

Санкт-Петербургский государственный университет

*Москаленко Иван Николаевич*

Выпускная квалификационная работа

# Масштабируемая визуальная локализация и картирование по аэрофотоснимкам

Уровень образования: бакалавриат

Направление *02.03.03 “Математическое обеспечение и администрирование  
информационных систем”*

Основная образовательная программа *СВ.5162.2020 “Технологии программирования”*

Научный руководитель:  
старший преподаватель, к.т.н. Ю. В. Литвинов

Консультант:  
инженер-исследователь, Mobile Robotics Lab, Skoltech А. В. Корнилова

Рецензент:  
научный сотрудник, Технический университет Мюнхена, к.ф.-м.н. А. В. Артёмов

Санкт-Петербург  
2024

Saint Petersburg State University

*Ivan Moskalenko*

Bachelor's Thesis

# Large-scale localization and mapping from aerial images

Education level: bachelor

Speciality *02.03.03 "Software and Administration of Information Systems"*

Programme *CB.5162.2020 "Programming Technologies"*

Scientific supervisor:  
C.Sc., senior lecturer Y.V. Litvinov

Consultant:  
research engineer at Mobile Robotics Lab, Skoltech A.V. Kornilova

Reviewer:  
C.Sc., research scientist at Technical University of Munich A.V. Artemov

Saint Petersburg  
2024

# Оглавление

|                                                 |           |
|-------------------------------------------------|-----------|
| <b>1. Введение</b>                              | <b>4</b>  |
| <b>2. Постановка цели и задач</b>               | <b>6</b>  |
| <b>3. Обзор</b>                                 | <b>7</b>  |
| 3.1. Методы визуальной геолокализации . . . . . | 7         |
| 3.2. Методы глобальной локализации . . . . .    | 8         |
| 3.3. Методы локального выравнивания . . . . .   | 11        |
| 3.4. Датасеты . . . . .                         | 11        |
| 3.5. Метрики . . . . .                          | 13        |
| <b>4. Реализация инструмента сравнения</b>      | <b>15</b> |
| 4.1. GeoReferencer . . . . .                    | 15        |
| 4.2. FeatureMatcher . . . . .                   | 16        |
| 4.3. VPRSystem . . . . .                        | 16        |
| 4.4. IndexSearcher . . . . .                    | 16        |
| 4.5. MapDownloader . . . . .                    | 17        |
| 4.6. Map . . . . .                              | 17        |
| 4.7. RetrievalSystem . . . . .                  | 17        |
| 4.8. HomographyEstimator . . . . .              | 17        |
| 4.9. LocalizationPipeline . . . . .             | 18        |
| <b>5. Экспериментальное сравнение</b>           | <b>19</b> |
| 5.1. Цель и вопросы эксперимента . . . . .      | 19        |
| 5.2. Тестовый стенд . . . . .                   | 19        |
| 5.3. Результаты . . . . .                       | 20        |
| 5.4. Выводы . . . . .                           | 29        |
| <b>6. Заключение</b>                            | <b>31</b> |
| <b>Список литературы</b>                        | <b>32</b> |

# 1. Введение

Локализация аэросъемки является важной задачей в современном мире [29]. Одним из возможных сценариев её использования является построение карты для локализации других летательных аппаратов [1, 43]. Также локализация аэросъёмки может применяться для построения точной 3D-модели местности и ее дальнейшего использования для анализа местности. Такие модели используются в различных областях, например, в геологии [60], строительстве [42] и многих других.

В большинстве случаев для локализации летательных аппаратов применяются такие системы глобального позиционирования как GPS [25] или ГЛОНАСС [22]. Однако данный подход требует стабильности спутникового сигнала и отсутствия помех. Это не всегда возможно, особенно в случае наличия отражений сигнала из-за водоёмов и/или сложного рельефа местности. Также стоит отметить, что покрытие сигнала глобальных спутниковых систем на планете неравномерно. Все эти причины вынуждают использовать другие подходы для локализации.

Одним из таких подходов является визуальная геолокализация — задача нахождения местоположения сделанного снимка. Поиск может осуществляться по-разному, однако чаще всего для этого используется заранее созданная база данных изображений и местоположений, им соответствующих [65]. Данный подход обычно делится на два этапа: глобальную локализацию и локальное выравнивание. Первый заключается в поиске ближайшего кадра из базы данных, а второй вычисляет точное положение системы, используя выбранный кадр. Глобальная локализация (или VPR — visual place recognition) также может выполняться по-разному. Некоторые методы предполагают использование дескрипторов — компактного представления изображений в базе данных, другие же используют ключевые точки изображений для сопоставления. Второй подход является более вычислительно сложным, однако способен делать более точные предсказания [7].

Использование визуальной геолокации в случае аэросъёмки может быть осложнено рядом факторов, такими как изменяющиеся по-

годные условия, смена времен года и времени суток. Задача локализации аэрофотоснимков требует карт достаточно большого размера, что также накладывает ограничения на этап глобальной локализации. При этом снимки из общедоступных наборов данных, таких как Google Maps API [26] и Sentinel [55], имеют разное разрешение. Для Sentinel максимальное разрешение составляет 10 м/пиксель, а в случае Google Maps оно варьируется от 0.15 м/пиксель до 5 м/пиксель.

На данный момент обзоры-сравнения VPR-систем сфокусированы на данных, снятых внутри и вне помещений [41, 23, 67, 18], и не затрагивают спутниковые изображения и аэрофотоснимки. Поэтому в рамках данной работы планируется провести обзор различных подходов к визуальной геолокализации, предложить метрики для оценки качества различных методов и определить, какие методы в связке с каким представлением карт дают наиболее точные результаты в случае аэрофотоснимков.

## 2. Постановка цели и задач

Целью работы является разработка инструмента для сравнения различных алгоритмов визуальной геолокализации и сравнение существующих подходов с его помощью.

Для достижения данной цели были поставлены следующие задачи.

1. Выполнить обзор существующих решений для визуальной геолокализации, а также глобальной локализации и локального выравнивания по отдельности.
2. Предложить и реализовать метрики для оценки результата работы различных методов локализации.
3. Спроектировать и разработать инструмент, позволяющий сравнивать различные подходы к визуальной локализации.
4. Провести экспериментальное исследование различных методов локализации, используя разработанный инструмент.

## 3. Обзор

В данном разделе будет проведён обзор существующих методов визуальной геолокализации, а также глобальной локализации и локального выравнивания по отдельности. Кроме того, будут описаны существующие датасеты аэрофотоснимков и метрики для оценки качества работы систем визуальной геолокализации.

### 3.1. Методы визуальной геолокализации

Существующие подходы для визуальной геолокализации аэросъёмки можно разделить на несколько больших групп: сравнение изображения-запроса с базой данных изображений [12, 49, 38, 28, 50], сравнение изображения-запроса с картой [58, 43, 34, 9, 27], а также использование сегментации [73, 21, 13, 24, 52].

Алгоритмы работы первой группы методов в свою очередь делятся на два этапа: глобальную локализацию и локальное выравнивание. Задача глобальной локализации заключается в поиске ближайшего кадра в базе данных изображений, а локальное выравнивание вычисляет точное положение системы, используя выбранный кадр. Общей проблемой предлагаемых сообществом решений является отсутствие сравнения конкретных компонентов систем визуальной геолокализации.

