УДК 004.8
К. Вергелес, Т. Ємел'яненко, канд. тех. наук, доц.
(Дніпровський національний університет імені Олеся Гончара, Україна)

ЗАСТОСУВАННЯ МОДЕЛІ GROUNDING DINO ДО РОЗВ'ЯЗАННЯ ЗАДАЧ КОМП'ЮТЕРНОГО ЗОРУ

K. Verheles, T. Yemelianenko, Ph.D, Assoc. Prof.

APPLICATION OF THE GROUNDING DINO MODEL FOR SOLVING COMPUTER VISION TASKS

Комп'ютерний зір – підгалузь області штучного інтелекту, що спеціалізується на зборі, обробці та аналізі комп'ютерними системами візуальних даних. Незважаючи на значний прогрес, зокрема, у задачі виявлення об'єктів на зображенні, більшість існуючих моделей покладаються на техніку навчання із вчителем, що потребує використання великих анотованих наборів даних для навчання. Основна проблема такого підходу полягає у відсутності гнучкості, оскільки для розширення переліку класів об'єктів, що можуть розпізнаватися моделлю, необхідно додатково збирати дані, оброблювати їх, позначаючи мітки та обмежувальні рамки для об'єктів, та навчати модель, що є часозатратним та трудомістким процесом.

Поява моделей zero-shot object detection сприяє подоланню описаних обмежень за рахунок використання відкритого словника категорій. Такий підхід дозволяє виявляти об'єкти нових класів без попереднього додаткового навчання моделі. Іншою перевагою описаного методу є можливість використання природної мови для взаємодії із моделлю. Даний підхід надає можливість використовувати довгі та нечіткі описи природною мовою, на відміну від традиційних методів виявлення об'єктів, що не здатні розпізнавати опис, довший за мітку класу, що зазвичай є іменником чи коротким словосполученням.

Для дослідження підходу zero-shot object detection було обрано задачу виявлення продуктів харчування на зображенні. Для вирішення даного завдання було застосовано метод Grounding DINO [1], який поєднує концепції DINO, моделі глибокого навчання на основі трансформерів, що відповідає за виявлення об'єктів та наскрізну оптимізацію, та GLIP, що займається поєднанням фраз із заданого тексту із відповідним візуальним представленням об'єкта на зображенні або відео. Алгоритм Grounding DINO реалізовано у відповідній бібліотеці [2] мови програмування python. У результаті обробки моделлю завантаженого вхідного зображення із продуктами та переліку можливих класів ('kiwi, apple, banana, mango, orange, lemon, lime, pineapple') було отримано вихідне зображення із обмежувальними рамками та мітками класів для виявлених об'єктів (рис. 1).

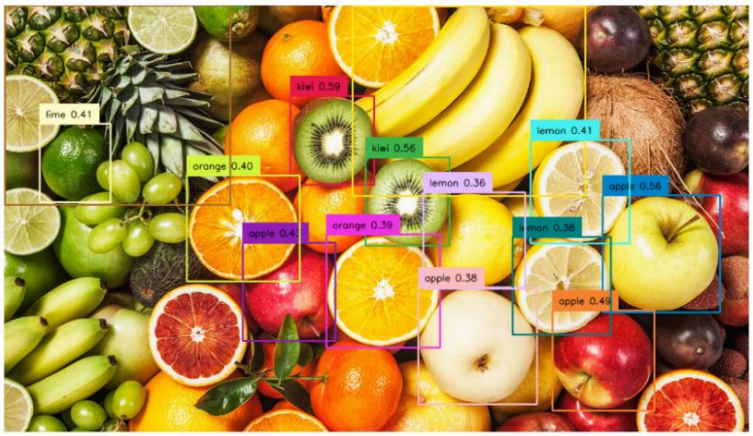

Рисунок 1. Використання Grounding DINO для виявлення продуктів на зображенні

Поєднання технологій DINO та GLIP дозволяє моделі Grounding DINO розпізнавати об’єкти на нових зображеннях без попереднього навчання моделі для виявлення нового класу. Таким чином, завдяки можливості ідентифікувати об’єкти поза межами тренувального набору, дана модель може досягати успіхів у різних сферах, зокрема, і для розв’язання розглянутої задачі виявлення продуктів харчування.

Література

1. Shilong Liu, Zhaoyang Zeng, Tianhe Ren, Feng Li, Hao Zhang, Jie Yang, Chunyuan Li, Jianwei Yang, Hang Su, Jun Zhu, Lei Zhang. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. 2023. DOI: https://doi.org/10.48550/arXiv.2303.05499.

2. Бібліотека Grounding Dino [Електронний ресурс] – Режим доступу: https://github.com/IDEA-Research/GroundingDINO.