УДК 004.932

ДЕТЕКЦИЯ МАСКИРОВАННЫХ ОБЪЕКТОВ НА СНИМКАХ БПЛА

Воловиков Г.А.[1], Левитан В.Р.[1], Ильин С.Ю.[1]
Научный руководитель – канд. физ. - мат. наук, Балахчи А. Г.[1]
[1]Университет ИГУ
jezv@yandex.ru

Введение

Детекция маскированных объектов на аэрофотоснимках с БПЛА — актуальная задача компьютерного зрения и дистанционного зондирования. Её сложность связана с камуфляжем, низким контрастом объектов, изменчивостью освещённости, тенями, атмосферными условиями и ограниченным разрешением сенсоров. Традиционные методы на основе глубинных сверточных нейронных сетей эффективны на стандартных датасетах, но требуют больших размеченных данных и теряют точность при переносе на специализированные съёмки БПЛА. Архитектуры YOLO обеспечивают баланс скорости и точности, однако их устойчивость к сложной маскировке ограничена [3].

Современные исследования в области самообучающихся и мультимодальных моделей открывают новые возможности для повышения обобщающей способности систем детекции. Модель CLIP демонстрирует способность формировать универсальные визуально-семантические представления за счёт обучения на больших корпусах изображений и текстовых описаний, что позволяет применять её в задачах zero-shot и few-shot детекции [1]. В свою очередь, подход DINOv3 обеспечивает получение устойчивых визуальных признаков без использования разметки, что делает его перспективным для доменов с ограниченными аннотированными данными [2]. Дополнительно, методы дистилляции знаний позволяют переносить информацию от крупных высокоточных моделей к компактным архитектурам, снижая вычислительные затраты без существенной потери качества распознавания [4]. Совмещение указанных направлений представляет собой перспективную стратегию для разработки эффективных систем детекции, пригодных для эксплуатации на борту БПЛА в условиях ограниченных вычислительных ресурсов.

Основная часть

Предлагаемое в данной работе решение основано на двухэтапной архитектуре, включающей высокообобщающую teacher-модель и компактную student-модель, предназначенную для развертывания на борту БПЛА. В качестве базового блока teacher-модели используется модель DINOv3. Данная модель формирует устойчивые визуальные представления, обладающие инвариантностью к изменениям масштаба, освещённости, частичным перекрытиям и текстурным искажениям, что особенно важно при анализе аэрофотоснимков. Для реализации механизма детекции по текстовому описанию визуальные признаки проецируются в общее латентное пространство совместно с текстовыми эмбеддингами модели CLIP. На основе косинусного сходства между визуальными и текстовыми представлениями формируются тепловые карты вероятности присутствия целевого объекта в различных областях изображения. Далее посредством алгоритма подавления немаксимумов осуществляется выделение ограничивающих рамок. При этом дообучение затрагивает исключительно проекционные слои, что позволяет сохранить обобщающую способность базовой самообучающейся модели и минимизировать вычислительные затраты.

На втором этапе осуществляется перенос знаний teacher-модели в компактную архитектуру YOLOv11-nano. Передача знаний реализуется посредством комбинированной процедуры дистилляции, включающей сопоставление промежуточных признаков, дистилляцию классификационных выходов и использование псевдоразметки, сформированной на неразмеченных данных. Такой подход позволяет сохранить высокую точность детекции, характерную для крупной модели, при значительном уменьшении числа параметров и вычислительной сложности.

Для интеграции в БПЛА модель экспортируется в формат ONNX и подвергается оптимизации с использованием TensorRT. Применяются методы квантования и структурного прунинга каналов с учётом сохранения информации, полученной в процессе дистилляции знаний.

Выводы

Разработан двухэтапный метод детекции маскированных объектов на аэрофотоснимках БПЛА, основанный на использовании самообучающейся teacher-модели и компактной student-модели с дистилляцией знаний. Предложенный подход обеспечивает высокую точность распознавания при ограниченных вычислительных ресурсах и позволяет реализовать инференс в реальном времени.

Практическое применение результатов возможно в системах мониторинга территорий, поисково-спасательных операциях, задачах наблюдения и анализа сложных сцен.

Литература

1. Radford A., Kim J. W., Hallacy C., et al. Learning Transferable Visual Models From Natural Language Supervision // Proceedings of the 38th International Conference on Machine Learning (ICML). 2021. Vol. 139. P. 8748–8763. URL: https://arxiv.org/abs/2103.00020
2. Oquab M., Darcet T., Moutakanni T., et al. DINOv2: Learning Robust Visual Features without Supervision // arXiv preprint. 2023. URL: https://arxiv.org/abs/2304.07193
3. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement // arXiv preprint. 2018. URL: https://arxiv.org/abs/1804.02767
4. Hinton G., Vinyals O., Dean J. Distilling the Knowledge in a Neural Network // arXiv preprint. 2015. URL: https://arxiv.org/abs/1503.02531