Вторая группа методов использует не базу данных изображений, а только лишь одно изображение-карту. В этом случае глобальная локализация не нужна, и процесс локализации состоит только из локального выравнивания. Данный подход применим только к картам малого размера и не является масштабируемым.

Третья группа методов выполняет сегментацию зданий/дорог на изображениях-запросах, а затем сравнивает полученную маску с сегментированной картой. Данный подход требователен к наличию ориентиров на изображении-запросе, и, соответственно, не подходит для малотекстурных сред. Также стоит отметить, что для работы подобных методов требуется довольно точная маска, которую сложно получить, используя текущие state-of-the-art алгоритмы сегментации.

В рамках данной работы подробно рассматривается первый подход к визуальной геолокализации, а также его составляющие — visual place recognition и локальное выравнивание.

## **3.2. Методы глобальной локализации**

В данном разделе будут описаны различные методы глобальной локализации.

### **3.2.1. Методы, использующие дескрипторы изображений**

Методы [45, 10, 20, 4, 5] из данной группы используют компактные представления изображений — дескрипторы. Хотя для каждого отдельно взятого алгоритма способ построения дескрипторов различен, идея одна. Она заключается в извлечении локальных признаков изображений и агрегации в вектор фиксированной длины. Преимуществом данного подхода является относительно высокая скорость, особенно, если размер вектора невелик. Отсюда вытекает и недостаток подобных методов — потеря важной информации при создании дескриптора. Поэтому текущие решения стремятся использовать такие компактные представления, которые бы содержали в себе как можно больше важной информации. Текущими современными решениями являются AnyLoc [5], CosPlace [10], EigenPlaces [20], MixVPR [4], NetVLAD [45], SALAD [32]. Новизна данных visual place recognition систем объясняет отсутствие независимых статей, которые бы сравнивали эти методы между собой.

AnyLoc — метод, который работает в широком диапазоне структурированных и неструктурированных сред без дополнительного дообучения. В статье демонстрируется, что признаки общего назначения, получаемые с помощью так называемых foundation models, обученных на больших объёмах данных, являются подходящей основой для создания такого универсального решения для визуальной геолокализации.

CosPlace — высокомасштабируемый метод обучения, который рассматривает обучение как задачу классификации. Это помогает избежать трудозатратного конструирования данных, необходимого для кон-

трастного обучения, которое требует для описания изображения не только позитивные примеры, но и негативные. В качестве базовой сети используется ResNet [17].

EigenPlaces — ещё один метод обучения, который во многом похож на CosPlace, но при этом он фокусируется на тренировке нейронной сети на изображениях, снятых с разных позиций. Таким образом, обучаемый дескриптор становится более устойчивым к изменению точки обзора.

MixVPR — техника агрегирования признаков изображений, которая принимает на вход карты признаков из промежуточных слоёв предварительно обученной базовой сети ResNet. Она объединяет отношения между элементами в каждой карте признаков в единое компактное представление, и, таким образом, устраняет необходимость в локальном агрегировании, которое применяется в NetVLAD.

В статье, посвященной NetVLAD, авторы предлагают архитектуру свёрточной нейронной сети, которая обучается для задачи визуальной геолокализации. Основным компонентом этой архитектуры является слой VLAD [6], который вдохновлён представлением изображения в виде вектора локально агрегированных дескрипторов, широко используемым в поиске изображений.

SALAD переформулирует мягкое назначение (soft assignment) локальных признаков кластерам в NetVLAD как оптимальную транспортную задачу. В SALAD учитываются отношения между признаками и кластерами, а также вводится «мусорный» кластер, предназначенный для выборочного отбрасывания признаков, которые считаются неинформативными, что повышает общее качество дескриптора. Кроме того, в качестве базовой сети используется DINOv2 [19], что позволяет расширить возможности описания локальных признаков и значительно сократить время обучения.

### **3.2.2. Методы, использующие ключевые точки**

Данная группа методов [7, 47, 14] для своей работы использует ключевые точки — точки, которые с большей вероятностью будут находиться на другом изображении того же объекта. Существуют раз-

личные подходы для обнаружения и описания ключевых точек. Самыми популярными [30] классическими методами являются ORB [46], SIFT [39] и SURF [56], в то время как текущее нейронное state-of-the-art [36] решение — SuperPoint [16]. Сопоставление ключевых точек также может выполняться по-разному. На данный момент самым популярным [7, 20, 48] является нейронный подход SuperGlue [57]. Также относительно недавно стал доступен сопоставитель точек LightGlue [35], который декларирует повышенную точность и производительность относительно SuperGlue, поэтому он также рассматривается в работе.

В SuperGlue соответствия между ключевыми точками рассчитываются решением дифференцируемой транспортной задачи, веса для которой предсказываются с помощью графовой нейронной сети. Гибкий механизм агрегирования контекстной информации позволяет SuperGlue получать базовое понимание о пространственных характеристиках сцены.

LightGlue — ещё одно решение для сопоставления ключевых точек. Его ключевой особенностью является способность адаптироваться к сложности задачи: предсказания происходят намного быстрее на парах изображений, которые интуитивно легко сопоставить, например, из-за большого визуального перекрытия.

Поиск наиболее подходящего изображения в базе данных выполняется с помощью сопоставления наборов ключевых точек. Выбранным изображением обычно является то, для которого нашлось наибольшее количество совпадений с ключевыми точками изображения-запроса. Преимуществом данного подхода является более высокая точность по сравнению с методами, использующими дескрипторы изображений. Однако и вычислительная сложность методов, использующих ключевые точки, выше.

### **3.2.3. Комбинированный подход**

Эффективность методов, работающих на ключевых точках, и производительность дескрипторных методов привели к увеличению популярности комбинированного подхода [7]. Он заключается в том, что с

помощью дескрипторных методов отбирается  $N$  лучших снимков, а выбор одного наилучшего производится с помощью более точных алгоритмов, работающими с ключевыми точками. Таким образом достигается баланс между производительностью и точностью.

### **3.3. Методы локального выравнивания**

Процесс локального выравнивания можно разделить на две части. В первой части строится отображение из изображения-запроса в соответствующий ему снимок из базы данных. Для построения такого отображения могут использоваться как разреженные ключевые точки, например как в реализации из библиотеки OpenCV [11], так и плотные карты характеристик [68].

Вторая часть заключается в определении геокоординат системы с помощью изображения-запроса. Используя построенное отображение, можно найти пиксельные координаты запроса в системе координат изображения из базы данных. Так как обычно для фрагмента карты из базы данных известны лишь геокоординаты угловых точек, также необходимо уметь вычислять геокоординаты для произвольного пикселя. Такие вычисления зависят от модели камеры, используемой для создания снимков в базе данных.

Подробно методы локального выравнивания рассмотрены в работе студентки второго курса СПбГУ математико-механического факультета по направлению «Технологии программирования» Николаевой Анны.

