УДК 004.8

БАГАТОШАРОВА МАТРИЧНА НЕЙРОННА МЕРЕЖА У ЯКОСТІ АЛЬТЕРНАТИВИ БАГАТОШАРОВОМУ ПЕРЦЕПТРОНУ ДЛЯ ЗАДАЧІ КЛАСИФІКАЦІХ ЗОБРАЖЕНЬ

Албасова А.І.
Науковий керівник – д.т.н., проф. Бодянський Є.В.
Харківський національний університет радіоелектроніки, каф.ШІ,
м. Харків, Україна
тел. +38(098) 436-81-26, e-mail: alina.albasova@nure.ua

This paper is devoted to the exploration of the advantages of using matrix neuron based layers as an alternative to fully connected layers in the head of image classification networks. The recent advances in the application of neural networks for computer vision tasks, including visual transformers and full MLP architectures for image processing, are considered. The advantages of using matrix layers, including decreased number of parameters and reducing information loss, are described. It has been found that replacing fully connected layers with matrix layers and a vectorization-free approach has the potential for improving neural networks performance.

Конволюційні нейронні мережі на протязі останніх десятиліть були де факто стандартом для вирішення широкого кола задач комп’ютерного зору. Однак в останні роки, після успіху у вирішенні задач обробки природного тексту, нейромережеві архітектури на основі трансформерів [1] привертають до себе все більше уваги, як з боку дослідників, так і збоку індустрії. Крім того, успіх трансформерів, що позбавлені індуктивних зміщень, властивих конволюційним нейромережам, у задачах комп’ютерного зору поставив під сумнів необхідність та важливість операції згортки для навчання ефективних моделей комп’ютерного зору.

У роботі [2] дослідники запропонували архітектуру MLP-Mixer, що основана виключно на багатошарових перцептронах. Результати, опубліковані у роботі свідчать про те, що при використанні досить великих навчальних наборів даних та потужних методів регуляризації, MLP-Mixer досягає конкурентних результатів, порівняних з результатами просунутих конволюційних моделей та трансформерів.

Автори [3] у своїй роботі поставили під сумнів ефективність трансформера як архітектури у задачах комп’ютерного зору і висунули гіпотезу про те, що саме використання патчів в якості вихідного представлення дозволяє візуальним трансформерам (ViT) досягати високих результатів. У роботі запропонована гібридна архітектура під назвою ConvMixer – конволюційна мережа, в якій застосовуються підходи, що запропоновані та використовуються у ViT та MLP-Mixer архітектурах.

Незалежно від того, за допомогою яких підходів вихідне зображення перетворювалося до того, як передаватися на вхід предиктора, у якості

предиктора у всіх архітектурах використовується один чи декілька повнозв'язаних шарів, що на вхід приймають дані, представлені у вигляді вектора. Крім того, більшість архітектур також використовує перетворення до векторної форми всередині інших блоків. З точки зору обробки матричних сигналів, більш природнім було б аналізувати їх у формі матриць зниженої розмірності, замість того, щоб векторизувати їх, адже під час векторизації втрачається частина інформації, що міститься у зв'язках між рядками та стовпцями матричного входу.

У роботі [4] запропонований так званий матричний нейрон (M-нейрон) як альтернатива перцептрону Розенблата, що є будівельним блоком традиційного повнозв'язаного шару. На вхід M-нейрон приймає двовимірний сигнал, який, за допомогою двох наборів параметрів, перетворює у скалярний сигнал. Комбінуючи M-нейрони у вигляді двовимірної решітки, можна формувати необхідний вихідний сигнал. На основі таких шарів можна як побудувати багатошарову мережу, що повністю складається із матричних шарів, так і використовувати матричні шари як альтернативу повнозв'язаним шарам у архітектурах, що розглянуті вище і їм подібним. Окрім збереження інформації, що знаходиться у зв'язках між рядками та стовпцями матричного сигналу, така заміна також призведе до зменшення кількості параметрів, що налаштовуються та, відповідно, оптимізації часу навчання. Для прикладу, для того, щоб перетворити матричний сигнал розміром 32×32 у вектор розміром 1×100, де 100 – кількість класів, у випадку повнозв'язаного шару буде використано 4,096,000 параметрів, тоді як матричний шар може виконати таке ж саме перетворення використовуючи лише 64,064 параметрів. Використання нейронних мереж на основі матричних нейронів також відкриває нові перспективи для онлайн навчання.

Список використаних джерел:

1. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv, abs/2010.11929*.

2. Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Keysers, D., Uszkoreit, J., Lucic, M., & Dosovitskiy, A. (2021). MLP-Mixer: An all-MLP Architecture for Vision. *ArXiv, abs/2105.01601*.

3. Trockman, A., & Kolter, J.Z. (2022). Patches Are All You Need? *ArXiv, abs/2201.09792*.

4. Albasova, A., Bodyanskiy, Y.V., Deineko, A.O., & Pliss, I. (2021). 2-D Neural Network Based on M-neurons and its Learning. 2021 11th International Conference on Advanced Computer Information Technologies (ACIT), 700-703.