

Т. В. Ермоленко^{1,2}, Р. С. Хакимов¹

¹Федеральное государственное бюджетное научное учреждение «Институт проблем искусственного интеллекта», г. Донецк 283048, г. Донецк, ул. Артема, 118 б

²Федеральное государственное бюджетное образовательное учреждение высшего образования «Донецкий государственный университет» г. Донецк, 283001, г. Донецк, ул. Университетская, 24

К ВОПРОСУ О ПРИМЕНЕНИИ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ЗАДАЧИ ПЕРЕКРЁСТНОЙ ГЕОЛОКАЛИЗАЦИИ¹

T. V. Yermolenko^{1,2}, R.S. Khakimov¹

¹Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems» 283048, Donetsk, Artema str, 118-b

²Federal State Educational Institution of Higher Education «Donetsk State University» 283001, Donetsk, University st, 24

ON THE QUESTION OF APPLICATION OF DEEP LEARNING FOR THE PROBLEM OF CROSS-VIEW GEOLOCALIZATION

В статье рассмотрена задача перекрёстной геолокализации с математической точки зрения, выделены основные подходы, проблемы и ключевые особенности применения глубокого обучения. Среди ключевых особенностей применения глубокого обучения можно выделить: 1) архитектуры на основе Трансформера позволяют более четко идентифицировать характерные объекты и игнорировать фоновую информацию; 2) результаты модели на основе Трансформера, обученной дважды, будут иметь большой разброс; 3) масштаб изображений со спутника фиксирован, в то время как высота полёта БПЛА динамически изменяется.

Ключевые слова: перекрёстная геолокализация, компьютерное зрение, машинное обучение, нейронные сети, БПЛА.

The article considers the problem of cross-geolocation from a mathematical point of view, highlights the main approaches, problems and key features of the application of deep learning. Problems such as 1) similarity of key objects in detail and uniform appearance are indicated; 2). image styles can vary greatly; 3) the average pool used in convolutional neural networks ignores the interaction between local objects. Among the key features of the application of deep learning are: 1) Transformer-based architectures make it possible to more clearly identify characteristic objects and ignore background information; 2) the results of a Transformer-based model trained twice will have a large spread; 3) the scale of satellite images is fixed, while the altitude of the UAV flight changes dynamically.

Keywords: cross-view geolocation, computer vision, machine learning, neural networks, UAVs.

¹ Исследование выполнено в рамках темы «Извлечение семантической информации из изображений для автономных систем навигации беспилотных летательных аппаратов» молодежной лаборатории ЛИСАД ФГБНУ «ИПКИ» (регистрационный номер проекта в ЕГИСУ НИОКТР 1023111200001-6-2.2.1), финансируемого из федерального бюджета в рамках национального проекта «Наука и университеты».

Введение

В последние годы беспилотные летательные аппараты (БПЛА) нашли широкое применение в таких сферах как сельскохозяйственная промышленность, аэрофото-съемка, логистика, навигация, обнаружение событий и т.п. Все вышеописанные сферы применения используют механизм геолокации по аэрофотоснимкам, который может помочь в различных ситуациях, когда система глобального позиционирования (Global Positioning System, GPS) либо недоступна, либо датчики стоят дорого, либо датчик GPS принимает зашумленные сигналы [1-3]. Эта проблема стоит остро особенно в городах, т.к. эффект городского каньона может привести к отклонениям в несколько метров.

Альтернативой спутниковой радионавигации для высокой точности определения местоположения является перекрестная локализация изображений (*cross-view geo-localization, CVGL*) [4-6], при которой для определения местоположения используется взаимосвязь между различными видами изображения, например, снимками с БПЛА и спутника. При этом полученные изображения сравниваются с базой данных спутниковых изображений с известными географическими позициями. Кроме этого, основным преимуществом CVGL является получение более детальной и точной информации об окружающей среде в определенной области.

Постановка задачи

Цель данной работы заключается в определении основных проблем в задаче перекрестной геолокации, а также в выявлении ключевых особенностей применения глубокого обучения к указанной выше задаче.

В работе были поставлены следующие задачи:

- 1) Дать математическое описание задаче CVGL
- 2) Провести анализ основных подходов CVGL
- 3) Выявить основные проблемы в задаче CVGL
- 4) Вывести ключевые особенности применения глубокого обучения к CVGL

В то время как классические подходы CVGL используют визуальные подсказки, такие как положение солнца, тени или погода, современные подходы все больше используют подходы на основе глубокого обучения.

В качестве входных данных CVGL, как правило, является: 1) снимок, полученный посредством БПЛА; 2) предзагруженная (сегментированная) карта местности. Дополнительно информация может быть дополнена ключевыми точками местности или описанием состояния БПЛА. Выходные данные представлены в виде положения БПЛА в пространстве в виде 3 или 6 степеней свободы: координаты (x, y, z) и углы ориентации (ϕ, θ, ψ) .

Задачи в CVGL условно можно разделить на два типа.

1. *Определение местоположения объекта с помощью дрона (drone view target localization)*. Задача состоит в том, чтобы найти наиболее похожее изображение со спутника, чтобы определить местоположение целевого здания на снимке.