### **3.4. Датасеты**

На данный момент количество датасетов с аэрофотоснимками, находящихся в открытом доступе, ограничено. Данная работа фокусируется на последовательных данных, в которых установлен порядок кадров и каждый следующий кадр имеет визуальное пересечение с предыдущим. Это позволяет симулировать непрерывную съёмку с летательных аппаратов, но ещё сильнее сужает перечень доступных да-

| Датасет         | Количество кадров | Длина траектории | Разрешение снимков |
|-----------------|-------------------|------------------|--------------------|
| VPAir query     | 2706              | 100 км           | 800x600            |
| VPAir reference | 2706              | 100 км           | 800x600            |
| ALTO            | 460               | 37 км            | 500x500            |

Таблица 1: Характеристики используемых тестовых датасетов

тасетов. Для оценки работы алгоритмов визуальной геолокализации в работах [5, 12, 59, 64, 38, 44, 71, 74, 72, 15] используются наборы данных University-1652 [69], SUES-200 [51], DenseUAV [63], CVUSA [66], VPAir [53], ALTO [2], Nardo-Air [5]. Однако датасеты University-1652, SUES-200 и CVUSA не являются последовательными, а набор данных DenseUAV хоть и подходит под необходимые критерии, но находится в закрытом доступе. Датасет Nardo-Air в свою очередь покрывает траекторию длиной всего в 700 метров. Поэтому в данной работе используются наборы данных ALTO и VPAir.

Набор данных ALTO представляет собой полет над широким набором различных местностей: леса, города, сельская местность, реки и озера. Камера направлена перпендикулярно плоскости Земли и обеспечивает беспрепятственный вид на территорию, над которой пролетает аппарат. Оба полета сняты в один и тот же день летом. Также датасет включает в себя высокоточные местоположения сделанных снимков.

Набор данных VPAir содержит две последовательности: query и reference. Query последовательность была снята с борта воздушного судна, в то время как reference последовательность содержит ортофотографии, полученные на основе аэрофотоснимков. Для каждого снимка также записаны эталонные положения системы. Набор данных охватывает траекторию длиной более ста километров над различными типами сложных ландшафтов, например, городскими, сельскохозяйственными и лесными угодьями между городом Бонн и горным массивом Эйфель.

Подробные характеристики датасетов можно увидеть в табл. 1.

## 3.5. Метрики

В задаче visual place recognition стандартной метрикой является Recall@k [18, 45, 62, 3, 47, 33, 37, 54, 8, 40], которая измеряет процент изображений-запросов, для которых хотя бы одно из k лучших изображений, выбранных методом, находится не дальше заданного порогового значения. Однако специфика используемых данных, а также необходимость оценивать качество работы глобальной локализации и всего конвейера визуальной геолокализации по отдельности требует небольшой корректировки данной метрики. Для этого было предложено две модификации оригинальной Recall@k метрики.

### 3.5.1. Recall для геопривязки

Данная метрика вычисляет местоположения тестовых снимков и сверяет их с эталонными координатами. Если разница меньше заданного порогового значения, то данное изображение считается правильно локализованным. Результат метрики — отношение правильно локализованных пар к их общему количеству. Таким образом, метрика позволяет оценить точность геопривязки, то есть конвейера визуальной геолокализации целиком. По сути эта метрика является подобием Recall@1, с той лишь разницей, что в качестве результирующего местоположения берутся не геокоординаты выбранного тайла карты, а результаты локального выравнивания относительно него.

### 3.5.2. Recall для систем глобальной локализации

Для оценки только систем глобальной локализации без участия локального выравнивания затруднительно определить результирующие геокоординаты, так как изображения в базе данных представляют собой плоские прямоугольные куски спутниковой карты, для которых по умолчанию доступны только угловые координаты. Поэтому данная метрика вычисляет лучшие N предсказаний из базы данных изображений с помощью системы глобальной локализации для каждого тестового снимка, а затем, если центр изображения-запроса попадает хотя бы в

один из выбранных снимков из базы данных, то данный результат засчитывается как правильный. Результат метрики — отношение правильных результатов к общему количеству тестовых изображений. Таким образом, метрика позволяет оценить качество работы систем глобальной локализации, не требует указания конкретного порогового значения и может работать с разными уровнями масштаба спутниковых карт.

## 4. Реализация инструмента сравнения

Поскольку все нейросетевые методы, используемые в данной работе, построены на Python [61] фреймворках, этот язык программирования был выбран и для реализации всего инструмента. Удобство прототипирования на Python является дополнительным плюсом. Как видно из UML-диаграммы на рис. 1, основны-

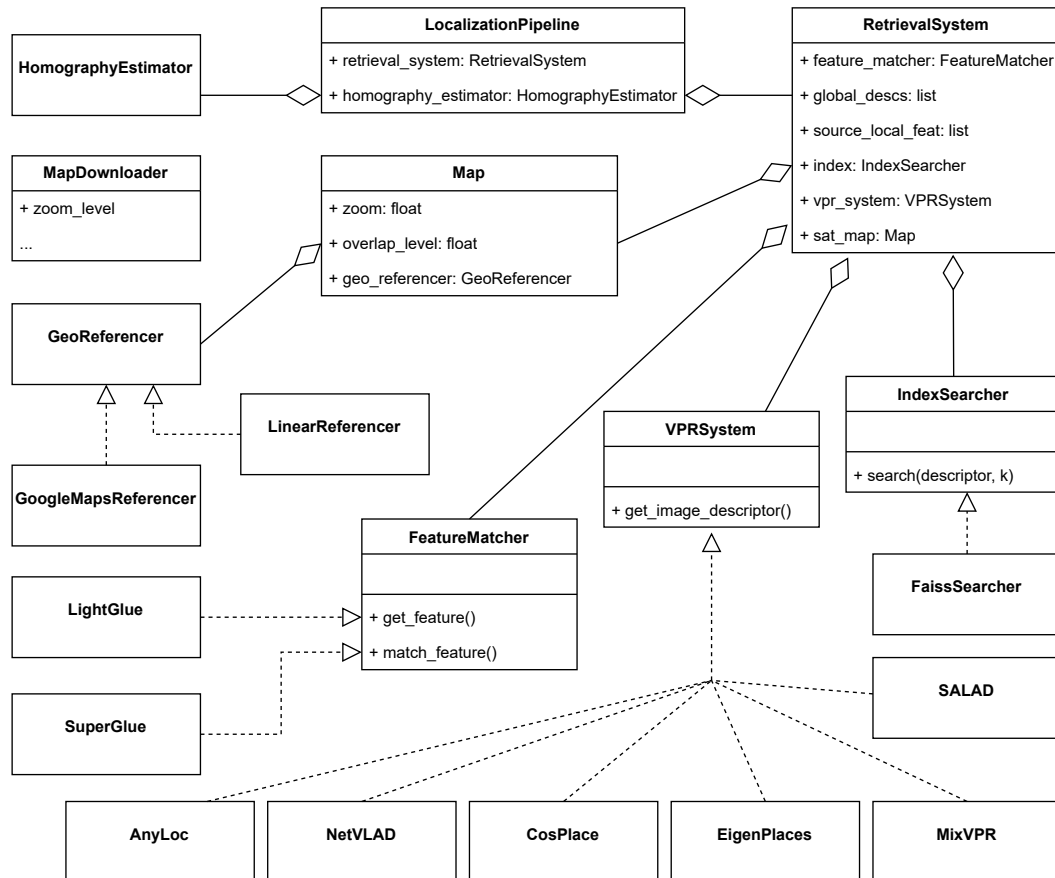


Рис. 1: UML-диаграмма реализованного инструмента

ми классами инструмента являются: GeoReferencer, FeatureMatcher, VPRSystem, IndexSearcher, MapDownloader, Map, RetrievalSystem, HomographyEstimator, LocalizationPipeline. Далее описаны детали того, за что отвечают эти классы.

