

РОЗПІЗНАВАННЯ ДІЙ У СПОРТИВНИХ ВІДЕО
Монастирський М. С., Колбасін В. О., Мезерна М. В.
Національний технічний університет
«Харківський політехнічний інститут», м. Харків

Розпізнавання дій у спортивних відео є важливою задачею, адже дає змогу автоматично збирати статистику спортивних подій, допомагати тренерам аналізувати виступи своїх підопічних і коригувати тренувальну програму, а також розробляти стратегію на наступні поєдинки, а також сприяє кращому розумінню спортивних трансляцій для покращення їх обробки методами машинного навчання.

У роботі було розглянуто підходи до вирішення задачі розпізнавання дій у спортивних відео на прикладі футболу і датасету SoccerNet. Цей датасет містить відеозаписи трансляцій 500 матчів сумарною тривалістю 764 години і 6637 анотацій подій, що представляють 3 великі класи (гол, жовта/червона картка і заміна). Цей датасет є однією із найбільших баз даних записів трансляцій футбольних матчів і підходить для вирішення різноманітних завдань машинного навчання в сфері розпізнавання дій у спортивних відео [1].

В поточній роботі до вирішення задачі застосовуються методи глибинного навчання із застосуванням згорткових і рекурентних нейронних мереж, оскільки вони дуже добре працюють із неструктурованими даними. Акцент робиться на застосуванні згорткових нейронних мереж до початкових відео, що дає змогу виокремити просторові залежності, з подальшою агрегацією отриманих ознак різними методами вздовж часової вісі і класифікацією. В даному підході розпізнавання дій відбувається в два етапи: на першому просторові ознаки виокремлюються попередньо тренованою згортковою нейронною мережею з початкових відео, а потім агрегуються за часовою віссю різними методами (пулінг та рекурентні нейронні мережі). Такий підхід не має переваг end-to-end навчання, проте дає переваги в ефективності використання ресурсів, при навчанні моделі, що в даному випадку є важливим, адже датасет є дуже великим; а також дає змогу експериментувати із різними методами агрегації ознак за часовою віссю і застосовувати принципи transfer learning.

У роботі порівнюються різні методи агрегації просторових ознак за часовою віссю, що засновані на принципах глибинного навчання (рекурентні нейронні мережі та методи пулінгу, що здатні до навчання), а також порівнюється їх ефективність відносно використання пам'яті і швидкості обробки моделлю вхідних даних, що є критично важливим для застосунків, що мають працювати в режимі реального часу.

Література:

1. S. Giancola, M. Amine, T. Dghaily, and B. Ghanem, "Soccernet: A scalable dataset for action spotting in soccer videos," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2018, pp. 1711-1721.