. Розділення дає змогу набути нейронну мережу за найбільш релевантними даними, і це дає нам доступ до класу шумів, який може бути використаний для покращення навчальних зразків.

Після того, як запис був розділений [5] на сигнальну хвилю і шумову хвилю, кожен розділяється на 3-секундні сегменти, які зберігаються на диску. Сегменти шуму можуть пізніше бути використані для посилення навчальних зразків, що показуються в мережі, які мають поліпшити узагальнення. Частина сигналу витягується спочатку, обчислюючи маску сигналу $\bar{v}$ для заданої звукової хвилі $\bar{x}$, а потім, використовуючи маску для витягання відповідної частини звукової хвилі, де $v_i = 0$ вказує, що x_i не є частиною сигналу, і $v_i = 1$ вказує, що це частина сигналу. Маска походить від бінарного зображення, яке обчислюється шляхом аналізу нормалізованої амплітудної спектрограми $\bar{x}$. Нехай $\bar{b}$ є бінарним зображенням, нехай $\bar{s}$ буде нормованою амплітудною спектрограмою $\bar{x}$, де $\bar{b}$ та $\bar{s}$ мають однакові розміри. Піксель за індексом (i, j) у бінарному зображенні потім встановлюється в один, якщо $s_j^{(i)}$ у t разів перевищує межу рядка та стовпця $\bar{s}$ у рядку і та стовпці j

$$b_j^{(i)} = \begin{cases} 1, & if\ s_j^{(i)} > t \times median\left(\bar{s}^{(i)}\right) \wedge s_j^{(i)} > t \times median\left(\bar{s}_j\right) \\ 0, & otherwise \end{cases}.$$

Двійкове зображення обробляється [6] додатково, застосовуючи бінарну ерозію, а потім двійкове розширення на зображенні, обидва з використанням розміру ядра від 4 до 4, що згладжує області, позначені як пташиний вокал, а маска сигналу $\bar{v}$ походить від бінарного зображення, встановлюючи v_j до одного, якщо стовпчик $\bar{b}_j$ містить один. Маску також згладжують шляхом виконання ще двох двійкових розбавлень (розмір ядра 4), а потім повторно треба масштабувати так, що $\left|\bar{v}\right| = \left|\bar{x}\right|$. Зображення на малюнку 2.1 були вручну порівняні для кожного етапу з відповідним зображенням, представленим Спренгелем для одного звукового файлу, і метод видає подібні результати.

Алгоритм Обчислення маски сигналу

1: procedure ComputeMask $(\bar{r}, t)$
2: P xx ← spectrogram $(\bar{r})$
3: P xx ← normalize (Pxx)
4: BinaryImage ← medianClipping (Pxx, t)
5: BinaryImage ← erosion (BinaryImage,(4, 4))
6: BinaryImage ← dilation(BinaryImage,(4, 4))
7: mask ← computeM ask(BinaryImage)
return mask

Маска сигналу обчислюється шляхом встановлення порога t = 3, а шум обчислюється шляхом встановлення t = 2,5, а потім інвертує маску наприкінці (фліпінг 0c-1c, а 1c-0c). Це може залишити частину хвилі, яка позначена як не сигнал і не шум (2.5-3). Вважається, що ці частини не сприяють випуску будь-якої відповідної інформації для мережі, їх просто ігнорують.

Експеримент. У експерименті було використано 18-шарову залишкову нейронну мережу, яка пройшла навчання на 120 епох, історія тренування якої показана на рисунках. На рисунку 1 показана зміна тренувань та втрат валідації.

А також зміна точності навчання та перевірки щодо навчальної епохи показана на рисунку 2.

На рисунку 3 показано кількість навчальних сегментів (синіх), зображених на правій осі у відношенні до 5% шматочків класів звуку, класифікованих за кількістю тренувальних сегментів у кожному 5-відсотковому шматочку. Він також показує середню кількість прогнозів, які отримують шматочки звукових класів (червоний колір) та очікувану середню кількість прогнозів для кож-