### 4.1. GeoReferencer

Как уже было упомянуто ранее, обычно для изображений в базе данных известны геокоординаты только для четырёх угловых точек.

Поэтому в инструменте реализована функциональность для определения геокоординат произвольного пикселя на необходимом кусочке карты. На данный момент поддерживается простая линейная модель, которая предполагает, что система координат карты — декартова. Она работает достаточно хорошо, если размер одного тайла карты невелик. Также реализована модель Web Mercator projection или WGS 84 Web Mercator [31], которая способна более точно описывать данные из Google Maps.

## 4.2. FeatureMatcher

Данный класс предоставляет возможность вычислять дескрипторы ключевых точек, а также сопоставлять их. В инструменте поддерживаются ключевые точки SuperPoint, а также ассоциаторы LightGlue и SuperGlue.

## 4.3. VPRSystem

Данный класс предоставляет функциональность для вычисления дескрипторов изображений с помощью различных систем визуальной геолокализации. Сейчас в инструменте поддерживаются AnyLoc, CosPlace, EigenPlaces, MixVPR, NetVLAD и SALAD.

## 4.4. IndexSearcher

Инструмент предоставляет возможность создания базы данных из дескрипторов изображений и поиска в ней. Для заданного дескриптора изображения-запроса возвращается  $N$  ближайших по расстоянию дескрипторов, где  $N$  — конфигурируемый пользователем параметр. Данная функциональность реализована с помощью библиотеки Faiss [70].

## 4.5. MapDownloader

Данный класс позволяет строить спутниковую карту из данных Google Maps [26] с помощью соответствующего API. Сама карта при этом представляет собой сетку из небольших кусочков, которые также называются тайлами. Инструмент позволяет устанавливать zoom level — масштаб скачиваемых изображений.

## 4.6. Map

Класс Map предназначен для работы с ранее скачанной спутниковой картой. Он ответственен не только за чтение файлов-тайлов с диска, но и за создание новых тайлов «на лету». Благодаря этому, пользователь может указать произвольный уровень перекрытия между соседними кадрами, а также масштаб. Стоит отметить, что высокий уровень приближения негативно сказывается на качестве тайлов, так как инструмент лишь обрезает исходные изображения. Поэтому целесообразно выбирать нужный масштаб сразу при скачивании и лишь немного корректировать его в дальнейшем.

## 4.7. RetrievalSystem

RetrievalSystem — класс, который позволяет производить поиск ближайших изображений в базе данных, комбинируя различные подходы к глобальной локализации. Первичный поиск может производиться с помощью различных систем визуальной локализации, которые поддерживаются инструментом. Далее результаты первой стадии могут быть отранжированы с помощью ассоциаторов ключевых точек LightGlue или SuperGlue.

## 4.8. HomographyEstimator

Данный класс позволяет рассчитать гомографию между изображением-запросом и выбранным глобальной локализацией

снимком из базы данных, используя соответствия между ключевыми точками. Вычисленная гомография применяется для нахождения пиксельных координат запроса в системе координат кусочка карты.

## 4.9. LocalizationPipeline

LocalizationPipeline — класс выбирает наиболее близкое изображение из базы данных с помощью класса RetrievalSystem. Затем, используя класс HomographyEstimator, строит гомографию между выбранным изображением и снимком-запросом. Наконец, инструмент вычисляет финальный результат в виде геокоординат системы с помощью класса GeoReferencer.

## 5. Экспериментальное сравнение

В данном разделе будут приведены цель и исследовательские вопросы проводимого экспериментального сравнения, рассказано об условиях, при которых оно проводилось. Также будут предоставлены результаты экспериментов и выводы из них.

### 5.1. Цель и вопросы эксперимента

Целью проводимого эксперимента является сравнение различных подходов к визуальной локализации. Для достижения данной цели были поставлены следующие исследовательские вопросы.

- RQ1: Какое значение  $N$ , отвечающее за количество предсказаний в метрике Recall для глобальной локализации, оптимальнее всего отображает качество работы алгоритмов?
- RQ2: Какой уровень приближения спутниковой карты позволяет добиться лучших результатов?
- RQ3: Какой уровень перекрытия между соседними тайлами карты является наиболее подходящим?
- RQ4: Какой из сопоставителей ключевых точек работает эффективнее других? Необходимо ли вообще применять методы, использующие ключевые точки, или достаточно дескрипторных?
- RQ5: Какие методы глобальной локализации являются наилучшими для визуальной геолокализации?
- RQ6: Какова производительность выбранных методов визуальной геолокализации?

### 5.2. Тестовый стенд

В качестве тестовых использовались наборы данных ALTO и VPAir. Для VPAir эксперименты проводились на обеих имеющихся последова-

| Датасет | Количество тайлов | Площадь              | Разрешение снимков |
|---------|-------------------|----------------------|--------------------|
| VPAir   | 4788              | 1079 км <sup>2</sup> | 1280x1230          |
| ALTO    | 2706              | 718 км <sup>2</sup>  | 1280x1230          |

Таблица 2: Характеристики используемых карт

тельность: query и reference.

Для обоих датасетов карта была выгружена с помощью соответствующей функциональности, встроенной в инструмент. Среди всех уровней приближения, доступных через Google Maps API, был выбран 17, так как этот уровень приближения соответствует масштабу, наиболее близкому к тестовым данным. Подробные характеристики выгруженных карт можно увидеть в табл. 2.

### 5.3. Результаты

В данном разделе будут представлены результаты экспериментов для каждого из исследовательских вопросов.

#### 5.3.1. Выбор параметра N метрики Recall для глобальной локализации

Главным параметром метрики Recall для глобальной локализации является значение N, которое отвечает за количество кадров, отбираемых тем или иным алгоритмом. Для того чтобы все дальнейшие сравнения были честными, на первом этапе было решено найти оптимальное значение N. Для этого значение метрики Recall было посчитано для всех N от 1 до 300. На графиках 2, 3 и 4 можно увидеть зависимость значения метрики от величины N для датасетов VPAir query, VPAir reference и ALTO соответственно. Для всех замеров не применялось изменение масштаба для тайлов карты, а уровень перекрытия между соседними кадрами был равен 25%.

