УДК 004.932.72 DOI 10.24412/2413-7383-2024-4-25-37

Я. С. Пикалёв
Федеральное государственное бюджетное научное учреждение
«Институт проблем искусственного интеллекта», г. Донецк
283048, г. Донецк, ул. Артема, 118 б

ОБНАРУЖЕНИЕ КЛЮЧЕВЫХ ОБЪЕКТОВ И ПЕРЕКРЁСТНАЯ ГЕОЛОКАЛИЗАЦИЯ: АНАЛИЗ НАБОРОВ ДАННЫХ И МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ*

Y. S. Pikalyov
Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»
283048, Donetsk, Artema str, 118-b

KEY OBJECTS DETECTION AND CROSS-VIEW GEOLOCATION: DATASET ANALYSIS AND METHODOLOGICAL ASPECTS

Данная работа посвящена проблеме создания системы для задачи перекрёстной геолокализации на основе нейросетевого подхода. Это направление актуально в связи с тем, что эти системы помогают БПЛА ориентироваться в сложных условиях, идентифицируя объекты, препятствия и маршруты, что позволяет снизить зависимость от операторов. К тому же, используя системы распознавания, БПЛА могут осуществлять сбор и анализ данных. В ходе данной работы был проведен анализ существующих наборов данных, применимых к задаче перекрестной геолокации, а также были выявлены общие положения и требования к используемым наборам данных. Среди всех наборов были выделены следующие: Objects 365, LVIS, VisDrone, DOTA, iSAID и GeoText. Выделены следующие основные проблемы рассматриваемой задачи: 1) предварительная обработка изображений; 2) извлечение признаков; 3) целесообразность предварительного обучения базовой модели; 4) учет семантических характеристик объектов и другие.

Ключевые слова: нейронные сети, наборы данных, перекрёстная геолокализация, компьютерное зрение, распознавание объектов, аугментация данных, автономные системы.

This work focuses on the challenge of developing a system for cross-view geo-localization using a neural network-based approach. This area is highly relevant, as such systems enable UAVs to navigate complex environments by identifying objects, obstacles, and routes, thereby reducing dependence on operators. Additionally, recognition systems allow UAVs to efficiently collect and analyze data.
In the course of this study, an analysis of existing datasets suitable for cross-view geo-localization tasks was conducted, and general principles and requirements for such datasets were identified. The following datasets were highlighted: Objects365, LVIS, VisDrone, DOTA, iSAID, and GeoText.
Moreover, the key challenges of the task were outlined, including: 1) image preprocessing; 2) feature extraction; 3) the feasibility of pre-training a base model; 4) accounting for the semantic characteristics of objects, among others.

Keywords: neural networks, datasets, cross-view geo-localization, computer vision, object recognition, data augmentation, autonomous systems.

* Работа выполнена в рамках федерального проекта «Развитие человеческого капитала в интересах регионов, отраслей и сектора исследований и разработок» национального проекта «Наука и университеты» по теме научной молодежной лаборатории «Извлечение семантической информации из изображений для автономных систем навигации беспилотных летательных аппаратов (FREN-2024-0002)».

Введение

Системы распознавания для беспилотных летательных аппаратов (БПЛА) становятся все более актуальными в свете быстрого развития технологий и распространения применения БПЛА в таких сферах как сельское хозяйство, логистика, мониторинг территорий и окружающей среды [1], [2]. Системы распознавания позволяют эффективно обрабатывать информацию и принимать решения в реальном времени. Одной из главных задач при разработке БПЛА является повышение их автономности. Системы распознавания помогают БПЛА ориентироваться в сложных условиях, идентифицируя объекты, препятствия и маршруты, что позволяет снизить зависимость от операторов. Помимо этого, при помощи систем распознавания БПЛА могут осуществлять сбор и анализ данных (данные о состоянии окружающей среды, инфраструктуры или объектов).

Таким образом, системы распознавания для БПЛА играют ключевую роль в их развитии и применении, обеспечивая более высокую степень автономности, безопасности и эффективности в выполнении различных задач.

Постановка задачи

Отдельным направлением является задача перекрёстной геолокализации (Cross-View Geo-Localization, CVGL). Суть этой задачи сводится к задаче сопоставления изображений, а именно изображений со спутника (как правило, заранее загруженных в модуль навигации БПЛА) с изображениями, полученных при помощи камер БПЛА [3-6]. Т.е. CVGL-системы позволяют определить географическое положение текущего изображения, полученного с БПЛА, путем сравнения его с базой данных геометок на спутниковых снимках.

Рассмотрим пример работы подобных систем на основе глубокого обучения на рис. 1.

