

**Хабарлак К.С.** асистент кафедри Системного аналізу та управління  
(Національний технічний університет «Дніпровська політехніка», м. Дніпро, Україна)

## НЕЙРО-МЕРЕЖЕВА СИСТЕМА КЛАСИФІКАЦІЇ ІЗ КОНФІГУРАЦІЄЮ ПІСЛЯ НАВЧАННЯ

Велика кількість практичних задач комп'ютерного зору потребують ефективного вирішення задачі класифікації зображень. Останніми роками найвищу якість класифікації демонструють згорткові нейронні мережі. Однак, такі мережі є обчислювально складними. Це є суттєвою проблемою на клієнтських пристроях (комп'ютерах, ноутбуках, мобільних телефонах), де обчислювальні можливості пристрою обмежені. Більш того, обчислювальні можливості комп'ютерів та смартфонів, бюджетних та коштовних гаджетів відрізняються на порядок [1]. Це шкодить більш широкому впровадженню нейронних мереж, адже в такому випадку одна нейронна мережа не може ефективно виконуватись на всіх категоріях вище перелічених пристроїв.

Для вирішення зазначеної проблеми ми пропонуємо адаптивну після навчання мобільну нейронну мережу [2], яку можна навчити один раз, а після цього конфігурувати, змінюючи швидкість її виконання. Це дозволяє динамічно адаптувати швидкість та якість мережі в залежності від потреби користувача.

Таку мережу засновано на Post-Train Adaptive (РТА) блоці, із використанням якого можна легко побудувати нейронну мережу. За допомогою спеціальної структури блоку, а також модифікованої процедури навчання, такий блок надає можливість конфігурувати нейронну мережу для пристроїв із значно відмінною швидкістю. Підкреслимо, що на відміну від існуючих нейронних мереж, запропонована мережа із РТА блоками навчається один раз, і вже після цього додатково конфігурується. Така конфігурація може відбуватися на основі бажаного часу, якості передбачення або ж навантаження на пристрій. Цікаво, що використання РТА блоків дозволило ще й значно покращити якість передбачення в задачі класифікації зображень.

**Висновки.** Розроблена нейронна мережа із РТА блоками дозволяє ефективно змінювати архітектуру нейронної мережі після навчання, що дозволяє динамічно обирати таку конфігурацію мережі, що виконується достатньо швидко на цільовому пристрої. Цікаво й те, що запропонована мережа демонструє вищу якість за рядом метрик у порівнянні з мережею без РТА блоків, навіть для конфігурацій із значно меншим часом виконання, що підтверджується цілою низкою практичних експериментів. Сподіваємось, що запропонований підхід збільшить можливість використання нейронних мереж на мобільних пристроях, та дозволить краще адаптувати нейронні мережі для них.

### ПЕРЕЛІК ПОСИЛАНЬ

1. Khabarlak K. Face detection on mobile: Five implementations and analysis / K. Khabarlak // CoRR. – 2022. – Vol. abs/2205.05572.
2. Khabarlak K. Post-Train Adaptive MobileNet for Fast Anti-Spoofing / K. Khabarlak // CEUR Workshop Proceedings – CEUR-WS.org, 2022. – Vol. 3156. – P. 44-53.