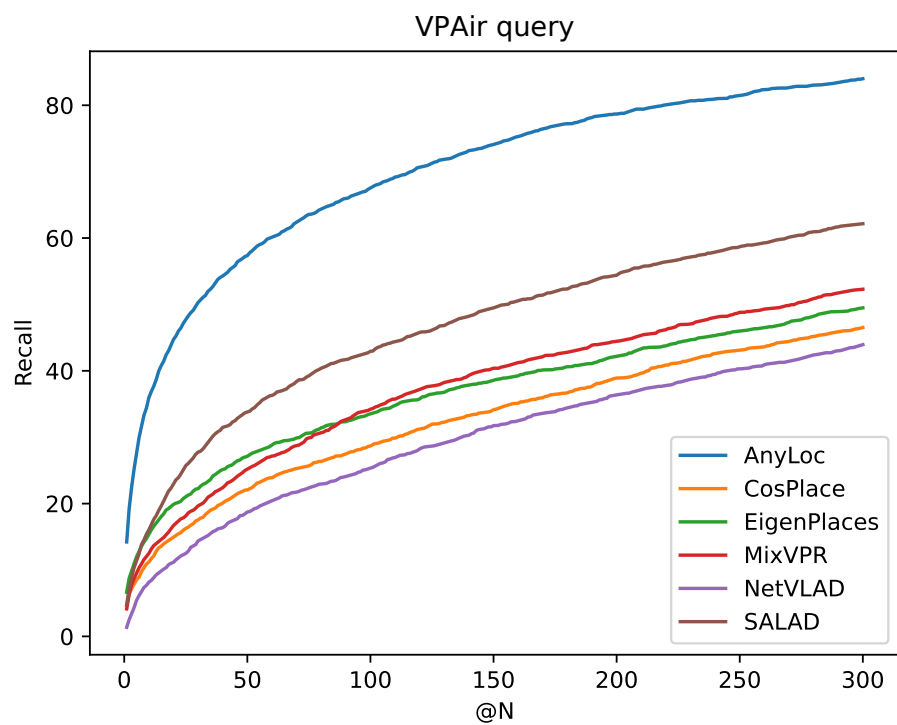


Рис. 2: Зависимость Recall для глобальной локализации от значения N, VPAir query

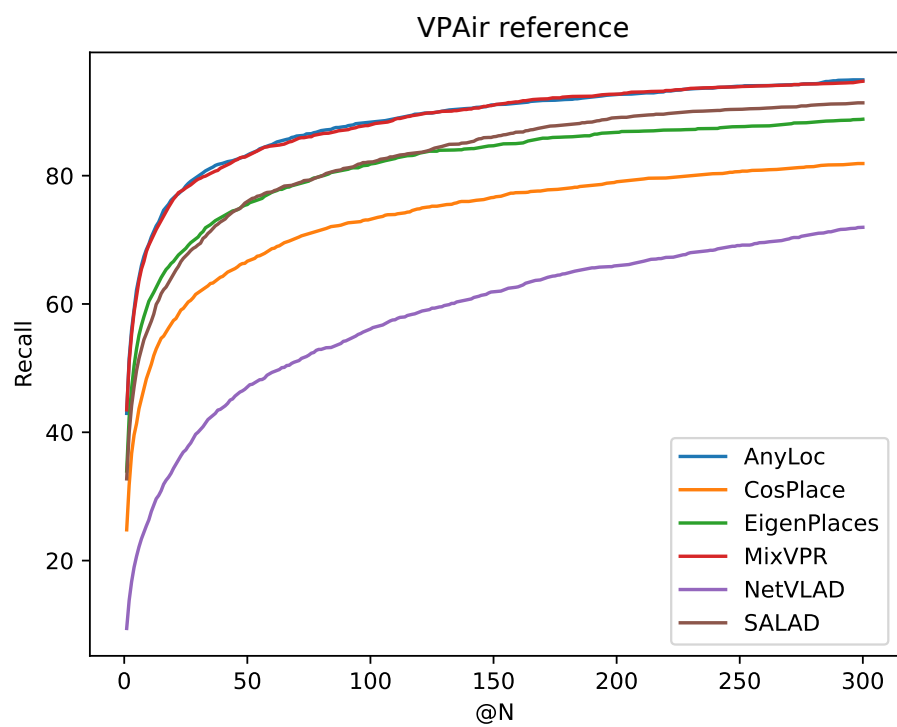


Рис. 3: Зависимость Recall для глобальной локализации от значения N, VPAir reference

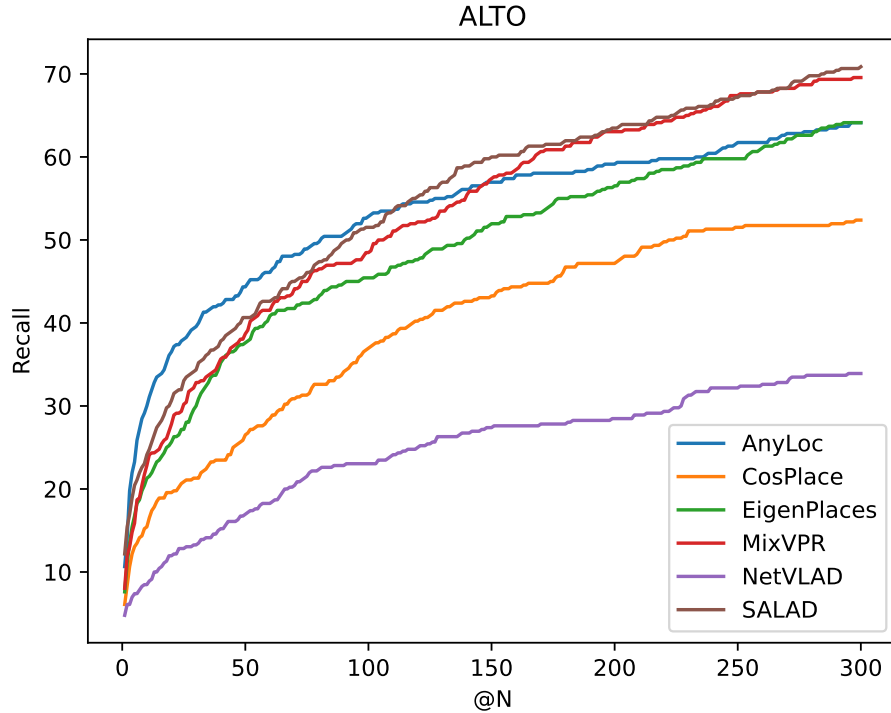


Рис. 4: Зависимость Recall для глобальной локализации от значения N, ALTO

Исходя из полученных результатов, было принято решение использовать значения 1, 10, и 50 для всех последующих сравнений, так как между ними происходят значительные скачки в качестве, а при величине  $N > 50$  рост значения Recall замедляется.

### 5.3.2. Выбор оптимального уровня приближения для спутниковых карт

Для максимально эффективной работы алгоритмов визуальной гео-локализации важно подобрать подходящий масштаб спутниковой карты. Изменение масштаба карты способно уравнивать площадь покрытия спутникового кадра и кадра-запроса. Однако стоит также отметить, что увеличение уровня приближения увеличивает и количество кадров в базе данных, что может «запутать» системы глобальной геолокализации. Поэтому на втором этапе экспериментов было сделано сравнение работы различных алгоритмов при разных уровнях масштаба. Для датасета ALTO использовались значения масштаба, равные 50%, 100%,

150%, 200% и 250%. Для двух последовательностей VPAir использовались те же значения кроме 250%. Например, уровень масштаба в 200% означает, что стороны исходного тайла карты уменьшены в два раза, а количество снимков в базе данных увеличено в 4 раза относительно количества исходных тайлов. Везде использовался уровень перекрытия между соседними кадрами, равный 25%. Результаты сравнения можно увидеть в таблицах 3, 4 и 5 для датасетов VPAir query, VPAir reference и ALTO соответственно.