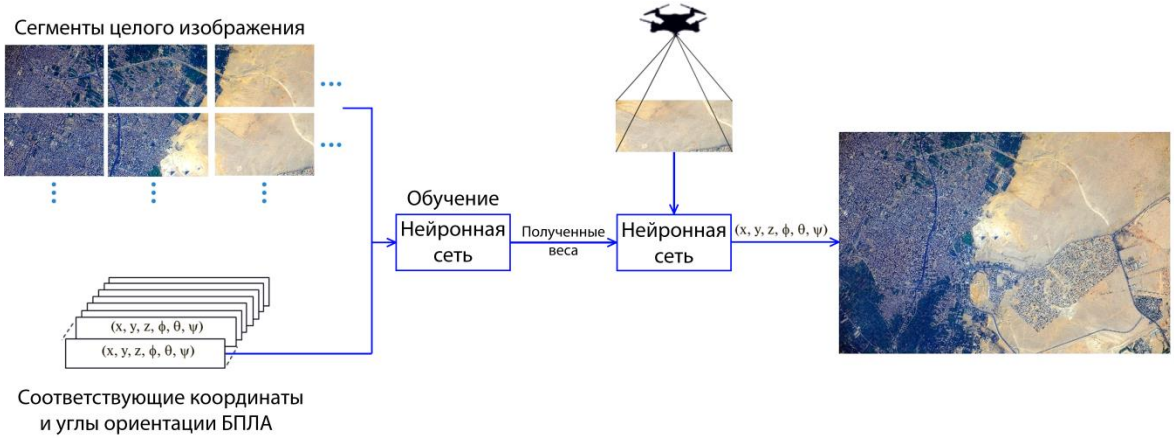

Рисунок 1 – Схема алгоритм навигации на основе сопоставления изображений с использованием нейронных сетей

Как видно из рис.1, CVGL основана на нейронной сети, которая, получая на вход изображение, формирует на выходе координаты и параметры ориентации БПЛА. При этом настройка весовых коэффициентов нейронной сети проводится на предварительно подготовленном наборе данных с имеющихся геопривязанными изображениями или изображениях, для которых точно известно местоположение и ориентация

объекта. Следовательно, отдельной проблемой при создании CVGL-систем на основе нейросетевого подхода является наличие аннотированного набора изображений с видом со спутника (satellite-view) и видом с дрона (drone-view).

Цель работы заключается анализе методологических аспектов для систем перекрёстной геолокации на основе нейросетевых подходов, а также выявления общих положений и требований к разработке подобных систем.

Исходя из вышеуказанных проблем**,** в работе были поставлены **следующие задачи**:

1) проанализировать существующих наборов данных, применимых к задаче CVGL
2) описать общие положения и требования к указанным выше наборам данных.

1. Описание существующих наборов данных

Анализ и обработка данных в области системного анализа и информационных технологий требует использования качественных и репрезентативных наборов данных [7], [8]. Доступ к существующим наборам данных позволяет исследователям разрабатывать, тестировать и оптимизировать алгоритмы, повышать точность моделей и ускорять процесс разработки решений для реальных задач. В данном разделе представлен обзор наиболее распространённых и значимых наборов данных аэро и космических снимков с учетом семантики для задач БПЛА.

Условно наборы данных для данной задачи можно разделить на 3 группы:

1) Наборы данных для распознавания объектов. Их целевое назначение заключается в предварительном обучении нейросетевых моделей с целью формирования внутренней семантики для объектов.

2) Наборы данных для распознавания объектов, полученных с БПЛА или спутника. В эту категорию входят наборы данных как без аннотаций, так и аннотированные для распознавания зданий или объектов местности.

3) Наборы данных для перекрёстной геолокализации. В эту группу входят наборы данных, предназначенные для сопоставления изображений полученных с дрона и со спутника, или для определения координат БПЛА.

1.1 Наборы данных для распознавания объектов

1) *ADE20K*. Набор данных включает в себя около 3000 именованных объектов, областей содержимого и деталей. Примечательно, что ADE 20K был аннотирован одним экспертом-аннотатором, что повышает согласованность, но также ограничивает размер набора данных. Из-за относительно небольшого количества изображений с аннотациями в большинстве категорий недостаточно данных для проведения как обучения, так и оценки.

2) *PASCAL VOC* (*Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes*) – первая попытка создать соревнование для исследователей. Имеет разметку для сегментации и детекции. Отсутствие разделения на тестирующую выборку, а также малое количество аннотированных данных – главная причина, почему он теряет свою актуальность.

3) *MS COCO* представляет собой крупномасштабный набор данных для обнаружения, сегментации и подписи объектов с 330 тыс. изображений. Он направлен на решение трех основных проблем понимания сцены: обнаружение неканонических

видов (или перспектив) объектов, контекстуальные рассуждения между объектами и точная двухмерная локализация объектов. COCO определяет 12 метрик для оценки производительности детектора, что дает более подробный и глубокий взгляд. Это самый популярный тест сегментации экземпляров для обычных объектов. Он содержит 80 категорий, которые попарно различны. Всего имеется 118 тыс. обучающих изображений, 5 тыс. проверочных изображений и 41 тыс. тестовых изображений. Все 80 категорий исчерпывающе аннотированы на всех изображениях (без учета ошибок в аннотациях), что приводит примерно к 1,2 миллионам масок сегментации экземпляров.

4) *ImageNet* – проект по созданию и сопровождению массивной базы данных аннотированных изображений, предназначенная для отработки и тестирования методов распознавания образов и машинного зрения. Создан в соответствии с иерархией WordNet.

5) *Open Images*. Это набор данных из около 9 млн изображений, снабженных аннотациями на уровне изображений, ограничивающими рамками объектов, масками сегментации объектов, визуальными взаимосвязями и локализованными описаниями: Он содержит в общей сложности 16 миллионов ограничивающих рамок для 600 классов объектов на 1,9 миллиона изображений, что делает его самым большим из существующих наборов данных с аннотациями местоположения объектов.