2. *Навигация с помощью дрона (drone navigation)*. По одному изображению, полученному со спутника, БПЛА намеревается найти наиболее подходящее место (изображения, полученные с помощью БПЛА), мимо которого он пролетал. В соответствии с историей полета, БПЛА может вернуться к заданному месту.

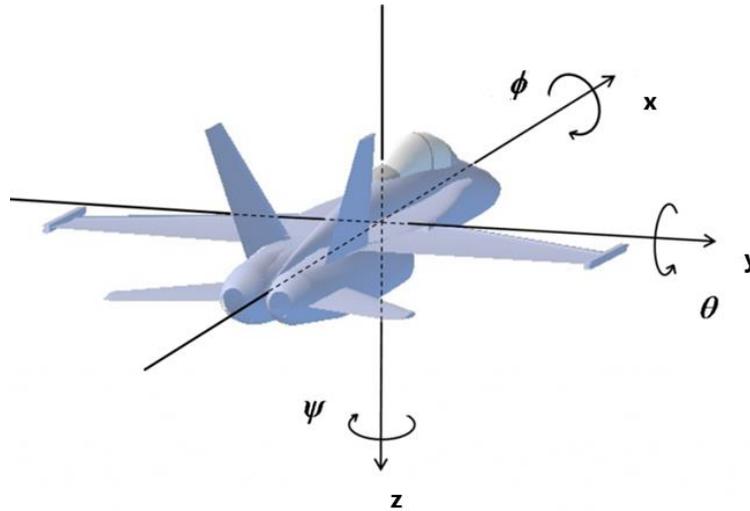


Рисунок 1 – Навигационные параметры БПЛА

1 Математическое описание задачи перекрёстной геолокализации

В математической форме задача геолокализации может быть описана следующим образом.

Входные данные:

z_{uav} – высота полета БПЛА, для измерения которой используется барометр;

$\phi_{uav}, \theta_{uav}, \psi_{uav}$ – углы ориентации БПЛА, полученные от блока инерциальных измерителей;

$I_{uav}(x_{uav}, y_{uav})$ – текущее изображение, снимок местности, полученный с БПЛА на высоте z_{uav} ;

$G = \{R_i(x_i, y_i, \Delta_i)\}$ – набор эталонных изображений, заданных в географических координатах (x_i, y_i) с разрешением Δ_i , определяющем масштаб изображения R_i . Такими изображениями могут быть карты, спутниковые снимки или заранее отснятые данные с известными геокоординатами;

T_{uav} – матрица трансформации, учитывающая высоту z_{uav} и углы ориентации БПЛА $\phi_{uav}, \theta_{uav}, \psi_{uav}$, необходимые для приведения снимка БПЛА к масштабу эталонов.

Промежуточные данные:

$I_{uav}^{Tr} = T(\Delta_i, I_{uav}, \phi_{uav}, \theta_{uav}, \psi_{uav}, z_{uav})$ – трансформированное изображение, полученное из изображения I_{uav} с помощью аффинных преобразований и масштабирования.

Целевая функция:

$S(x_i, y_i) = \text{similar}(R_i(x_i, y_i, \Delta_i), I_{uav}^{Tr})$ – функция подобия для оценки схожести двух изображений I_{uav}^{Tr} и R_i с координатами (x_i, y_i) , т.е. оценка соответствия текущих координат (x_{uav}, y_{uav}) координатам эталонного изображения R_i .

Выходные данные:

Географические координаты (x^*, y^*) эталонного изображения R^* , при котором функция $S(x_i, y_i)$ достигает максимума:

$$(x^*, y^*) = \arg \max_G S(x_i, y_i)$$

Основная трудность при разработке таких методов и алгоритмов заключается в том, что сопоставляемые изображения (текущий снимок и эталонные изображения) существенно различаются вследствие различных условий съемки (пространственное положение БПЛА что влияет на положение камеры и изменение масштаба, а также на положение целевого объекта на изображении; перекрытие одних объектов другими (окклюзии); ракурс; бортовые и атмосферные помехи; время суток, что влияет на освещение; сезонные изменения местности; аффинные преобразования; появление новых объектов на изображении и т.д.). Для решения этих проблем используются несколько подходов.

2. Подходы к решению задач перекрестной геолокализации

Рассматриваемые алгоритмы навигации на основе перекрестной геолокализации можно отнести к одному из трех различных подходов:

- корреляционно-экстремальный;
- использование ключевых точек;
- использование нейронных сетей как единственного элемента алгоритма сопоставления с эталоном.

Корреляционно-экстремальные методы относятся к самым ранним. Название «корреляционно-экстремальный» обусловлено использованием при сопоставлении изображений (эталонного и формируемого на борту БПЛА) критерия в виде корреляционной функции, т.е. основная идея заключается в поиске максимума взаимной корреляционной функции изображений [7-9]. При этом, как правило, сопоставляются цельные изображения попиксельно.

Для более надежного сравнения изображений могут использоваться контуры, области, точки и другие их характерные элементы, извлекаемые на основе как классических, так и нейросетевых алгоритмов обработки изображений [10-16].

На рис. 2 приведена типовая схема работы алгоритмов этого класса.