| Метод локализации | Zoom 50% |      |             | Zoom 100%  |             |             | Zoom 150%   |             |             | Zoom 200% |      |      | Лучший масштаб |
|-------------------|----------|------|-------------|------------|-------------|-------------|-------------|-------------|-------------|-----------|------|------|----------------|
|                   | R@1      | R@10 | R@50        | R@1        | R@10        | R@50        | R@1         | R@10        | R@50        | R@1       | R@10 | R@50 |                |
| AnyLoc            | 5.1      | 22.8 | 54.4        | 14.2       | 36.0        | 57.4        | <b>21.7</b> | <b>43.6</b> | <b>61.2</b> | 18.9      | 39.4 | 57.2 | 150%           |
| CosPlace          | 2.5      | 10.5 | <b>25.0</b> | <b>4.7</b> | <b>11.2</b> | 22.1        | 4.3         | 9.3         | 16.1        | 2.0       | 6.0  | 11.0 | 100%           |
| EigenPlaces       | 2.3      | 11.0 | 25.2        | 6.7        | <b>15.4</b> | <b>27.1</b> | <b>7.1</b>  | 15.2        | 22.4        | 3.5       | 9.9  | 17.8 | 150%           |
| MixVPR            | 1.1      | 8.0  | <b>26.1</b> | 4.1        | 12.5        | 25.2        | <b>7.2</b>  | <b>14.5</b> | 24.2        | 5.5       | 10.6 | 17.4 | 150%           |
| NetVLAD           | 1.0      | 7.3  | <b>21.9</b> | <b>1.4</b> | <b>8.2</b>  | 18.7        | 0.9         | 4.0         | 9.5         | 0.6       | 2.3  | 5.2  | 100%           |
| SALAD             | 2.1      | 14.0 | <b>36.4</b> | 4.8        | 16.0        | 33.8        | <b>6.9</b>  | <b>17.3</b> | 30.0        | 5.8       | 14.1 | 24.1 | 150%           |

Таблица 3: Сравнение различных масштабов карты на датасете VPAir query, значение Recall в %

| Метод локализации | Zoom 50% |      |      | Zoom 100%  |             |             | Zoom 150%   |             |             | Zoom 200%   |             |      | Лучший масштаб |
|-------------------|----------|------|------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|------|----------------|
|                   | R@1      | R@10 | R@50 | R@1        | R@10        | R@50        | R@1         | R@10        | R@50        | R@1         | R@10        | R@50 |                |
| AnyLoc            | 13.3     | 42.8 | 72.0 | 42.9       | 69.3        | 83.2        | 59.5        | 76.3        | <b>86.3</b> | <b>59.7</b> | <b>76.9</b> | 84.5 | 200%           |
| CosPlace          | 8.8      | 29.1 | 50.8 | 24.8       | 49.4        | 66.7        | <b>28.8</b> | <b>54.7</b> | <b>70.0</b> | 24.3        | 48.2        | 63.9 | 150%           |
| EigenPlaces       | 7.5      | 28.5 | 56.5 | 34.0       | 60.4        | 75.5        | <b>48.4</b> | <b>72.5</b> | <b>82.0</b> | 46.9        | 69.1        | 79.7 | 150%           |
| MixVPR            | 5.7      | 22.8 | 50.8 | 43.5       | 69.2        | 83.1        | <b>67.2</b> | <b>85.1</b> | <b>91.9</b> | 66.0        | 84.1        | 90.5 | 150%           |
| NetVLAD           | 2.1      | 13.1 | 39.1 | <b>9.4</b> | <b>26.3</b> | <b>47.1</b> | 9.2         | 22.6        | 40.7        | 6.8         | 17.1        | 30.0 | 100%           |
| SALAD             | 10.4     | 34.2 | 61.6 | 32.7       | 56.3        | 76.0        | <b>47.2</b> | <b>70.0</b> | <b>82.1</b> | 47.0        | 67.4        | 78.9 | 150%           |

Таблица 4: Сравнение различных масштабов карты на датасете VPAir reference, значение Recall в %

| Метод локализации | Zoom 50% |      |             | Zoom 100%  |             |             | Zoom 150% |      |             | Zoom 200%   |             |             | Zoom 250% |      |      | Лучший масштаб |
|-------------------|----------|------|-------------|------------|-------------|-------------|-----------|------|-------------|-------------|-------------|-------------|-----------|------|------|----------------|
|                   | R@1      | R@10 | R@50        | R@1        | R@10        | R@50        | R@1       | R@10 | R@50        | R@1         | R@10        | R@50        | R@1       | R@10 | R@50 |                |
| AnyLoc            | 3.7      | 19.1 | 40.0        | 10.7       | 30.0        | 44.3        | 19.3      | 34.1 | <b>48.5</b> | <b>20.0</b> | <b>35.4</b> | 46.5        | 18.9      | 32.6 | 43.7 | 200%           |
| CosPlace          | 5.0      | 12.6 | <b>28.7</b> | <b>6.1</b> | <b>15.4</b> | 26.5        | 5.0       | 13.3 | 26.1        | 6.1         | 12.8        | 21.7        | 4.1       | 11.5 | 20.2 | 100%           |
| EigenPlaces       | 4.6      | 12.0 | 31.1        | <b>7.6</b> | <b>21.3</b> | <b>37.6</b> | 7.4       | 18.9 | 35.0        | 6.5         | 16.1        | 27.6        | 4.8       | 15.7 | 27.0 | 100%           |
| MixVPR            | 2.6      | 14.6 | 27.4        | 8.0        | 22.8        | 38.7        | 17.4      | 35.4 | 50.2        | <b>20.7</b> | <b>38.9</b> | <b>53.5</b> | 17.0      | 34.6 | 50.4 | 200%           |
| NetVLAD           | 2.6      | 7.2  | 14.3        | <b>4.8</b> | <b>8.5</b>  | <b>17.0</b> | 0.9       | 5.9  | 16.1        | 0.9         | 5.2         | 10.4        | 0.9       | 3.3  | 6.7  | 100%           |
| SALAD             | 6.1      | 18.7 | 37.8        | 12.2       | 24.1        | 40.7        | 16.7      | 29.8 | 45.0        | <b>18.3</b> | <b>31.7</b> | <b>46.5</b> | 15.9      | 28.9 | 42.8 | 200%           |

Таблица 5: Сравнение различных масштабов карты на датасете ALTO, значение Recall в %

Исходя из результатов, можно сделать вывод, что оптимальный уровень приближения зависит не только от тестового датасета, но и от используемого метода локализации. В целом, для методов CosPlace,

EigenPlaces и NetVLAD более подходит уровень приближения в 100% - 150%, а алгоритмы AnyLoc, MixVPR и SALAD лучше работают с масштабом в 150% - 200%. Для выбора наилучшего масштаба в первую очередь учитывались результаты Recall@1. Во всех дальнейших экспериментах, для каждого алгоритма использовался только соответствующий ему оптимальный уровень масштаба.