6) Авторы набора *LVIS* [9] основывались на подходах создания вышеуказанных наборов данных, а также на наборах данных, которые сосредоточены на уличных сценах и пешеходах. Ключевой особенностью LVIS является представление набора данных как интегрированного (federated) набора. Т.е. эта особенность заключается в том, чтобы желаемый протокол оценки не требовал от исследователя исчерпывающего аннотирования всех изображений по всем категориям. Вместо этого требуется, чтобы для каждой категории $\boldsymbol{c}$ существовало два непересекающихся подмножества всего набора данных $\boldsymbol{D}$, для которых выполняются следующие положения. В среднем к каждому изображению прилагается 11,2 экземпляра из 3,4 категорий. Самое большое количество экземпляров на изображение – 294.

7) *Objects365* [10] – крупномасштабный высококачественный набор данных для обнаружения объектов, который фокусируется на трех аспектах: масштабе, качестве и обобщении. Objects365 значительно больше, чем существующие тесты обнаружения объектов, такие как PASCAL и COCO. Он содержит 365 категорий, 638 тыс. изображений и 10 млн ограничивающих рамок. Objects365 содержит в 5 раз больше изображений, в 4 раза больше категорий и в 10 раз больше ограничивающих рамок, чем COCO. Набор данных Objects365 предоставляет лучшую альтернативу для изучения объектов. Набор данных разделен на обучающий (600 тыс.), проверочный (38 тыс.) и тестовый (100 тыс.) наборы.

В табл. 1 приведена статистика существующих наборов обнаружения объектов вместе с Objects365. Objects365 содержит изображений примерно в 60 раз больше, чем PASCAL VOC, и в 5 раз больше, чем COCO. По сравнению с набором данных ImageNet, Objects365 содержит большее количество блоков на изображение: 15,8 против 1,1 (2,3 для плотного набора). По сравнению с OpenImages, Objects365 полностью аннотирован людьми, при этом по сравнению с OpenImages он обладает большим распределением ограничивающих рамок на изображении: 15.8 против 9.8.

Таблица 1 – Сравнение статистики набора данных с существующими критериями обнаружения объектов.

Набор данных	Кол-во изображений, тыс.	Кол-во ограничивающих рамок, тыс.	Кол-во категорий	Отношение количества рамок/ изображений	Полностью аннотирован
Pascal VOC	11.5	27	20	2.4	+
ImageNet All	477	534	200	1.1	+
ImageNet Dense	80	186	200	2.3	+
COCO	123	896	80	7.3	+
OpenImages	1515	14815	600	9.8	Частично
Objects365	638	10101	365	15.8	+

1.2 Наборы данных для распознавания объектов, полученных с БПЛА или спутника

1) *Eurosat* – это набор данных и ориентир для глубокого изучения землепользования и классификации растительного покрова. Набор данных состоит из 10 различных классов по 2000 – 3000 изображений в каждом классе. Всего в наборе данных 27 тыс. изображений. Размер участков 64 × 64 пикселей. В наборе данных присутствуют такие классы: шоссе, жилые и промышленные здания, река, море, озеро, лес и травянистая растительность.

2) *iSAID* [11] – эталонный набор данных для сегментации аэрофотоснимков, который сочетает в себе задачи обнаружения объектов на уровне экземпляра и сегментации на уровне пикселей. Набор данных содержит 655451 экземпляров объектов для 15 категорий на 2806 изображениях высокого разрешения. Такие точные примечания к пикселям для каждого экземпляра обеспечивают точную локализацию, что важно для детального анализа сцены.

3) *XView* – это один из крупнейших общедоступных наборов данных спутниковых снимков, который содержит изображения сложных сцен по всему миру, аннотированных с помощью ограничивающих рамок. Он содержит более 1 млн экземпляров объектов, представленных 60 классами на территории более 1400 км². Данные собраны со спутников WorldView-3 с разрешением 0.3 метра на пиксель, что обеспечивает более высокое разрешение по сравнению с большинством общедоступных наборов спутниковых изображений.

4) *VisDrone* [12] – это специальный набор данных, состоящий из изображений с дронов. Всего в нем 8599 изображений, в том числе 6471 для обучения, 548 для обнаружения 18 объектов с помощью сверточных нейронных сетей. Набор данных был собран с использованием различных платформ дронов (т. е. дронов разных моделей), в разных сценариях (в 14 разных городах, раскинувшихся на тысячи километров) и при различных погодных условиях и условиях освещения. Максимальное разрешение статических изображений составляет 2000 × 1500.

5) *UAVOD-10* – набор данных обнаружения объектов БПЛА из 10 категорий, созданный для облегчения дистанционного обнаружения небольших слабых объектов на снимках БПЛА. Набор данных UAVOD-10 содержит 844 изображения и в общей сложности 18234 экземпляра, каждый из которых аннотирован горизонтальными ограничивающими рамками (HBB). Изображения имеют ширину от 1000 до 4800 пикселей с приблизительным разрешением 0,15 метра.

6) *Dataset of Object deTection in Aerial images (DOTA)* [13] – набор данных, который содержит 1793658 экземпляров объектов, охватывающих 18 различных кате-

горий, все из которых аннотированы с помощью аннотаций ориентированных ограничивающих рамок (OBB). Эти аннотации были собраны из 11268 аэрофотоснимков. Изображения, используемые в DOTA-v2.0, получены из трех различных источников изображений: изображения Google Earth, спутниковые изображения GF-2 и JL-1 (GF&JL) и воздушные изображения CycloMedia. Другой ценной метаинформацией является расстояние выборки на земле (GSD), которое измеряет расстояние между центрами пикселей на Земле. GSD ценно для расчета фактических размеров объектов, которые, в свою очередь, могут использоваться для идентификации неправильно маркированных или неправильно классифицированных экземпляров.

