

РОЗРОБКА МЕТОДІВ СЕМАНТИЧНОЇ СЕГМЕНТАЦІЇ ЩО ПРАЦЮЮТЬ У РЕАЛЬНОМУ ЧАСІ ТА Є УЗГОДЖЕНИМИ У ЧАСІ

Шугаєнко І. Д., Ямковий К. С.

***Національний технічний університет
«Харківський політехнічний інститут», м. Харків***

Задача семантичної сегментації полягає у визначенні піксельних масок областей зображення, що містять об'єкти з різною семантикою, і є важливою у багатьох областях комп'ютерного зору, включаючи медичну діагностику, автономне водіння та розпізнавання об'єктів на зображеннях.

Класичний підхід вирішення задачі семантичної сегментації, коли сегментується лише одне зображення без урахування інформації із сусідніх зображень, може стикатися з проблемами, коли об'єкти змінюють свою форму або розмір у часі, особливо під час руху камери. Методи семантичної сегментації, які є узгодженими у часі, дозволяють отримувати більш точні та стабільні результати сегментації.

Відеоспостереження є однією з найбільш поширених областей застосування задачі семантичної сегментації, узгодженої у часі, де стійка сегментація відео є важливим етапом для виявлення та відстеження об'єктів на відео. У даній роботі буде розглянуто вирішення задачі семантичної сегментації для відеоаналізу на наборі даних Cityscapes dataset.

Cityscapes dataset є одним з найбільш популярних наборів даних для задачі семантичної сегментації. Він містить розмічені зображення високої якості міських сцен, що включають різноманітні об'єкти, такі як дороги, будівлі, автомобілі, пішоходи та інші.

Задля забезпечення узгодженості передбачень у часі використовується спеціальна функція витрат, що має назву temporal loss. Роль temporal loss в задачі реального часу сегментації відео з узгодженістю в часі полягає в тому, щоб забезпечити стабільність сегментації об'єктів протягом часу.

Під час вирішення задачі було прийнято рішення звернутися до техніки дистиляції знань (knowledge distillation) через те, що моделі, що вирішують задачу семантичної сегментації, мають досить велику кількість параметрів і не працюють у режимі реального часу. Суть ідеї полягає у тому, що в процесі передачі знань модель з більшою кількістю параметрів, яка називається "вчительською моделлю", використовується для навчання моделі з меншою кількістю параметрів, яка називається "студентською моделлю", з метою підвищення її точності та ефективності.

У процесі виконання роботи була зібрана і натренована модель, яка вирішує задачу семантичної сегментації у реальному часі з узгодженням у часі.

Література:

1. "Efficient Video Segmentation Models with Per-frame Inference", 2022 [Електронний ресурс]: Режим доступу: <https://arxiv.org/pdf/2202.12427.pdf>.