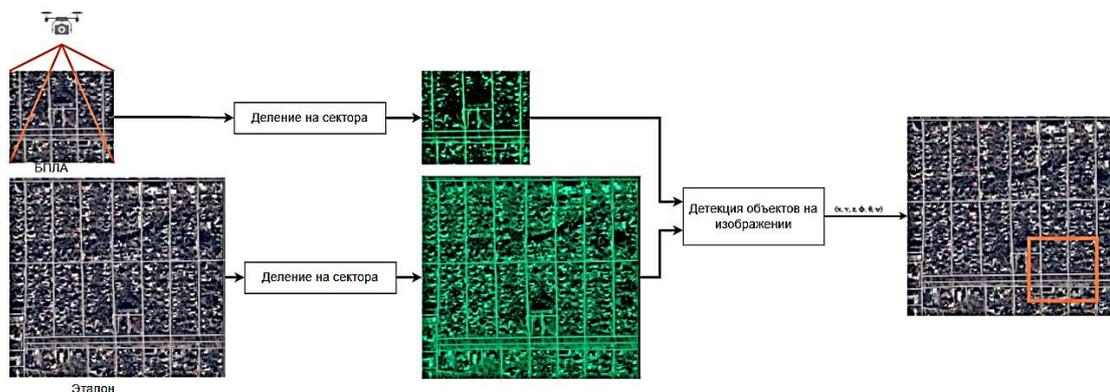


Рисунок 2 – Схема работы корреляционно-экстремального метода

Методы с использованием ключевых точек, в которых изображения сопоставляются по픽сельно работают приемлемо, практически, только в идеальных ситуациях. Причины – различные условия съемки. Собственно, плохие качества этих методов обусловлены их основной идеей, т.е. тем, что в признаки изображения вносит вклад каждая точка изображения, каким бы плохим этот вклад не был.

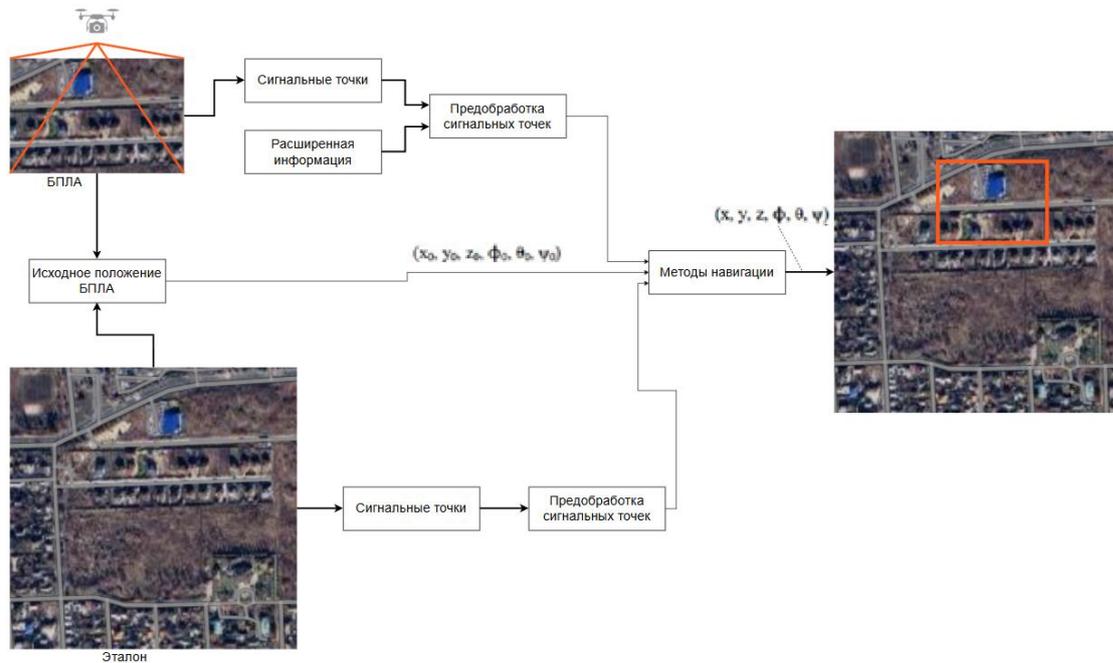


Рисунок 3 – Схема работы методов навигации на основе сопоставления изображений с использованием ключевых точек

Для обхода таких проблем нужно либо выбирать точки, вносящие вклад в признаки, либо, ещё лучше, выделять некоторые особые (ключевые) точки и сравнивать их. Суть состоит в том, что мы заменяем изображение некоторой моделью – набором его ключевых точек. Ключевой точкой считается такая точка, которая имеет некие признаки, существенно отличающие ее, от основной массы точек. Например, это могут быть края линий, небольшие круги, резкие перепады освещенности, углы и т.д. Предполагая, что ключевые точки присутствуют на эталоне всегда, можно поиск эталона свести к поиску на сцене его ключевых точек. А поскольку ключевые точки сильно отличаются от основной массы точек, то их число будет существенно меньше, чем общее число точек эталона. Схема работы этой группы методов представлена на рис. 3.