7) *AeroScapes* – набор данных из 3269 изображений воздушных сцен (снятых с помощью флота дронов), аннотированных с помощью плотной семантической сегментации. AeroScapes состоит из изображений, полученных с помощью дронов на высоте 5 – 50 метров. Набор данных содержит 12 различных классов, включая фон, растительность, дорогу и другие: человек, препятствие, сооружение, велосипед, автомобиль, небо, дрон, животное и лодка. В наборе данных есть 2 разделения: train (2621 изображение) и val (648 изображений).

1.3 Наборы данных для перекрёстной геолокализации

1) *CVUSA* состоит из пар наземных (ground-view) и аэрофотоснимков со всей территории США. Ground-view снимки были собраны как с Google Street View, так и с Flickr. Для каждого наземного снимка был сформирован аэрофотоснимок размером 800 × 800 с центром в этом месте из Bing Maps в нескольких пространственных масштабах (уровни масштабирования 14, 16 и 18). После учета совпадений в результате получается в общей сложности 35532 пар изображений для обучающего набора и 8884 для тестового набора.

2) В *CVACT* ground-view снимки также собирались с помощью Google Street View и охватывают географическую область площадью 300 кв. миль при уровне масштабирования 2. Разрешение изображений составляет 1664 × 832. Спутниковые снимки (satellite-view) получены с помощью Google Map. Для каждого ground-view было получено соответствующее спутниковое изображение в соответствии с местоположением ground-view, полученным с помощью GPS, с максимальным увеличением 20. Разрешение satellite-view снимков составляет 1200 × 1200. Разрешение ground-view снимков составляет 0,12 метра на пиксель. Тестовый набор изображений составляет 92802 пар перекрестных изображений.

3) VIGOR состоит из 90618 аэрофотоснимков, охватывающие города США. Этот набор данных собран при помощи Google Maps и Google Street-View. Все GPS-координаты панорамных снимков уникальны и типичный интервал между экземплярами составляет около 30 м. Для полученных данных была проведена балансировка по исходным панорамам, чтобы убедиться, что на каждом аэрофотоснимке не более 2 положительных панорам. В результате этой процедуры было получено 105214 панорам для экспериментов по геолокации. Масштаб спутниковых снимков составляет 20, а разрешение наземной съемки - около 0.114 м. Размеры необработанных изображений для satellite-view и ground-view составляют 640 × 640 и 2048 × 1024 соответственно. В метаданых этого набора используются GPS-координаты.

4) *University-1652* состоит из: 1) снимков, сделанных со спутника; 2) снимков, полученных при помощи БПЛА (drone-view); 3) снимков ground-view для каждого местоположения. Для этого было выбрано 1652 архитектурных сооружений из 72 университетов по всему миру в качестве целевых объектов. Для получения изображений с беспилотника, из-за недоступной стоимости полета в реальном времени,

авторы использовали 3D-модели, предоставляемые Google Earth, для имитации вида с БПЛА. University-1652 состоит из 51355 drone-view, 951 satelite-view, 2921 ground-view.

5) *SUES-200* – это набор данных для задачи CVGL Изображения были собраны из нескольких источников: сделанные со спутников, и соответствующих изображения с БПЛА в 200 местах по всему Шанхайскому университету инженерии и науки. Для того чтобы модель могла распознавать характерные признаки на разных высотах, были собраны снимки, сделанные с помощью дрона с высоты 150, 200, 250 и 300 м. Чтобы предотвратить потерю информации из-за разрешения изображения, снимки с беспилотника в SUES-200 используют исходное разрешение 1080 × 1080, а спутниковые снимки – разрешение 512 × 512. Набор данных включает 200 местоположений с 50 снимками с беспилотника и 1 соответствующим спутниковым изображением для каждого местоположения. Обучающий набор состоит из 24 тыс. drone-view и 120 satelite-view, тестовый набор состоит из 56 тыс. drone-view и 280 stelite-view.

6) GeoText-1652 [14] является расширением набора данных University-1652, используя помимо набора изображений, текстовые описания к ним на разных уровнях, а также метаданные об ограничивающих рамках, где расположены объекты на изображении. Соответствующие текстовые описания были получены при помощи полуавтоматической процедуры аннотирования, в результате которой были получены 276045 экземпляров описаний для органичивающих рамок и 316335 общих описаний. Подробные текстовые аннотации представлены для каждого изображения в виде 3 глобальных описаний для 2.62 ограничивающих рамок, поскольку были удалены некоторые ограничивающие рамки низкого качества. В частности, каждое глобальное описание, включающее сведения как на уровне изображения, так и на уровне региона, содержит в среднем 70.23 слова. Описания на уровне региона, полученные для совпадений с ограничивающими рамками, содержат в среднем 21.6 слова.

Таблица 2 – Общая характеристика набора GeoText

Dataset	Images	Image-level	Region-level
Train drone-view	37854	13562	113367
Train satellite-view	701	2103	1709
Train ground-view	11663	34989	14761
Test drone-view	51355	154065	140179
Test satellite-view	951	2853	2006
Test ground-view	2921	8763	4024

2 Общие проблемы и требования к созданию системы для задачи перекрёстной геолокализации на основе нейросетевого подхода

Ключевые подходы к классификации объектов на аэроснимках, включает использование нейронных сетей, таких как свёрточные нейронные сети (CNN), и методы обучения с учителем для выделения и распознавания объектов [15-19]. Это может включать примеры по классификации построек, дорог, водных объектов и растительности. Будет полезно также затронуть вопросы предварительной обработки данных, например аугментацию снимков для улучшения результатов классификации.