### 5.3.3. Выбор оптимального уровня перекрытия между соседними тайлами карты

Уровень перекрытия между соседними тайлами также важен для эффективной работы различных алгоритмов визуальной геолокализации. Чем он больше, тем выше шанс, что для очередного кадра-запроса найдётся максимально схожее по покрытию изображение из базы данных. Однако высокий уровень перекрытия также существенно увеличивает количество тайлов в базе данных. Поэтому на третьем этапе было проведено сравнение различных уровней перекрытия, а именно 0%, 25% и 50%. Результаты сравнения можно увидеть в таблицах 6, 7 и 8 для датасетов VPAir query, VPAir reference и ALTO соответственно.

| Метод локализации | Overlap 0% |      |      | Overlap 25% |            |             | Overlap 50% |             |             | Лучший уровень |
|-------------------|------------|------|------|-------------|------------|-------------|-------------|-------------|-------------|----------------|
|                   | R@1        | R@10 | R@50 | R@1         | R@10       | R@50        | R@1         | R@10        | R@50        |                |
| AnyLoc            | 12.9       | 30.3 | 48.9 | 21.7        | 43.6       | 61.2        | <b>26.3</b> | <b>49.2</b> | <b>66.0</b> | 50%            |
| CosPlace          | 2.6        | 9.3  | 19.3 | 4.7         | 11.2       | <b>22.1</b> | <b>4.9</b>  | <b>11.6</b> | 21.8        | —              |
| EigenPlaces       | 2.6        | 8.6  | 15.9 | 7.1         | 15.2       | 22.4        | <b>7.2</b>  | <b>15.5</b> | <b>23.8</b> | —              |
| MixVPR            | 3.6        | 9.6  | 18.9 | <b>7.2</b>  | 14.5       | <b>24.2</b> | 6.9         | <b>14.8</b> | <b>24.2</b> | —              |
| NetVLAD           | 1.4        | 6.9  | 17.3 | 1.4         | <b>8.2</b> | <b>18.7</b> | <b>1.7</b>  | 6.6         | 15.9        | —              |
| SALAD             | 4.7        | 12.8 | 24.9 | 6.9         | 17.3       | 30.0        | <b>9.3</b>  | <b>21.7</b> | <b>33.3</b> | —              |

Таблица 6: Сравнение различных уровней перекрытия между тайлами на датасете VPAir query, значение Recall в %

| Метод<br>локализации | Overlap 0% |      |      | Overlap 25% |      |             | Overlap 50% |             |             | Лучший<br>уровень |
|----------------------|------------|------|------|-------------|------|-------------|-------------|-------------|-------------|-------------------|
|                      | R@1        | R@10 | R@50 | R@1         | R@10 | R@50        | R@1         | R@10        | R@50        |                   |
| AnyLoc               | 44.1       | 64.6 | 76.0 | 59.7        | 76.9 | 84.5        | <b>69.1</b> | <b>81.1</b> | <b>86.8</b> | 50%               |
| CosPlace             | 21.8       | 41.8 | 57.0 | 28.8        | 54.7 | 70.0        | <b>39.0</b> | <b>60.3</b> | <b>73.4</b> | —                 |
| EigenPlaces          | 34.3       | 53.8 | 66.7 | 48.4        | 72.5 | 82.0        | <b>60.5</b> | <b>77.5</b> | <b>84.8</b> | —                 |
| MixVPR               | 45.5       | 65.7 | 77.8 | 67.2        | 85.1 | 91.9        | <b>78.9</b> | <b>90.7</b> | <b>94.5</b> | 50%               |
| NetVLAD              | 7.9        | 22.1 | 40.0 | 9.4         | 26.3 | <b>47.1</b> | <b>11.6</b> | <b>28.9</b> | 45.8        | —                 |
| SALAD                | 31.1       | 50.8 | 64.9 | 47.2        | 70.0 | 82.1        | <b>65.7</b> | <b>80.7</b> | <b>88.5</b> | 50%               |

Таблица 7: Сравнение различных уровней перекрытия между тайлами на датасете VPAir reference, значение Recall в %

| Метод<br>локализации | Overlap 0% |      |      | Overlap 25% |      |             | Overlap 50% |             |             | Лучший<br>уровень |
|----------------------|------------|------|------|-------------|------|-------------|-------------|-------------|-------------|-------------------|
|                      | R@1        | R@10 | R@50 | R@1         | R@10 | R@50        | R@1         | R@10        | R@50        |                   |
| AnyLoc               | 14.6       | 28.7 | 39.6 | 20.0        | 35.4 | 46.5        | <b>25.7</b> | <b>39.1</b> | <b>48.0</b> | 50%               |
| CosPlace             | 4.1        | 10.9 | 22.0 | 6.1         | 15.4 | 26.5        | <b>7.2</b>  | <b>16.3</b> | <b>28.0</b> | —                 |
| EigenPlaces          | 7.0        | 17.4 | 32.2 | 7.6         | 21.3 | <b>37.6</b> | <b>11.1</b> | <b>23.7</b> | 36.1        | —                 |
| MixVPR               | 13.5       | 24.8 | 40.7 | 20.7        | 38.9 | 53.5        | <b>32.2</b> | <b>49.3</b> | <b>61.1</b> | 50%               |
| NetVLAD              | 2.6        | 9.6  | 17.0 | <b>4.8</b>  | 8.5  | 17.0        | 4.6         | <b>11.7</b> | <b>20.2</b> | —                 |
| SALAD                | 9.8        | 20.9 | 35.4 | 18.3        | 31.7 | 46.5        | <b>23.9</b> | <b>40.9</b> | <b>56.7</b> | 50%               |

Таблица 8: Сравнение различных уровней перекрытия между тайлами на датасете ALTO, значение Recall в %

Результаты показывают, что увеличение уровня перекрытия между кадрами ведёт к повышению качества работы систем глобальной локализации, а главным ограничением в данной ситуации становится время и память, необходимые для создания базы данных. Опираясь на результаты сравнения, было принято решение проводить дальнейшие эксперименты только на методах, которые показали себя лучше всего. Для датасетов VPAir reference и ALTO это методы AnyLoc, MixVPR и SALAD, а для VPAir query — только AnyLoc.

#### 5.3.4. Сравнение различных сопоставителей ключевых точек

Как уже было упомянуто ранее, существуют разные подходы к глобальной локализации. В данной работе рассматривается два из них: первый подход выбирает один лучший кадр только с помощью дескрипторных методов, а другой использует несколько лучших предсказаний из дескрипторных методов, чтобы выбрать наилучший кадр с помощью

методов на ключевых точках. Поэтому целью следующего эксперимента было не только сравнение различных сопоставителей ключевых точек, но и исследование целесообразности их использования. Для достижения цели ранее отобранные методы были протестированы в трёх конфигурациях. Первая из них просто возвращает один лучший кадр с помощью дескрипторных подходов, вторая и третья переранжирует 10 и 50 кадров соответственно и возвращает наиболее подходящий снимок. В качестве сопоставителей ключевых точек использовались SuperGlue и LightGlue. Результаты сравнения можно увидеть в таблицах 9, 10 и 11 для датасетов VPAir query, VPAir reference и ALTO соответственно.

| Метод<br>локализации | SuperGlue   |      |      | LightGlue   |      |      |
|----------------------|-------------|------|------|-------------|------|------|
|                      | R@1         | R@10 | R@50 | R@1         | R@10 | R@50 |
| AnyLoc               | <b>26.3</b> | 25.5 | 24.5 | <b>26.3</b> | 20.0 | 17.8 |