Методы с использованием нейронных сетей как единственного элемента алгоритма сопоставления с эталоном имеют в своей основе глубокую нейронную сеть, которая, получая на вход изображение, формирует на выходе координаты и параметры ориентации БПЛА. Обучение нейронной сети проводится на наборе имеющихся геопривязанных изображений или изображениях, для которых точно известно местоположение и ориентация объекта, при этом минимизируется взвешенная сумма квадратов разностей координат и параметров ориентации – предсказанных нейронной сетью и истинных (соответствующих набору имеющихся изображений). По полученным на этапе обучения весам сети осуществляется навигация (рис. 4).

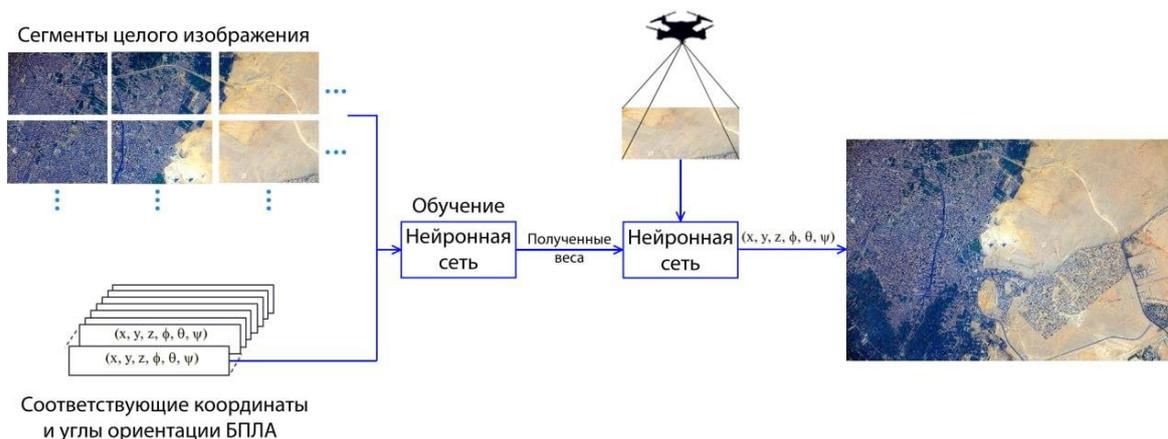


Рисунок 4. Схема алгоритма навигации на основе сопоставления изображений с использованием нейронных сетей

3. Основные проблемы в задаче CVGL

1) Как было указано ранее, ключевой целью, лежащей в основе CVGL, является извлечение отличительных признаков, что остается сложной задачей из-за небольших различий между классами. В частности, это связано с тем, что большинство ключевых объектов имеют схожие детали и однородный внешний вид.

2) CVGL является сложной задачей, поскольку точки обзора между снимками, сделанными с БПЛА, и спутниковыми снимками различны. Кроме того, стили изображений могут сильно отличаться, поскольку они получены из разных источников (доменов).

3) Методы глубокого обучения рассматривают CVGL как задачу поиска изображений (image retrieval), и их цель состоит в том, чтобы приблизить соответствующую пару изображений друг к другу в пространстве высокой размерности. Большинство традиционных подходов сосредоточено на применении механизмов пространственного внимания для извлечения географических объектов. Эти функции описывают основную цель или контекстуальную информацию о цели, но не учитывают дополнительную информацию, получаемую от БПЛА. При этом, данные о полете БПЛА имеют большое значение для захватываемого изображения. Кроме того, для извлечения глобальных объектов после базовой сети используется среднестатистический пул, игнорирующий взаимодействие между локальными объектами, расположенными между уровнями, и глобальными объектами. Эти недостатки осложняют разработку подобных систем с различными сценами и условиями окружающей среды.

4) Большинство традиционных наборов данных содержат только 360-градусные изображения с видом на окружение (street- или ground-view). Этот факт усложняет CVGL в тех случаях, когда необходимо сопоставлять объекты, не привязанные к ориентации в пространстве.

5) Дополнительной проблемой является идеальное согласование положения изображения с центром спутникового снимка. Для разрешения этой проблемы требуются дополнительные наборы данных с явно не выровненными центральными позициями и неизвестной географической ориентацией между изображением со спутника и ground-view. Кроме того, угол обзора не должен превышать 120 градусов.

6) Общей характеристикой существующих наборов данных является сосредоточенность на городской среде. Т.е. эти наборы не отражают разнообразие ландшафтов. Поэтому необходимы такие наборы данных, которые содержали бы различные виды местности, чтобы увеличить разнообразие, полезность и практичность подобных данных [17].

4. Ключевые особенности глубокого обучения для CVGL

1) По сравнению с архитектурами CVGL, основанными на свёрточных нейронных сетях (Convolutional Neural Network, CNN), архитектуры на основе Трансформера позволяют более четко идентифицировать характерные объекты, такие как здания и дороги, и игнорировать фоновую информацию, такую как деревья.

2) Большинство методов встраивают механизм внимания в глубокую сверточную сеть, что в определенной степени усиливает контекстуальные связи [18].