Можно сосредоточиться на нескольких ключевых этапах и аспектах классификации объектов на аэроснимках:

Подготовка данных: идет предварительной обработки изображений. Поскольку аэроснимки часто содержат искажения, важно упомянуть о коррекции яркости, улучшении контраста и фильтрации шума. Также можно рассмотреть аугментацию данных, включая вращение, масштабирование и сдвиги, чтобы повысить обобщающую способность модели [20].

Извлечение признаков. Здесь стоит объяснить, что свёрточные нейронные сети (CNN) извлекают пространственные и текстурные признаки из изображений. Исследуются, как слои свёрточных фильтров улавливают особенности изображений, которые затем используются для классификации. Например, начальные слои могут находить простые границы и углы, в то время как более глубокие слои обнаруживают более сложные объекты, такие как здания или дороги.

Выделение низко-, средне- и высокоуровневых признаков. Низкоуровневые признаки – это простые признаки, которые включают контуры**,** углы**,** текстуры **и** цветовые градиенты. Например, на начальных уровнях свёрточных нейронных сетей (CNN) фильтры распознают только контуры, например, границы зданий или очертания дорог. Эти признаки помогают модели отличить один объект от другого, выявляя общие границы и формы. Среднеуровневые признаки – это признаки более сложных форм, такие как крыши зданий**,** реки или засаженные участки земли**.** Например, нейронная сеть может научиться распознавать крыши благодаря текстуре и форме – они часто прямоугольные и имеют определённый цвет. Дороги, напротив, вытянуты и имеют характерные края. Высокоуровневые признаки извлекаются на самых глубоких уровнях сети (ближе к выходу), и представляют собой комбинацию низко- и среднеуровневых признаков. Они позволяют сети идентифицировать сложные объекты, такие как "жилой квартал" (комбинация домов, дорог, деревьев) или "парк" (растительность, дорожки). Эти признаки уже ближе к семантическому пониманию иерархии объектов на изображении, что полезно, когда необходимо понять контекст и расположение объектов относительно друг друга.

Для этапа выделения признаков целесообразно предварительно обучить базовую сеть для извлечения визуальных признаков (и, как правило, языковую модель для извлечения текстовых признаков) на большом наборе данных для классификации и распознавания изображений.

Учет семантических характеристик объектов: Для задач навигации важно не просто выделить объекты, но и понять их взаиморасположение. Например, дороги обычно соединены между собой, здания находятся на суше, а водоёмы рядом с природной средой. Эти семантические особенности можно интегрировать с помощью гибридных моделей, комбинируя CNN с графовыми нейронными сетями (GNN), чтобы учитывать логические связи и улучшать точность классификации. На этом этапе необходимо выполнить тонкую настройку (fine-tuning), адаптировав нейронную сеть под целевой вид набора данных (например, вид с БПЛА и со спутника).

Исходя из анализа наборов данных для задачи CVGL можно сделать вывод о том, что использование БПЛА для естественного сбора наборов данных является ресурсозатратной задачей, и для ее решения учёные используют вспомогательный инструментарий, такой как Bing Maps, AutoNavi Map, Google Maps, Google Street View, Google Earth.

Дополнительной задачей является разработка дополнительных метаданных, таких как кодирование в виде текстовых описаний положений об угле наклоне камеры, высоты и т.п. В качестве инструментов для аннотирования данных следует использовать предварительно обученные модели для распознавания объектов из открытых словарей (с возможностью тонкой настройки).

Среди рассмотренных наборов данных можно выделить следующие наборы данных:

– Из всех расмотренных наборов данных для предобучения нейросетевых архитектур лучшие показатели демонстрирует Objects 365. Это подтверждается тем, что данный набор полностью аннотирован вручную, содержит достаточное распределение классов и достаточный объем данных, в то же время среднее количество классов и ограничивающих рамок превосходят существующие аналоги, что позволит сделать модель более устойчивой.

– Дополнительно для тестирования и обучения нейросетевых моделей актуальным является использование LVIS, т.к. этот набор данных специально предназначен для тестирования модели на способность распознавать ранее неизвестные классы, а общее число классов в LVIS превышает тысячу.

– Для задач, связанных с дронами, после предварительного обучения нейросетевой модели, следует использовать для обучения и тестирования целевые наборы данных VisDrone и S-ODv2**.**

– DOTA и iSAID для тонкой настройки нейронной сети для распознавания классов рельефа, водных поверхностей, строений, растительности и дорог.

– GeoText в качестве настройки нейронной сети на задачу перекрёстной геолокализации с учётом многомодальных данных.

Стоит отметить, что количество наборов данных для перекрёстной геолокализации, по сравнению с остальными задачами, относительно мало. Это связано со сложностью и недостаточным исследованием данной задачи. При этом эти наборы данных содержат визуальные признаки для таких стран как Китай и США, в то время как для Европы, а в частности для РФ, подобные данные отсутствуют, что делает задачу создания подобных наборов данных актуальной, учитывая востребованность задачи перекрёстной геолокализации. Дополнительной задачей также является разработка перечня ключевых классов, участвующих в перекрёстной геолокализации, а также, как указывалось ранее, стратегии расширения данных для обеспечения робастности (устойчивости к шумам) нейросетевой модели.

