УДК 528.8.044
П. С. Панчишин; М. І. Паламар, д.т.н., професор.
(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

ПІДВИЩЕННЯ ЯКОСТІ РОЗПІЗНАВАННЯ ОБ'ЄКТІВ У ВІДЕОПОТОЦІ В РЕАЛЬНОМУ ЧАСІ

UDC 528.8.044
P. Panchyshyn; M. I. Palamar Dr., Prof.

IMPROVING THE QUALITY OF OBJECT RECOGNITION IN A REAL-TIME VIDEO STREAM

Розпізнавання об'єктів у реальному часі є важливим завданням у багатьох галузях, включаючи автономні системи, безпілотні літальні апарати, системи відеоспостереження та роботи. Технології, такі як нейромережа **YOLO (You Only Look Once)**, вже демонструють високу швидкість і точність, але складні умови, такі як погане освітлення, шуми або мала роздільна здатність відео, вимагають додаткових підходів для підвищення якості розпізнавання.

Розглянемо методи та алгоритми, які покращують точність та стабільність розпізнавання об'єктів у відеопотоці, для моделі YOLO, зокрема використання механізмів уваги, такі як **SE-блоки (Squeeze-and-Excitation)**, що дозволить моделі зосередитися на важливих частинах зображення, ігноруючи другорядні деталі. Це важливо для поліпшення точності в умовах складних сцен, де на зображенні можуть бути як значні, так і незначні об'єкти.

SE-блоки (Squeeze-and-Excitation):

Блоки працюють шляхом зменшення просторової інформації через глобальне середнє зменшення (**squeeze**) та потім виділяють важливі канали в мережі (**excitation**).

В результаті модель фокусується на найбільш інформативних ознаках, покращуючи детекцію об'єктів, що важливі для поставленої задачі.

Як працюють механізм уваги SE?

1. Після надходження зображення до нейронної мережі.
2. **Крок 1 — Squeeze (скорочення):**

Використовує глобальне середнє згортання по простору, щоб отримати згорнуті ознаки для кожного каналу.

Для кожного каналу *c* підраховуємо середнє значення по всіх просторових координатах:

$$z_c = \sigma\left(\frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{ijc}\right)$$

де x_{ijc} – значення пікселя на координатах *i*,
j для каналу *c*,
H та *W* – висота та ширина карти ознак,
а σ – сигмоїдальна функція активації для отримання вихідного значення в межах [0,1].

3. **Крок 2 – Excitation (збудження):**

Після отримання згорнутого вектора для кожного каналу, механізм генерації коефіцієнтів важливості кожного каналу через два повнозв'язкові шари:

$$s_c = \sigma(W_2\delta(W_1 z_c))$$

де W_1, W_2 – це ваги для двох лінійних шарів,
δ – *ReLU* активація для ненегативних значень,
S_c – вихідний коефіцієнт важливості для каналу c, що визначає, на скільки цей канал повинен бути посилений чи ослаблений.

4. **Крок 3 – Оновлення ознак:**

Множимо оригінальну карту ознак на відповідні коефіцієнти важливості S_c для кожного каналу:

$$\hat{x}_{ijc} = s_c x_{ijc}$$

Таким чином, канали з високими значеннями S_c будуть посилені, а канали з низькими значеннями – ослаблені.

5. **Вихід:**

Модифікована карта ознак, готова для подальшої обробки.

Переваги:

Покращена точність завдяки зосередженню уваги на важливих частинах зображення, модель може більш точно визначати об'єкти навіть у складних ситуаціях.

Зниження помилок зменшення ймовірності пропуску або неправильної детекції об'єктів, особливо при великій кількості фону або маленьких об'єктах.

Таким чином застосування механізму уваги **SE** у моделі **YOLO** є потужним інструментом для покращення ефективності глибоких нейронних мереж. Він дозволяє зосередити увагу на найбільш важливих каналах і забезпечує значне підвищення точності розпізнавання об'єктів, сегментації та інших задач. Впровадження SE-блоків дозволяє значно покращити загальну ефективність моделі без значного збільшення обчислювальних витрат.

Література

1. Xuan-Phung Huynh and Yong-Guk Kim. Discrimination between genuine versus fake emotion using long-short term memory with parametric bias and facial landmarks. In Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017.
2. David Daniel Cox and Thomas Dean. Neural networks and neuroscience-inspired computer vision. Current Biology, 2014.
3. Bin Yang, Junjie Yan, Zhen Lei, and Stan Z Li. Convolutional channel features. In Proceedings of the IEEE international conference on computer vision, 2015.
4. Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018
5. Behzad Hasani and Mohammad H Mahoor. Facial expression recognition using enhanced deep 3d convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017