

**Ілюшик Т.С.**

*студент,*

*Національний технічний університет України*

*«Київський політехнічний інститут імені Ігоря Сікорського»*

## **ОПТИМІЗАЦІЯ НЕЙРОННИХ МЕРЕЖ ДЛЯ КЛАСИФІКАЦІЇ ЗОБРАЖЕНЬ РАКУ ПРОСТАТИ**

Глибокі нейронні мережі досягли великого прогресу за останнє десятиліття в кількох завданнях машинного навчання, таких як комп'ютерний зір [1]. Не виключенням є задачі класифікації зображень. У випадку класифікації ракових утворень ми маємо справу з високоякісними цифровими зображеннями біопсії. У телемедицині ці зображення називаються whole slide images (WSI). У зв'язку з високою роздільною здатністю зображення можуть досягати до 1 Гб пам'яті. Більш того для побудови точної нейронної мережі потрібно тисячі таких зображень. Тому виникає проблема в оптимізації алгоритму нейронної мережі для прискорення процесу навчання. Однією з найвідоміших методик оптимізації є пакетна нормалізація (batch normalization) [2].

У цій роботі ми визначимо вплив пакетної нормалізації на швидкість та ефективність роботи нейронної мережі. Пакетна нормалізація – це операція, яка може бути введена між шарами для нормалізації виводу кожного шару, і було показано, що вона надзвичайно ефективна для стабілізації та прискорення навчання глибоких нейронних мереж. Отже, сьогодні вона є стандартом у багатьох сучасних архітектурах.

Перш ніж переходити до пакетної нормалізації, давайте розберемося з терміном нормалізація. Нормалізація – це інструмент попередньої обробки даних, який використовується для приведення числових даних до загального масштабу без спотворення їх форми. Як правило, коли ми вводимо дані до алгоритму глибокого навчання, ми схильні змінювати значення на збалансовану шкалу. Причина, чому ми нормалізуємо, частково полягає в тому, щоб переконатися, що наша модель може належним чином передбачати мітки даних. Повертаючись до пакетної нормалізації, це процес, щоб зробити нейронні мережі швидшими та стабільнішими за допомогою додавання додаткових шарів у глибоку нейронну мережу. Новий шар виконує операції стандартизації та нормалізації над входом шару, що надходить із попереднього шару. Типова нейронна мережа навчається за допомогою зібраного набору вхідних даних, який називається batch (пакет).

Пакетна нормалізація це двохетпний процес. Спочатку нормалізуються дані які подаються на вхід, а пізніше виконується масштабування та зміщення. На першому кроці ми маємо batch, який подається на вхід із шару  $h$ , спочатку нам потрібно обчислити середнє значення цієї прихованої активації. Після того як ми обчислили середнє значення нам потрібно знайти стандартне відхилення для цього шару активації. Далі, оскільки у нас готові середнє значення і стандартне відхилення. За допомогою цих значень ми нормалізуємо приховані активації. Для цього ми віднімемо середнє з кожного введення і розділимо ціле значення на суму стандартного відхилення та коефіцієнта згладжування. Коефіцієнт згладжування забезпечує чисельну стабільність операції, уникаючи поділ на нульове значення. В результаті отримуємо функцію нормалізації за формулою (1).

$$h_{i(norm)} = \frac{(h_i - \mu)}{\sigma + \varepsilon} \quad (1)$$

де  $\varepsilon$  – коефіцієнт згладжування,  $\mu$  – середнє значення,  $\sigma$  – стандартне відхилення,  $h_i$  – вхідний батч.

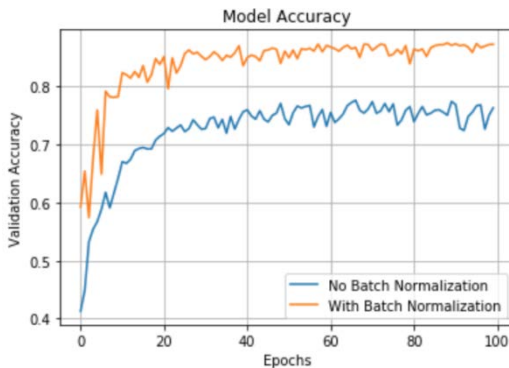
Під час останнього кроку відбувається повторне масштабування та зміщення вхідного пакету. Тут впливають два компоненти алгоритму batch normalization,  $\gamma$  (гамма) і  $\beta$  (бета). Ці параметри використовуються

для повторного масштабування ( $\gamma$ ) і зсуву ( $\beta$ ) вектора, що містить значення з попередніх операцій.

Ці два параметри під час навчання нейронна мережі коригуються забезпечуючи використання оптимальних значень  $\gamma$  і  $\beta$ . Це дозволить точно нормалізувати кожен батч.

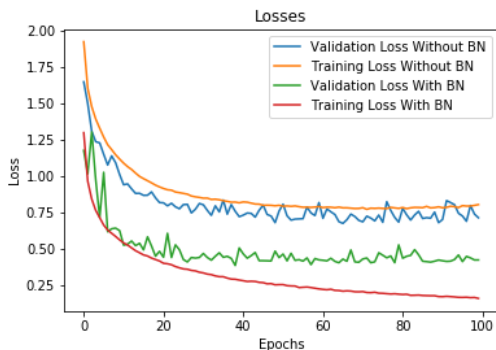
За основу для класифікації зображень раку простати було обрано архітектуру VGG16 з ваговими коефіцієнтами, попередньо натренованими в Imagenet для вилучення ознак. На етапі навчання модель тренувалася протягом 100 епох за допомогою оптимізатора SGD. Швидкість навчання ( $\eta$ ) була ініціалізована як  $1 \cdot 10^{-2}$ , значення якої зменшується в 10 раз після половини ітерацій. Функцією активації було обрано ReLU. Запропонована модель була виконана за допомогою мови програмування Python та бібліотеки Pytorch.

Зазвичай шари з batch normalization застосовуються перед функціями активації, проте деякі останні дослідження показують, що застосування нормалізації після функції активації може бути ефективніше [3]. Ми побудували 2 моделі, перша використовує batch normalization інша ні. За допомогою метрик для оцінки точності класифікації ми отримали результати цих моделей на тестовій вибірці на рис. 1. Більш того ми оцінили динаміку зміни значення функції втрат на навчальній та тестовій вибірці для обох моделей на рис. 2.



**Рис. 1.** Точність моделей на тестовій вибірці

*Джерело: розроблено авторами*



**Рис. 2. Функції втрат для навчальної і тестової вибірки**

*Джерело: розроблено авторами*

На графіках ми бачимо, що завдяки batch normalization точність моделі зросла на 10%. Також значення функції втрат для моделі з використанням нормалізації значно менше. Це дає нам можливість зробити висновок, що нормалізація допомагає набагато більше мінімізувати функцію втрат. На основі нашого експерименту видно, що batch normalization є ефективним методом для оптимізації нейронних мереж.

### Список використаних джерел:

1. Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q. (eds.), *Advances in Neural Information Processing Systems*, volume 25, pp. 1097–1105. Curran Associates, Inc., 2012. URL: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
2. Ioffe, S. and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Bach, F. and Blei, D. (eds.), *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pp. 448–456. Lille, France, 07–09 Jul 2015. PMLR. URL: <http://proceedings.mlr.press/v37/ioffe15.html>
3. Chen, G., Chen, P., Shi, Y., Hsieh, C.-Y., Liao, B., and Zhang, S. Rethinking the usage of batch normalization and dropout in the training of deep neural networks. arXiv preprint arXiv:1905.05928, 2019.