Заключение

В данной работе была описана задача перекрёстной геолокализации. Описаны основные наборы данных, применяемые в задачах распознавания объектов, в том числе распознавания таких объектов как здания, сооружения, водные поверхности, рельеф, растительность. Дополнительно описаны наборы данных, используемые непосредственно в задаче перекрёстной геолокализации. Среди всех наборов были выделены следующие: Objects 365, LVIS, VisDrone, DOTA, iSAID и GeoText.

Описаны общие проблемы и требования к созданию наборов данных для задачи перекрёстной геолокализации. Среди основных проблем выделены такие проблемы и требования: 1) предварительная обработка изображений (коррекции яркости, улучшении контраста и фильтрации шума, а также аугментация данных); 2) извлечение признаков (на каком уровне следует извлекать признаки); 3) целесообразность предварительного обучения базовой модели; 4) Учет семантических характеристик объектов; 5) ресурсозатратность сбора наборов данных без использования инструментов для 3D-моделирования; 6) использование дополнительных метаданных; 7) аннотирование данных при помощи тонконастроенной модели распознания объектов из открытого словаря.

Список литературы

1. Ронжин, А.Л. Оптимизация технологической карты допустимых системотехнических решений задачи видеоаналитики аквакультуры [Текст] / А.Л. Ронжин, В.Н., Ле, Н.С. Шувалов// Вестник Южно-Уральского университета. Серия «Математика. Механика. Физика» . – 2024. – № 2(16). – С. 50-58 – ISSN 2075-809X. – DOI 10.14529/mmph240205
2. Ронжин, А. Л. Интеллектуализация и роботизация научного оборудования для междисциплинарных исследований [Текст] / А.Л. Ронжин // Проблемы искусственного интеллекта. – 2024. – № 1(28). – С. 4-10 – ISSN 2413-7383. – DOI ?????.
3. Durgam A. et al. Cross-view geo-localization: a survey //arXiv preprint arXiv:2406.09722. – 2024.
4. Zhang X. et al. Understanding image retrieval re-ranking: A graph neural network perspective //arXiv preprint arXiv:2012.07620. – 2020.
5. Lin J. et al. Joint representation learning and keypoint detection for cross-view geo-localization //IEEE Transactions on Image Processing. – 2022. – Т. 31. – С. 3780-3792.
6. Али Б. Алгоритмы навигации беспилотных летательных аппаратов с использованием систем технического зрения [Текст] / Б. Али., Р.Н. Садеков, В.В. Цодокова //Гироскопия и навигация. – 2022. – Т. 30. – №. 4 (119). – С. 87). – ISSN 0869-7035. – DOI 10.17285/0869-7035.00105
7. Пикалёв, Я.С. Разработка системы нормализации текстовых корпусов [Текст] / Я.С. Пикалёв // Проблемы искусственного интеллекта. – 2022. – № 2(25). – С. 64-78. – ISSN 2413-7383
8. Хакимов, Р. С. К вопросу о разработке системы аннотирования данных для задач компьютерного зрения [Текст] / Р. С. Хакимов, О. Л. Нижникова, М. В. Близно // Проблемы искусственного интеллекта. – 2024. – № 3 (34). – С. 70–79. – ISSN 2413-7383. – DOI 10.24412/2413-7383-2024-3-70-79
9. Gupta A., Dollar P., Girshick R. Lvis: A dataset for large vocabulary instance segmentation //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – С. 5356-5364.
10. Shao S. et al. Objects365: A large-scale, high-quality dataset for object detection //Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – С. 8430-8439. of the IEEE International Conference on Computer Vision. – 2019. – Vols. 2019-October.
11. Waqas Zamir S. et al. isaid: A large-scale dataset for instance segmentation in aerial images //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. – 2019. – С. 28-37.
12. Cao Y. et al. VisDrone-DET2021: The vision meets drone object detection challenge results //Proceedings of the IEEE/CVF International conference on computer vision. – 2021. – С. 2847-2854.
13. Xia G. S. et al. DOTA: A large-scale dataset for object detection in aerial images //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – С. 3974-3983.
14. Chu M. et al. Towards natural language-guided drones: GeoText-1652 benchmark with spatial relation matching //European Conference on Computer Vision. – Springer, Cham, 2025. – С. 213-231.
15. Ермоленко, Т.В. Классификация ошибок в тексте на основе глубокого обучения [Текст] / Т.В. Ермоленко // Проблемы искусственного интеллекта. – 2019. –№ 3(14). – С. 47-57.. – ISSN 2413-7383
16. Зуев, В. М. Сравнение обнаружения объектов средствами искусственного интеллекта в сравнении с классическими методами [Текст] / Зуев В. М. // Проблемы искусственного интеллекта. – 2024. – № 3(34). –. С. 4-10 – ISSN 2413-7383. – DOI 10.24412/2413-7383-2024-3-30-35.
17. Пикалёв, .Я. С. О нейронных архитектурах извлечения признаков для задачи распознавания объектов на устройствах с ограниченной вычислительной мощностью [Текст] / Я.С. Пикалёв, Т.В. Ермоленко // Проблемы искусственного интеллекта.. – 2023. – № 3(30). –. С. 44-54 – ISSN 2413-7383. – DOI 10.34757/2413-7383.2023.30.3.004
18. Павленко, Б. В. Интеллектуально-алгоритмический метод калибровки прицелов [Текст] / Б. В. Павленко, В. И. Бондаренко // Проблемы искусственного интеллекта. – 2024. – № 3 (34). – С. 55–63. – ISSN 2413-7383. – DOI 10.24412/2413-7383-2024-3-55-63
19. Кришнан, Ш. Р.. Улучшение обнаружения аномалий на видео с помощью усовершенствованной технологии UNET и техники каскадного скользящего окна. [Текст] / Ш. Р. Кришнан, П. Амудха //Информатика и автоматизация. – 2024. – № 6 (23). – С. 1899-1930. – ISSN 2713-3192. –DOI 10.15622/ia.23.6.12
20. Сойфер, В. А. Калмановская фильтрация одного класса изображений динамических объектов [Текст] / В. А. Сойфер, В. А. Фурсов, С. И. Харитонов //Информатика и автоматизация. – 2024. – №. 4.(23) – С. 953-968. – ISSN 2713-3192. –DOI 10.15622/ia.23.4.1