3) Детализированная информация (высокоинформативные признаки) очень важна для поиска информации. Операции понижающей выборки, т.е. объединение в пул и пошаговая свертка метода, основанного на CNN, могут снизить разрешение изображения, в то же время меняя структуру детализированной информации.

4) CVGL используется для поиска соответствующей информации в разных контекстах. Изображения из разных областей могут изменяться в зависимости от положения, например, поворачиваться, масштабироваться и смещаться. Поэтому необходимо полностью понимать семантическую информацию глобального контекста. Методы, основанные на CNN, фокусируются на небольших областях распознавания из-за гауссовского распределения эффективных рецептивных полей.

5) Учитывая ограничения методов, основанных исключительно на CNN, модули были сосредоточены на детализированной информации о различных частях, что помогает модели изучать более полные характеристики. Кроме того, при разделении карт объектов и контроле за ними будут полностью выделены второстепенные объекты на изображении. Мелкозернистые области могут быть созданы вручную человеком, но также могут быть автоматически изучены с помощью контролируемых методов. А детализированные функции, основанные на деталях, доказали свою надежность в задаче поиска [19].

6) У моделей на основе Трансформера есть отрицательные стороны. Например, если модель с одинаковыми настройками будет обучена дважды, полученные результаты будут иметь большой разброс. Основная причина заключается в том, что в снимках со спутника есть только одно изображение для каждой категории, в результате чего одновременно отображается только одно изображение из других видов. В этом случае будет наблюдаться дисбаланс между спутниковыми снимками и другими изображениями. Поэтому для решения проблемы несбалансированности выборки необходимо применять стратегию множественной выборки. Т.е. необходимо задать гиперпараметр k , означающий количество выборок. После обрабатываем изображение со спутника и генерируем k расширенных спутниковых изображений. Методы расширения (аугментации) [20] включают в себя случайное смещение, случайное вырезание, случайное улучшение цвета и т.п. В то же время случайным образом выбираются k изображений с других ракурсов, которые относятся к той же категории, что и соответствующий спутниковый ракурс.

7) Зависимость от размера входного изображения. Изображение небольшого размера сжимает мелкозернистую информацию и искажает все характеристики исходного изображения. Крупномасштабные изображения часто позволяют добиться более высокой точности, поскольку они сохраняют исходную мелкозернистую информацию. Напротив, крупномасштабные изображения часто требуют больших ресурсов памяти и более длительного времени вывода во время обучения и тестирования.

8) Влияние высоты полёта БПЛА. Масштаб изображений со спутника фиксирован, в то время как высота полёта БПЛА динамически изменяется. В соответствии с этим необходимо наличие в наборе данных снимков с разных высот: большой, средней и низкой для обеспечения возможности модели для адаптации к масштабированию.

Заключение

В работе приведено математическое описание задачи CVGL: описаны входные, промежуточные и выходные данные, а также целевая функция. Было рассмотрено три подхода в задаче CVGL: 1) корреляционно-экстремальный; 2) использование ключевых точек; 3) использование нейронных сетей как единственного элемента алгоритма сопоставления с эталоном. Можно сделать вывод, что актуально использовать гибридный подход на основе последних двух вышеперечисленных. Рассмотрены основные проблемы и ключевые особенности CVGL. На основе этого были предложены решения некоторых проблем. 1) Результаты модели на основе Трансформер, обученной дважды, имеют большой разброс. Для решения этой проблемы необходимо при обработке изображения со спутника генерировать k расширенных спутниковых изображений и выбирать k изображений с других ракурсов, которые относятся к той же категории. 2) Масштаб изображений со спутника фиксирован, в то время как высота полёта БПЛА динамически изменяется. В соответствии этим необходимо наличие в наборе данных снимков с разных высоты: большой, средней и низкой для обеспечения возможности модели для адаптации к масштабированию. 3) Подходы в CVGL сосредоточены на применении механизмов пространственного внимания, описывая основную цель или контекстуальную информацию о цели, но не учитывают дополнительную информацию, получаемую от самого БПЛА. Для разрешения этой проблемы необходимо передавать модели данные о полете БПЛА (например, в виде текстовых шаблонов).