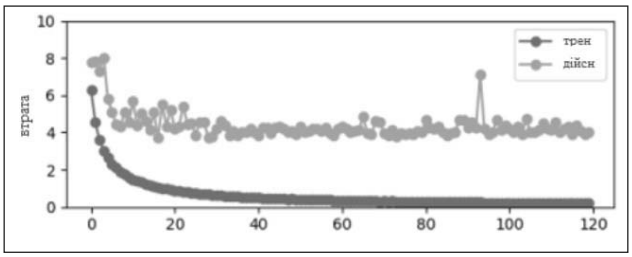

Рис. 1. Зміна тренувань та втрат валідації

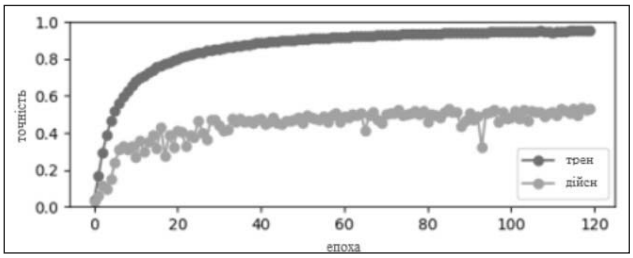

Рис. 2. Зміна точності навчання та перевірки щодо навчальної епохи

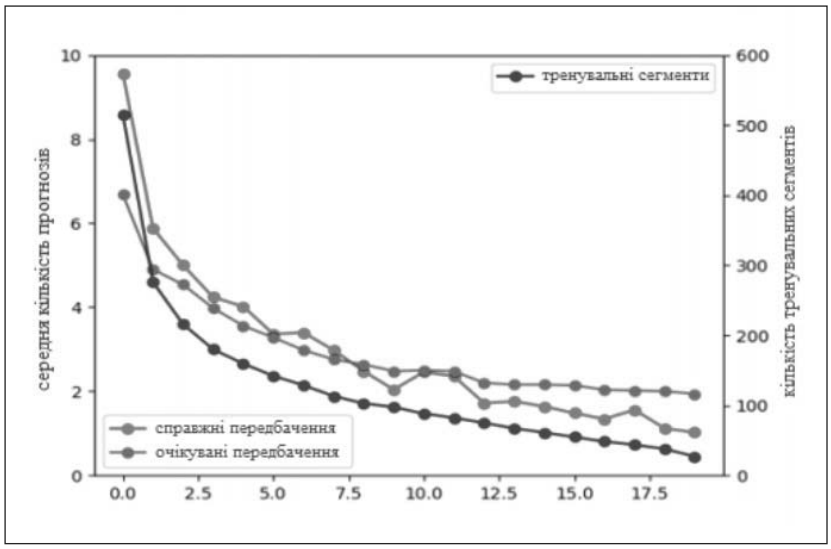

Рис. 3. Передбачені та тренувальні сегменти

ного шматка (зеленого кольору), нанесеного на ліву вісь У.

Висновок. Нейронна мережа, яка може розпізнавати види живих птахів, дуже корисна для ефективного отримання даних у сфері екології, але для того щоб нейронна мережа могла бути корисною, спочатку треба її навчити, як відрізняти звуки. Для цього треба обробити дані на подання навчання для нейромережі. Тому якщо не надати нейронній мережі досить даних для навчання, ми не будемо завжди отримувати правильні відповіді.

Список літератури:

1. Sergey Zagoruyko and Nikos Komodakis. Wide Residual Networks. Arxiv, 2016.

2. Dan Stowell, Mike Wood, Yannis Stylianou, and Hervé Glotin. Bird detection in audio: a survey and a challenge. 2016.

3. Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. Audio Based Bird Species Identification using Deep Learning Techniques. 2016

4. Jaderick P. Pabico, Anne Muriel V. Gonzales, Mariann Jocel S. Villanueva, and Arlene a. Mendoza. Automatic identification of animal breeds and species using bioacoustics and artificial neural networks. arXiv preprint, pages 1–17, 2015.

5. Peter Jancovic and Munevver Kokuer. Acoustic recognition of multiple bird species based on penalised maximum likelihood. IEEE Signal Processing Letters, 22(10):1–1, 2015.

6. Jason Wimmer, Michael Towsey, Paul Roe, and Ian Williamson. Sampling environmental acoustic recordings to determine bird species richness. Ecological Applications, 23(6):1419–1428, 9 2013.

7. Alexis Joly, Herve Goeau, Herve Glotin, Concetto Spampinato, Pierre Bonnet, Willem-Pier Vellinga, Robert Planque, Andreas Rauber, Robert Fisher, and Henning Müller. LifeCLEF 2014: Multimedia Life Species Identification Challenges. (ii):229–249, 2014.

8. Mario Lasseck. Bird song classification in field recordings: Winning solution for NIPS4B 2013 competition. Proc. of int. symp. Neural Information Scaled . . ., pages 1–6, 2013.

9. Herve Goeau, Herve Glotin, Willem Pier Vellinga, Robert Planque, Andreas Rauber, and Alexis Joly. LifeCLEF bird identification task 2015. CEUR Workshop Proceedings, 1391, 2015.

10. Yann LeCun, Yoshua Bengio, and Hinton Geoffrey. Deep learning. Nature Methods, 13(1):35–35, 2015.

ВЫДЕЛЕНИЕ ЗВУКА ПТИЦЫ В ЖИВОЙ ПРИРОДЕ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

В статье освещаются практические аспекты использования нейронной сети для выделения звука птиц в живой природе с помощью нейронной сети. В частности, обучение нейронной сети с помощью нужных данных. В статье будет рассматриваться, как подается информация для обучения нейронной сети, а также как нейронная сеть обучается. Также мы рассмотрим предыдущие актуальные системы выделения звука птицы в живой природе с помощью нейронной сети. В конце статьи будут показаны результаты эксперимента.

Ключевые слова: *нейронная сеть, спектрограмма, обучение, функция, алгоритм.*

ALLOCATION OF THE SOUND OF BIRDS IN WILDLIFE THROUGH THE NEURAL NETWORK

The article covers practical aspects of using the neural network for the allocation of the sound of birds in wildlife through the neural network. In particular, training the neural network with the necessary data. The article will consider how to provide information for training the neural network and also how the neural network learns. Also, we will consider the previous up-to-date systems for the allocation of bird's sound in wildlife through the neural network. The last one in the article will show the results of the experiment.

Key words: *neural network, spectrograph, training, function, algorithm.*