References

1. Ronzhin, A.L. Optimizaciya tehnologicheskoj karty dopustimyh sistemotehnicheskih reshenij zadachi videoanalitiki akvakultury [Text] / A.L. Ronzhin, V.N., Le, N.S. Shuvalov // «Bulletin of the South Ural State University». Ser. «Mathematics. Mechanics. Physics». – 2024. – № 2(16). – P. 50-58 – ISSN 2075-809X. – DOI 10.14529/mmph240205
2. Ronzhin, A. L. Intellektualizaciya i robotizaciya nauchnogo oborudovaniya dlya mezhdisciplinarnyh issledovanij [Text] / A.L. Ronzhin // Problems of Artificial Intelligence. – 2024. – № 1(28). – P. 4-10 – ISSN 2413-7383. – DOI ?????.
3. Durgam A. et al. Cross-view geo-localization: a survey //arXiv preprint arXiv:2406.09722. – 2024.
4. Zhang X. et al. Understanding image retrieval re-ranking: A graph neural network perspective //arXiv preprint arXiv:2012.07620. – 2020.
5. Lin J. et al. Joint representation learning and keypoint detection for cross-view geo-localization //IEEE Transactions on Image Processing. – 2022. – T. 31. – C. 3780-3792.
6. Ali B. Algoritmy navigacii bespilotnyh letatelnyh apparatov s ispolzovaniem sistem tehnicheskogo zreniya [Text] / B. Ali., R.N. Sadekov, V.V. Tsodokova // Gyroscopy and navigation. – 2022. – Vol. 30. – №. 4 (119). – P. 87. – ISSN 0869-7035. – DOI 10.17285/0869-7035.00105
7. Pikalyov, Ya.S. Razrabotka sistemy normalizacii tekstovyh korpusov [Text]/ Ya.S. Pikalyov // Problems of Artificial Intelligence. – 2022. – № 2(25). – P. 64-78. – ISSN 2413-7383. – DOI ?????.
8. Khakimov R.S. K voprosu o razrabotke sistemy annotirovaniya dannyh dlya zadach kompyuternogo zreniya [Text] / R.S. Khakimov, O.L. Nizhnikova, M.V. Blizno // Problems of Artificial Intelligence. – 2024. – № 3 (34). – P. 70–79. – ISSN 2075-1087. – DOI 10.24412/2413-7383-2024-3-70-79
9. Gupta A., Dollar P., Girshick R. Lvis: A dataset for large vocabulary instance segmentation //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – C. 5356-5364.
10. Shao S. et al. Objects365: A large-scale, high-quality dataset for object detection //Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – C. 8430-8439. of the IEEE International Conference on Computer Vision. – 2019. – Vols. 2019-October.
11. Waqas Zamir S. et al. isaid: A large-scale dataset for instance segmentation in aerial images //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. – 2019. – C. 28-37.
12. Cao Y. et al. VisDrone-DET2021: The vision meets drone object detection challenge results //Proceedings of the IEEE/CVF International conference on computer vision. – 2021. – C. 2847-2854.
13. Xia G. S. et al. DOTA: A large-scale dataset for object detection in aerial images //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – C. 3974-3983.
14. Chu M. et al. Towards natural language-guided drones: GeoText-1652 benchmark with spatial relation matching //European Conference on Computer Vision. – Springer, Cham, 2025. – C. 213-231.
15. Yermolenko, T.V. Klassifikaciya oshibok v tekste na osnove glubokogo obucheniya [Text]/ T.V. Yermolenko // Problems of Artificial Intelligence. – 2019. –№ 3(14). – P. 47-57.. – ISSN 2413-7383
16. Zuev V.M. Sravnenie obnaruzheniya obektov sredstvami iskusstvennogo intellekta v sravnenii s klassicheskimi metodami [Text] / Zuev V.M. // Problems of Artificial Intelligence. – 2024. – № 3(34). –. P. 4-10 – ISSN 2413-7383. – DOI 10.24412/2413-7383-2024-3-30-35.
17. Pikalyov, Ya.S. O nejronnyh arhitekturah izvlecheniya priznakov dlya zadachi raspoznavaniya obektov na ustrojstvah s ogranichennoj vychislitelnoj moshnostyu [Text] / Ya.S. Pikalyov, T.V. Yermolenko // Problems of Artificial Intelligence. – 2023. – № 3(30). –. P. 44-54 – ISSN 2413-7383. – DOI 10.34757/2413-7383.2023.30.3.004
18. Pavlenko, B.V. Intellektualno-algoritmicheskij metod kalibrovki pricelov [Text] / B.V.Pavlenko, V.I. Bondarenko // Problems of Artificial Intelligence. – 2024. – № 3 (34). – P. 55–63. – ISSN 2413-7383. – DOI 10.24412/2413-7383-2024-3-55-63
19. Krishna, Sh. R. Uluchshenie obnaruzheniya anomalij na video s pomoshyu usovershenstvovannoj tehnologii UNET i tehniki kaskadnogo skolzyashego okna [Text] / Sh. R. Krishann, P. Amudha //Informatics and automatisation. – 2024. – № 6 (23). – P. 1899-1930. – ISSN 2713-3192. –DOI 10.15622/ia.23.6.12
20. Soyfer, V.A., Kalmanovskaya filtraciya odnogo klassa izobrazhenij dinamicheskih obektov [Text] / V.A. Soyfer, V.A. Fursov, S.I. Kharitonov // Informatics and automatisation. – 2024. –№. 4.(23) – P. 953-968. – ISSN 2713-3192. –DOI 10.15622/ia.23.4.1