Список литературы

1. Deuser F. Orientation-Guided Contrastive Learning for UAV-View Geo-Localisation / F. Deuser, K. Habel, M. Werner, N. Oswald // UAVM 2023 - Proceedings of the 2023 Workshop on UAVs in Multimedia: Capturing the World from a New Perspective, Co-located with MM 2023. – 2023. – P. 7-11.
2. М. Н. Фаворская, А. И. Пахирка, “Восстановление аэрофотоснимков сверхвысокого разрешения с учетом семантических особенностей”, Информатика и автоматизация, 23:4 (2024), 1047–1076
3. Understanding Image Retrieval Re-Ranking: A Graph Neural Network Perspective / X. Zhang, M. Jiang, Z. Zheng, X. Tan, E. Ding, Y. Yang // Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2020. — P. 4911–4920. — DOI: 10.1109/CVPR42600.2020.00496.
4. Lin J. Joint Representation Learning and Keypoint Detection for Cross-View Geo-Localization / J. Lin, Z. Zheng, Z. Zhong, Z. Luo, S. Li, Y. Yang, N. Sebe // IEEE Transactions on Image Processing. – 2022. – Vol. 31. – № May. – P. 3780-3792.
5. Deuser F. Sample4Geo: Hard Negative Sampling For Cross-View Geo-Localisation / F. Deuser, K. Habel, N. Oswald // Proceedings of the IEEE International Conference on Computer Vision. – 2023. – P. 16801-16810.
6. Ронжин А.Л. Оптимизация технологической карты допустимых аквакультуры / А.Л. Ронжин, В.Н. Ле, Н. Шувалов. – 2024. – Том. 16. – № 2. – С. 50-58.
7. Fam K.B. Fuzzy Model of Diagnostics of Technological Processes / K.B. Fam, P.M. Murashev, V.N. Bogatkov // Pattern Recognition and Image Analysis. – 2024. – Vol. 34. – № 3. – P. 632-638.
8. Дворников, С. В., & Васильева, Д. В. (2024). Повышение достоверности выявления аномалий на изображениях при формировании их векторов признаков в базисах вейвлетов. *Информатика и автоматизация*, 23(6), 1698-1729. <https://doi.org/10.15622/ia.23.6.5>
9. Сойфер, В. А., Фурсов, В. А., & Харитонов, С. И. (2024). Калмановская фильтрация одного класса изображений динамических объектов. *Информатика и автоматизация*, 23(4), 953-968. <https://doi.org/10.15622/ia.23.4.1>
10. Верхотуров, А. Л., Степанов, А. С., & Илларионова, Л. В. (2024). Использование радиолокационных данных для мониторинга состояния посевов сельскохозяйственных культур на юге Дальнего Востока России. *Информатика и автоматизация*, 23(4), 1221-1245. <https://doi.org/10.15622/ia.23.4.11>

11. Dai M. A Transformer-Based Feature Segmentation and Region Alignment Method for UAV-View Geo-Localization / M. Dai, J. Hu, J. Zhuang, E. Zheng // *IEEE Transactions on Circuits and Systems for Video Technology*. – 2022. – Vol. 32. – № 7. – P. 4376-4389.
12. 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.
13. 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
14. 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
15. Кришнан, Ш. Р.. Улучшение обнаружения аномалий на видео с помощью усовершенствованной технологии UNET и техники каскадного скользящего окна. / Ш. Р. Кришнан, П. Амудха // *Информатика и автоматизация*. – 2024. – № 6 (23). – С. 1899-1930. – ISSN 2713-3192. – DOI 10.15622/ia.23.6.12
16. Али Б.. Алгоритмы навигации беспилотных летательных аппаратов с использованием систем технического зрения / Б. Али, Р. Н. Садеков, В. В. Цодокова // *Гироскопия и навигация*. – Том 30 – №4 (119) – С. 87-105, 2022
17. 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
18. Ронжин А.Л. Интеллектуализация и роботизация отечественного научного оборудования для междисциплинарных исследований/ А.Л. Ронжин. – 2023. – Том. 1. – № 28. – С. 4-10.
19. Durgama A., Paheding S., Dhiman V., Devabhaktuni V. Cross-view geo-localization: a survey // *Ad Hoc Networks*. – 2021. – Vol. 29, no. 3. – P. 1519–1541. – DOI: 10.1016/j.adhoc.2021.101843.
20. Хакимов, Р. С. Обзор расширенных техник аугментации для набора данных изображений [Текст] / Р. С. Хакимов, Б. В. Павленко, Я. С. Пикалев // *Донецкие чтения 2024: образование, наука, инновации, культура и вызовы современности: Материалы IX Международной научной конференции (Донецк, 15–17 октября 2024 г.)*. – Т. 2: Физические, химические, технические и компьютерные науки. Часть 2 / Под общей редакцией проф. С.В. Беспаловой. – Донецк : Изд-во ДонГУ, 2024. – 296 с. – С. 272–275. – ISSN 2664-7362 (Print); ISSN 2664-7370 (Online)

References

1. Deuser F. Orientation-Guided Contrastive Learning for UAV-View Geo-Localization / F. Deuser, K. Habel, M. Werner, N. Oswald // *UAVM 2023 - Proceedings of the 2023 Workshop on UAVs in Multimedia: Capturing the World from a New Perspective, Co-located with MM 2023*. – 2023. – P. 7-11.
2. Favorskaia M. N., Pakhirka A. I. Restoration of Ultra-High-Resolution Aerial Images Considering Semantic Features // *Informatics and Automation*. – 2024. – Vol. 23. – No. 4. – P. 1047–1076.
3. Zhang X. Understanding Image Retrieval Re-Ranking: A Graph Neural Network Perspective / X. Zhang, M. Jiang, Z. Zheng, X. Tan, E. Ding, Y. Yang // *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2020. – P. 4911–4920. – DOI: 10.1109/CVPR42600.2020.00496.
4. Lin J. Joint Representation Learning and Keypoint Detection for Cross-View Geo-Localization / J. Lin, Z. Zheng, Z. Zhong, Z. Luo, S. Li, Y. Yang, N. Sebe // *IEEE Transactions on Image Processing*. – 2022. – Vol. 31. – No. May. – P. 3780-3792.
5. Deuser F. Sample4Geo: Hard Negative Sampling for Cross-View Geo-Localization / F. Deuser, K. Habel, N. Oswald // *Proceedings of the IEEE International Conference on Computer Vision*. – 2023. – P. 16801-16810.
6. Ronzhin A. L. Optimization of Technological Map of Permissible Aquaculture / A. L. Ronzhin, V. N. Le, N. Shuvalov // 2024. – Vol. 16. – No. 2. – P. 50–58.
7. Fam K. B. Fuzzy Model of Diagnostics of Technological Processes / K. B. Fam, P. M. Murashev, V. N. Bogatkov // *Pattern Recognition and Image Analysis*. – 2024. – Vol. 34. – No. 3. – P. 632–638.

