

УДК 004.021

ПОВЫШЕНИЕ ТОЧНОСТИ РАСПОЗНАВАНИЯ МАЛОРАЗМЕРНЫХ ОБЪЕКТОВ

Краснов Д.И. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Гусев А.А.
(ИТМО)

Введение. Задача распознавания объектов размером 10-40 пикселей на изображении встречается достаточно часто в системах безопасности и контроля периметра, а так же в системах анализа биомедицинских изображений. Малоразмерный объект содержит крайне мало семантической информации, а частое наличие сложного фона (облаков и инфраструктуры) значительно усложняет локализацию объекта.

Одним из путей решения данной задачи является использование инфракрасного (ИК) диапазона, где малый объект является ярко светящейся точкой. Для детекции объектов используются различные алгоритмы подавления фона [1] или сверточные нейронные сети (СНС) [2]. Использование ИК диапазона позволяет с высокой точностью детектировать малые объекты в системах безопасности, но приводит к невозможности классификации объектов, к значительному понижению разрешения и повышению стоимости оборудования.

Другим путем является использование видимого диапазона с высоким разрешением и обработкой СНС либо одного кадра, либо последовательности кадров. Использование последовательности кадров позволяет повысить точность локализации и классификации за счет анализа траекторий движения объектов [3]. Однако точность распознавания малоразмерных объектов на изображениях видимого диапазона обычно ниже. Для повышения точности распознавания часто использует метод повышения локального разрешения [4] или вспомогательные блоки для подавления фона [3].

В данной работе исследовано влияние блоков пространственного и межканального внимания для подавления фона на точность семантической сегментации малоразмерных летательных аппаратов (самолетов, вертолетов и дронов).

Основная часть. Для оценки влияния блоков внимания на точность сегментации малых объектов был собран набор данных из открытых источников [5-7]. Итоговый датасет содержит объекты размером 5-70 пикселей (самолеты, вертолеты и дроны) и состоит из 12242 изображений с соответствующими сегментационными масками. Датасет был разделен на обучающую, валидационную и тестовую выборки.

В данной работе исследуются блоки межканального и пространственного внимания, которые являются самостоятельными единицами и могут быть внедрены практически в любую часть СНС. Принцип работы блока межканального внимания состоит в преобразовании входного тензора в вектор весов с помощью операции \max pooling или average pooling . Полученный вектор весов проходит через несколько сверточных слоев с одиночным ядром и умножается на входной тензор, в результате чего часть каналов подавляется, а часть усиливается. Принцип работы блока пространственного внимания состоит в преобразовании входного тензора в маску с помощью сверточного слоя с одним каналом. Полученная маска поэлементно умножается на входной тензор, в результате чего часть областей усиливается, а часть подавляется.

В качестве базовой архитектуры использовалась нейронная сеть ESNet [8], состоящая из энкодера и декодера. В качестве базового способа повышения точности использовалось смешение признаков (feature fusion) между тремя последними блоками декодера. Исследуемые блоки межканального и пространственного внимания были добавлены в базовую архитектуру по отдельности и одновременно. Таким образом, было получено 5 архитектур СНС. Эти модели были обучены с разрешением 512x512 с параметрами: оптимизатор – Adam, количество эпох – 200, размер пакета – 16, функция потерь – dice loss . Обученные модели были проверены на тестовой выборке с построением precision-recall и ROC кривых и сбором метрик: F-мера, мера Жаккара (IoU), $\text{average precision (AP)}$ и ROC

AUC.

Результаты показали, что использование модуля feature fusion позволяет увеличить точность по AP и IoU на 8% (AP 0,571 и IoU 0,547 против AP 0,613 и IoU 0,595), а совместное использование модулей межканального и пространственного внимания более чем на 10% (AP 0,571 и IoU 0,547 против AP 0,639 и IoU 0,615). При этом отдельное использование этих блоков дает результат хуже, чем смешение признаков. Для оценки возможности подавления сложного фона были построены тепловые карты путем усреднения предпоследних тензоров для базовой СНС и для СНС с модулями внимания. Результаты показали, что базовая нейронная сеть обращает лишнее внимание на объекты инфраструктуры и облака, в то время как сеть с модулями внимания подавляет даже сложный фон и выделяет более информативную область объекта.

Выводы. В работе проведено исследование влияния модулей внимания на точность семантической сегментации малоразмерных объектов с помощью СНС. Результаты валидации обученных моделей на тестовой выборке показали, что модули внимания способны значительно увеличить точность распознавания малоразмерных объектов (более чем на 10%) на изображениях видимого диапазона. При этом описанные модули могут быть внедрены в топологию СНС практически любой сложности без значительного увеличения вычислительной сложности. Полученные результаты будут полезны исследователям и инженерам в качестве одного из способа повышения точности при распознавании маленького объекта на сложном фоне.

Список использованных источников:

1. Chen CLP, Li H, Wei Y, Xia T, Tang YY. A Local Contrast Method for Small Infrared Target Detection. *IEEE Transactions on Geoscience and Remote Sensing* 2014; 52(1): 574-581. DOI: 10.1109/TGRS.2013.2242477.
2. Li B, Xiao C, Wang L, Wang Y, Lin Z, Li M, An W, Guo Y. Dense Nested Attention Network for Infrared Small Target Detection. *IEEE Transactions on Image Processing* 2022; 99: 1-1. DOI: 10.1109/TIP.2022.3199107.
3. Ashraf M, Sultani W, Shah M. Dogfight: Detecting Drones from Drones Videos. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 2021; 7063-7072. DOI: 10.1109/CVPR46437.2021.00699.
4. Wang Z, Wang C, Chen Y, Li J. Target detection algorithm based on super-resolution color remote sensing image reconstruction. *Journal of Measurements in Engineering* 2023; 12(1): 15. DOI: 10.21595/jme.2023.23510.
5. Li J, Ye DH, Kolsch M, Wachs JP, Bouman CA. Fast and Robust UAV to UAV Detection and Tracking From Video. *IEEE Transactions on Emerging Topics in Computing* 2022; 10(3): 1519-1531. DOI: 10.1109/TETC.2021.3104555.
6. Svanström F, Englund C, Alonso-Fernandez F. Real-Time Drone Detection and Tracking With Visible, Thermal and Acoustic Sensors. *25th International Conference on Pattern Recognition (ICPR)* 2020: 7265-7272. DOI: 10.1109/ICPR48806.2021.9413241.
7. Rozantsev A, Lepetit V, Fua P. Flying objects detection from a single moving camera. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2015: 4128-4136. DOI: 10.1109/CVPR.2015.7299040.
8. Wang Y, Zhou Q, Xiong J, Wu X, Jin X. ESNet: An Efficient Symmetric Network for Real-Time Semantic Segmentation. *Pattern Recognition and Computer Vision. (PRCV)* 2019: 41-52. DOI: 10.1007/978-3-030-31723-2_4.