RESUME

Ya. S. Pikalyov

Key objects detection and cross-view geolocalization: Dataset analysis and methodological aspects

The article addresses the challenges of creating a system for cross-view geo-localization using a neural network-based approach. This area is particularly relevant because such systems enable UAVs to navigate complex environments, identify objects, obstacles, and routes, thereby reducing dependence on operators. Additionally, recognition systems allow UAVs to collect and analyze data effectively.

–The article also describes datasets specifically used for cross-view geo-localization. Key points regarding these datasets include:

–Objects365 delivers the best performance for pretraining neural network architectures.

–LVIS is recommended for testing and training neural network models.

–Specialized datasets like VisDrone and S-ODv2 should be used for drone-related tasks after model pretraining.

–DOTA and iSAID are suitable for fine-tuning models for tasks involving terrain recognition, water surfaces, buildings, vegetation, and roads.

–GeoText is ideal for cross-view geo-localization tasks using multimodal data.

– Using UAVs for natural data collection is a resource-intensive process. To optimize this, tools such as Bing Maps, AutoNavi Map, Google Maps, Google Street View, and Google Earth are extensively utilized.

Currently, datasets for cross-view geo-localization are limited, reflecting the complexity and insufficient exploration of this task. These datasets primarily cover visual features for China and the United States, while similar data for Europe, particularly Russia, is lacking. This highlights the importance of creating analogous datasets for these regions.

Promising research directions include:

–Developing a list of key classes critical for cross-view geo-localization.

–Designing data expansion strategies to enhance the robustness of neural network models.

РЕЗЮМЕ

Я. С. Пикалёв

Обнаружение ключевых объектов и перекрёстная геолокализация: Анализ наборов данных и методологические перспективы

В статье рассмотрена проблема создания системы для задачи перекрёстной геолокализации на основе нейросетевого подхода. Это направление актуально в связи с тем, что эти системы помогают БПЛА ориентироваться в сложных условиях, идентифицируя объекты, препятствия и маршруты, что позволяет снизить зависимость от операторов. К тому же, используя системы распознавания, БПЛА могут осуществлять сбор и анализ данных. Дополнительно описаны наборы данных, используемые непосредственно в задаче перекрёстной геолокализации. Среди рассмотренных наборов данных можно выделить следующие ключевые моменты:

- Objects 365 демонстрирует лучшие результаты для предобучения нейросетевых архитектур.
- Для тестирования и обучения нейросетевых моделей рекомендуется LVIS.

- Для задач с использованием дронов после предобучения модели следует применять специализированные наборы данных, такие как VisDrone и S-ODv2.
- Для тонкой настройки моделей на задачи распознавания рельефа, водных поверхностей, строений, растительности и дорог актуальны наборы DOTA и iSAID.
- GeoText подходит для задач перекрестной геолокации с использованием многомодальных данных.

Кроме того, использование БПЛА для естественного сбора данных является ресурсоемким процессом. Для его оптимизации активно применяются инструменты вроде Bing Maps, AutoNavi Map, Google Maps, Google Street View и Google Earth.

Наборы данных для перекрестной геолокации представлены в ограниченном количестве, что связано со сложностью и недостаточной исследованностью данной задачи. Эти данные в основном охватывают визуальные признаки для Китая и США, тогда как для Европы, в частности России, такие данные отсутствуют. Это подчеркивает актуальность создания аналогичных наборов для указанных регионов.

Перспективными направлениями исследования являются:

- Разработка перечня ключевых классов, важных для перекрестной геолокации.
- Разработка стратегий расширения данных для повышения робастности нейросетевых моделей.

Пикалёв Ярослав Сергеевич – кандидат техн. наук, старший научный сотрудник лаборатории интеллектуальных систем и анализа данных, Федерального государственного бюджетного научного учреждения «Институт проблем искусственного интеллекта», г. Донецк. *Область научных интересов:* Цифровая обработка сигналов, анализ данных, распознавание образов, обработка естественного языка, компьютерное зрение, машинное обучение, нейронные сети, эл. почта i@pikaliov.ru, адрес: 283085, ДНР, г. Донецк, ул. Отважных, д. 19, кв.85, телефон: +7949 4287388.

Статья поступила в редакцию 03.06.2024.