8. Dvornikov S. V., Vasilieva D. V. Improving the Reliability of Anomaly Detection in Images by Forming Feature Vectors in Wavelet Bases // Informatics and Automation. – 2024. – Vol. 23. – No. 6. – P. 1698–1729. – DOI: 10.15622/ia.23.6.5.
9. Soifer V. A., Fursov V. A., Kharitonov S. I. Kalman Filtering for a Class of Images of Dynamic Objects // Informatics and Automation. – 2024. – Vol. 23. – No. 4. – P. 953–968. – DOI: 10.15622/ia.23.4.1.
10. Verkhotorov A. L., Stepanov A. S., Illarionova L. V. Using Radar Data to Monitor Crop Conditions in the South of the Russian Far East // Informatics and Automation. – 2024. – Vol. 23. – No. 4. – P. 1221–1245. – DOI: 10.15622/ia.23.4.11.
11. Dai M. A Transformer-Based Feature Segmentation and Region Alignment Method for UAV-View Geo-Localization / M. Dai, J. Hu, J. Zhuang, E. Zheng // IEEE Transactions on Circuits and Systems for Video Technology. – 2022. – Vol. 32. – No. 7. – P. 4376–4389.
12. Zuev V. M. Comparison of Object Detection Using Artificial Intelligence vs. Classical Methods // Problems of Artificial Intelligence. – 2024. – No. 3(34). – P. 4–10. – ISSN: 2413-7383. – DOI: 10.24412/2413-7383-2024-3-30-35.
13. Pikalev Y. S. On Neural Architectures for Feature Extraction in Object Recognition Tasks on Devices with Limited Computational Power / Y. S. Pikalev, T. V. Ermolienko // Problems of Artificial Intelligence. – 2023. – No. 3(30). – P. 44–54. – ISSN: 2413-7383. – DOI: 10.34757/2413-7383.2023.30.3.004.
14. Pavlenko B. V. Intellectually Algorithmic Method for Sight Calibration / B. V. Pavlenko, V. I. Bondarenko // Problems of Artificial Intelligence. – 2024. – No. 3(34). – P. 55–63. – ISSN: 2413-7383. – DOI: 10.24412/2413-7383-2024-3-55-63.
15. Krishnan S. R. Improving Anomaly Detection in Video Using Enhanced UNET Technology and Cascade Sliding Window Technique / S. R. Krishnan, P. Amudha // Informatics and Automation. – 2024. – No. 6(23). – P. 1899–1930. – ISSN: 2713-3192. – DOI: 10.15622/ia.23.6.12.
16. Ali B. Navigation Algorithms for Unmanned Aerial Vehicles Using Vision Systems / B. Ali, R. N. Sadekov, V. V. Tsodokova // Gyroscopy and Navigation. – 2022. – Vol. 30. – No. 4(119). – P. 87–105.
17. Khakimov R. S. On the Development of a Data Annotation System for Computer Vision Tasks / R. S. Hakimov, O. L. Nizhnikova, M. V. Blizno // Problems of Artificial Intelligence. – 2024. – No. 3(34). – P. 70–79. – ISSN: 2413-7383. – DOI: 10.24412/2413-7383-2024-3-70-79.
18. Ronzhin A. L. Intellectualization and Robotization of Domestic Scientific Equipment for Interdisciplinary Research / A. L. Ronzhin. – 2023. – Vol. 1. – No. 28. – P. 4–10.
19. Durgama A., Paheding S., Dhiman V., Devabhaktuni V. Cross-view Geo-Localization: A Survey // Ad Hoc Networks. – 2021. – Vol. 29. – No. 3. – P. 1519–1541. – DOI: 10.1016/j.adhoc.2021.101843.
20. Khakimov R. S. Overview of Advanced Augmentation Techniques for Image Datasets / R. S. Khakimov, B. V. Pavlenko, Ya. S. Pikalyov // Donetsk Readings 2024: Education, Science, Innovations, Culture and Challenges of Modernity: Proceedings of the IX International Scientific Conference (Donetsk, October 15–17, 2024). – Vol. 2: Physical, Chemical, Technical, and Computer Sciences. Part 2 / Edited by Prof. S. V. Bepalova. – Donetsk: DonGU Publishing, 2024. – 296 p. – P. 272–275. – ISSN: 2664-7362 (Print); ISSN: 2664-7370 (Online).

