

**ПОРІВНЯННЯ ПРЕТРЕНОВАНИХ МОДЕЛЕЙ ДЛЯ ДЕТЕКЦІЇ ОБ'ЄКТІВ**

UDC 004.6

**O. Sorokivskyi, I. Lytvynenko****COMPARATION OF THE PRETRAINED MODELS FOR OBJECT DETECTION**

Детекція об'єктів – це одне з найважливіших завдань комп'ютерного бачення, завдання якого знаходження об'єктів певних класів на заданих зображеннях. Для вирішення цієї не простої задачі існує два підходи:

1. Використання власноруч налаштованих фільтрів для виділення певних об'єктів на зображеннях. Цей спосіб не є ефективним, адже, зазвичай, об'єкти візуально відрізняються як і кольорами так і формою. Налаштовування фільтрів під кожен випадок може тривати дуже довго і не завжди успішно;

2. Використання моделей машинного навчання, які автоматично налаштовують безліч фільтрів використовуючи для цього промарковані зображення. Цей підхід є більш універсальним, адже він враховує всі можливі варіанти об'єктів, а також, володіє властивістю генералізації, що допомагає справлятися з раніше небаченими об'єктами.

Враховуючи те, що перший метод займає багато часу та потребує багато зусиль – для вирішення задачі детекції об'єктів буде використовуватись, в основному, другий спосіб.

Ця доповідь стосується теми «Порівняння претренованих моделей для детекції об'єктів». Для того, щоб отримати модель, яка здатна знаходити потрібні об'єкти на зображенні потрібно багато промаркованих даних, а також потужні обчислювальні ресурси. Для економії ресурсів та для покращення точності моделей використовується техніка під назвою «Передача навчання». Згідно з дослідженням Proper Reuse of Image Classification Features Improves Object Detection (1, Cristina Vasconcelos, Vighnesh Birodkar, Vincent Dumoulin, 2022) з правильним використанням попередньо тренованих моделей можна досягнути кращих результатів швидше чим без них.

Під час дослідження готових імплементацій моделей для детекції об'єктів було знайдено багато рішень які відрізняються як і датою публікації, точністю і швидкодією. Для вирішення поточної задачі не потрібно враховувати швидкодію, адже швидкість роботи не залежить від реального часу. Для порівняння точності моделей машинного навчання потрібно використовувати одні й ті ж дані. Одним з найпопулярніших відкритих наборів даних з готовими промаркованими об'єктами є Microsoft COCO (2, Tsung-Yi Lin Michael Maire Serge Belongie Lubomir Bourdev Ross Girshick, James Hays Pietro Perona Deva Ramanan C. Lawrence Zitnick Piotr Dollar, 2015) – це набір даних, який містить 328 тисяч картинок повсякденних об'єктів та людей.

Для порівняння було обрано такі моделі:

- InternImage;
- EVA;
- Co-DETR;
- FD-SwinV2-G.

Для кращого визначення моделі також буде враховуватись набір даних для семантичної сегментації ADE20K (3, Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso and Antonio Torralba. Computer Vision and Pattern Recognition (CVPR), 2017).

Для порівняння на COCO наборі даних використовувалась метрика box AP, а для порівняння на ADE20K – Validation mIoU.

На рисунку 1 зображена порівняльна таблиця моделей на обох наборах даних.

Model name	COCO test-dev	ADE20K
InternImage	65.0	62.9
EVA	64.7	62.3
Co-DETR	64.5	-
FD-SwinV2-G	64.2	61.4
YOLO-P6	55.4	-

**Рисунок 1.** Порівняльна таблиця

Як видно з таблиці, більшість порівнюваних моделей дуже близькі по значеннях. Модель YOLOR-P6 має найгірший результат, але моделі YOLO є найпопулярнішими моделями, які використовуються для детекції об'єктів через те, що вони мають декілька конфігурацій з різною швидкодією, що дозволяє налаштувати їх під потрібну задачу.

Отже, є багато моделей та наборів даних для їх порівнянь. Попри те, що є безліч моделей для знаходження об'єктів потрібно використовувати ресурси саме на ті, які дають хороші результати на схожих задачах. Всі моделі, які розглянуті вартують подальшого дослідження вже в рамках поточної задачі знаходження спортивних об'єктів на кадрах відеозапису.

### Літератури

1. Cristina Vasconcelos, Vighnesh Birodkar, Vincent Dumoulin. Proper Reuse of Image Classification Features Improves Object Detection, 2022. URL: <https://arxiv.org/abs/2204.00484>.
2. Microsoft COCO: Common Objects in Context, Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár, 2015. URL: <https://arxiv.org/abs/1405.0312>.
3. Semantic Understanding of Scenes through the ADE20K Dataset, Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, Antonio Torralba, 2018. URL: <https://arxiv.org/abs/1608.05442>.
4. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions, Wenhai Wang, Jifeng Dai, Zhe Chen, Zhenhang Huang, Zhiqi Li, Xizhou Zhu, Xiaowei Hu, Tong Lu, Lewei Lu, Hongsheng Li, Xiaogang Wang, Yu Qiao, 2022. URL: <https://arxiv.org/abs/2211.05778>.
5. EVA: Exploring the Limits of Masked Visual Representation Learning at Scale, Yuxin Fang, Wen Wang, Binhui Xie, Quan Sun, Ledell Wu, Xinggang Wang, Tiejun Huang, Xinlong Wang, Yue Cao, 2022. URL: <https://arxiv.org/abs/2211.07636v1>.
6. DETRs with Collaborative Hybrid Assignments Training, Zhuofan Zong, Guanglu Song, Yu Liu, 2022. URL: <https://arxiv.org/abs/2211.12860v1>.
7. Contrastive Learning Rivals Masked Image Modeling in Fine-tuning via Feature Distillation, Yixuan Wei, Han Hu, Zhenda Xie, Zheng Zhang, Yue Cao, Jianmin Bao, Dong Chen, Baining Guo, 2022. URL: <https://arxiv.org/abs/2205.14141v3>.
8. You Only Learn One Representation: Unified Network for Multiple Tasks, Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao, 2021. URL: <https://arxiv.org/abs/2105.04206v1>.