RESUME

T.V. Yermolenko, R. S. Khakimov

On the question of application deep learning for the problem of cross-view geolocation

The article is devoted to the problems and key aspects of the application of deep learning in the task of cross-geolocation. The paper provides a mathematical description of the problem of cross geolocation and considers three approaches: 1) correlation-extreme; 2) the use of key points; 3) the use of neural networks as the only element of the comparison algorithm with the standard.

The paper considers the main problems and key features of CVGL. Based on this, solutions to some problems have been proposed. 1) The results of the Transformer-based model, trained twice, have a large spread. To solve this problem, when processing satellite images, it is necessary to generate k extended satellite images and select k images from other angles that belong to the same category. 2) The scale of satellite images is fixed, while the altitude of the UAV flight changes dynamically. In accordance with this, it is necessary to

have images from different heights in the dataset: large, medium and low to allow the model to adapt to scaling. 3) CVGL approaches focus on the application of spatial attention mechanisms, describing the main goal or contextual information about the goal, but do not take into account additional information received from the UAV itself. To solve this problem, it is necessary to transfer UAV flight data to the model (for example, in the form of text templates).

As a result, the following conclusions can be drawn.

1) When developing effective systems for the task of cross-geolocation, it is necessary to use a hybrid approach based on the key point method and the use of neural networks as the only element of the comparison algorithm with the reference.

2) Using Transformer architecture allows you to more clearly identify characteristic objects such as buildings and roads, and ignore background information such as trees.

3) When training models on existing datasets, their focus on the urban environment should be taken into account. Therefore, an urgent direction is to collect and create data sets that would contain different types of terrain in order to increase the diversity, usefulness and practicality of such data.

4) When creating datasets, it is worth considering the dependence of models on the size of the input image and the altitude of the UAV.

РЕЗЮМЕ

Т.В. Ермоленко, Р. С. Хакимов

К вопросу о применении глубокого обучения к задаче перекрёстной геолокализации

Статья посвящена проблемам и ключевым аспектам применения глубокого обучения в задаче перекрёстной геолокализации. В работе дано математическое описание задачи перекрёстной геолокализации и рассмотрено три подхода: 1) корреляционно-экстремальный; 2) использование ключевых точек; 3) использование нейронных сетей как единственного элемента алгоритма сопоставления с эталоном.

В работе рассмотрены основные проблемы и ключевые особенности CVGL. На основе этого были предложены решения некоторых проблем. 1) Результаты модели на основе Трансформер, обученной дважды, имеют большой разброс. Для решения этой проблемы необходимо при обработке изображения со спутника генерировать к расширенным спутниковым изображениям и выбирать к изображениям с других ракурсов, которые относятся к той же категории. 2) Масштаб изображений со спутника фиксирован, в то время как высота полёта БПЛА динамически изменяется. В соответствии этим необходимо наличие в наборе данных снимков с разных высоты: большой, средней и низкой для обеспечения возможности модели для адаптации к масштабированию. 3) Подходы в CVGL сосредоточены на применении механизмов пространственного внимания, описывая основную цель или контекстуальную информацию о цели, но не учитывают дополнительную информацию, получаемую от самого БПЛА. Для разрешения этой проблемы необходимо передавать модели данные о полете БПЛА (например, в виде текстовых шаблонов).

В результате можно сделать следующие выводы.

1) При разработке эффективных систем для задачи перекрёстной геолокализации необходимо использовать гибридный подход на основе метода ключевых точек и использования нейронных сетей как единственного элемента алгоритма сопоставления с эталоном.

2) Использование архитектуры Трансформер позволяет более четко идентифицировать характерные объекты, такие как здания и дороги, и игнорировать фоновую информацию, такую как деревья.

3) При обучении моделей на существующих наборах данных следует учитывать их сосредоточенность на городской среде. Поэтому актуальным направлением является сбор и создание наборов данных, которые содержали бы различные виды местности, чтобы увеличить разнообразие, полезность и практичность подобных данных.

4) При создании наборов данных стоит учитывать зависимость моделей от размера входного изображения и высоты полёта БПЛА.

Ермоленко Татьяна Владимировна – кандидат техн. наук, доцент, доцент кафедры компьютерных технологий физико-технического факультета ФГБОУ ВО "ДонГУ", заместитель заведующего лабораторией интеллектуальных систем и анализа данных ФГБНУ "ИПИИ". Область научных интересов: математическая логика, нечеткая логика, обработка естественного языка, глубокое обучение, распознавание образов, цифровая обработка сигналов. Число научных публикаций – более 80. t.ermolenko@donnu.ru; ул. Университетская., д. 24, г. Донецк, 283001, Донецкая Народная Республика, РФ.

Хакимов Ренат Саитович – младший научный сотрудник, лаборатория интеллектуальных систем и анализа данных ФГБНУ "ИПИИ". Область научных интересов: компьютерное зрение, машинное обучение, нейронные сети. Число научных публикаций – более 5. khakimov.ru@mail.ru; ул. Артема, д. 118 б, г. Донецк, 283048, Донецкая Народная Республика, РФ.

Статья поступила в редакцию 03.06.2024.