Таблица 9: Сравнение различных сопоставителей ключевых точек на датасете VPAir query, значение Recall в %

| Метод<br>локализации | SuperGlue |      |             | LightGlue |             |      |
|----------------------|-----------|------|-------------|-----------|-------------|------|
|                      | R@1       | R@10 | R@50        | R@1       | R@10        | R@50 |
| AnyLoc               | 69.1      | 72.4 | <b>72.8</b> | 69.1      | 72.3        | 72.0 |
| MixVPR               | 78.9      | 80.3 | 79.0        | 78.9      | <b>82.2</b> | 80.5 |
| SALAD                | 65.7      | 72.7 | <b>75.5</b> | 65.7      | 72.6        | 75.2 |

Таблица 10: Сравнение различных сопоставителей ключевых точек на датасете VPAir reference, значение Recall в %

| Метод<br>локализации | SuperGlue |      |             | LightGlue |      |             |
|----------------------|-----------|------|-------------|-----------|------|-------------|
|                      | R@1       | R@10 | R@50        | R@1       | R@10 | R@50        |
| AnyLoc               | 25.7      | 29.1 | 31.3        | 25.7      | 30.7 | <b>31.7</b> |
| MixVPR               | 32.2      | 34.6 | 37.4        | 32.2      | 36.7 | <b>39.1</b> |
| SALAD                | 23.9      | 28.9 | <b>39.3</b> | 23.9      | 29.6 | 38.0        |

Таблица 11: Сравнение различных сопоставителей ключевых точек на датасете ALTO, значение Recall в %

Результаты показывают, что единый ответ на поставленный вопрос отсутствует. В случае датасета VPAir query подход с использованием

только дескрипторного метода AnyLoc показывает наилучший результат, а для последовательности VPAir reference SuperGlue и LightGlue дают прирост, но он весьма незначителен. В свою очередь, результаты на датасете ALTO говорят о том, что и SuperGlue, и LightGlue дают прирост, при этом LightGlue в среднем немного лучше. Поэтому в последующих экспериментах было принято решение не использовать методы, работающие на ключевых точках, на последовательностях набора данных VPAir, но применять LightGlue на датасете ALTO.

### 5.3.5. Сравнение методов визуальной геолокализации

В ходе предыдущих экспериментов были отобраны лучшие методы глобальной локализации и найдены оптимальные конфигурации для них. На данном же этапе было проведено их сравнение с помощью метрики Recall для геопривязки. В качестве пороговых использовались значения 10 метров, 50 метров и 100 метров. Результаты сравнения можно увидеть в таблицах 12, 13 и 14 для датасетов VPAir query, VPAir reference и ALTO соответственно.

| Метод локализации | 10 метров | 50 метров | 100 метров |
|-------------------|-----------|-----------|------------|
| AnyLoc            | 1.8       | 7.0       | 10.8       |

Таблица 12: Сравнение методов визуальной геолокализации на датасете VPAir query, значение Recall в %

| Метод локализации | 10 метров   | 50 метров   | 100 метров  |
|-------------------|-------------|-------------|-------------|
| AnyLoc            | 42.4        | 58.8        | 62.4        |
| MixVPR            | <b>50.3</b> | <b>65.5</b> | <b>69.3</b> |
| SALAD             | 43.5        | 56.7        | 59.3        |

Таблица 13: Сравнение методов визуальной геолокализации на датасете VPAir reference, значение Recall в %

| Метод<br>локализации | 10 метров  | 50 метров   | 100 метров  |
|----------------------|------------|-------------|-------------|
| AnyLoc + LightGlue   | 3.0        | 24.3        | 27.0        |
| MixVPR + LightGlue   | <b>4.6</b> | <b>32.4</b> | <b>35.7</b> |
| SALAD + LightGlue    | 3.5        | 31.3        | 33.9        |

Таблица 14: Сравнение методов визуальной геолокализации на датасете ALTO, значение Recall в %

### 5.3.6. Производительность методов визуальной геолокализации

На последнем этапе экспериментов было проведено сравнение производительности ранее выбранных методов. Для этого для каждой последовательности и метода было измерено время обработки каждого кадра-запроса, а затем было посчитано среднее и стандартное отклонение. Для запуска методов использовалась машина: AMD Ryzen Threadripper 3970X, 128 Gb RAM, NVIDIA GeForce RTX 3090, Ubuntu 22.04. Результаты сравнения можно увидеть в таблицах 15, 16 и 17 для датасетов VPAir query, VPAir reference и ALTO соответственно.

| Метод<br>локализации | Время вычисления (сек.) |
|----------------------|-------------------------|
| AnyLoc               | $2.78 \pm 0.11$         |

Таблица 15: Производительность методов визуальной геолокализации на датасете VPAir query

| Метод<br>локализации | Время вычисления (сек.)           |
|----------------------|-----------------------------------|
| AnyLoc               | $3.97 \pm 0.09$                   |
| MixVPR               | <b><math>0.80 \pm 0.09</math></b> |
| SALAD                | $0.97 \pm 0.08$                   |

Таблица 16: Производительность методов визуальной геолокализации на датасете VPAir reference

| Метод<br>локализации | Время вычисления (сек.)           |
|----------------------|-----------------------------------|
| AnyLoc + LightGlue   | $2.18 \pm 0.08$                   |
| MixVPR + LightGlue   | <b><math>0.78 \pm 0.09</math></b> |
| SALAD + LightGlue    | $0.88 \pm 0.07$                   |

Таблица 17: Производительность методов визуальной геолокализации на датасете ALTO

Результаты показывают, что быстрее других работает метод MixVPR, также в скорость 1 кадр/сек. укладывается SALAD. Применение LightGlue не оказывает существенного влияния на время работы методов.

## 5.4. Выводы

Результаты экспериментального сравнения показывают, что не на все вопросы о формировании базы данных из спутниковых снимков можно ответить однозначно. Оптимальный уровень масштаба тайлов карты зависит не только от масштаба кадров-запросов, но и от используемых методов глобальной локализации. При этом на вопрос об оптимальном уровне перекрытия между кадрами можно ответить однозначно — его увеличение ведёт к улучшению результатов. Однако изменение этих параметров существенно влияет на количество изображений в базе данных, что может оказаться критичным в условиях ограниченных ресурсов.

Необходимость использования методов, использующих ключевые точки, также спорна. В большинстве случаев это действительно ведёт к росту в качестве, иногда весьма значительному. Однако в некоторых ситуациях их применение лишь ухудшает результаты. Стоит отметить, что использование нового относительно SuperGlue сопоставителя LightGlue не оказывает существенного влияния на общее время работы конвейера визуальной геолокализации.

Методы глобальной локализации CosPlace, EigenPlaces и NetVLAD показали себя существенно хуже других и были отброшены на ранних

этапах экспериментов. Среди оставшихся особо выделяется MixVPR, который показывает самые высокие результаты в качестве и производительности на двух из трёх датасетов. На VPAir query лучше других работает AnyLoc, но при этом время обработки одного кадра приближается к трём секундам.

## 6. Заключение

В рамках работы были выполнены следующие задачи.

1. Выполнен обзор существующих решений для визуальной геолокации, а также глобальной локализации и локального выравнивания по отдельности.
2. Предложены и реализованы метрики для оценки результата работы различных методов локализации.
3. Спроектирован и разработан инструмент, позволяющий сравнивать различные подходы к визуальной локализации.
4. Проведено экспериментальное исследование различных методов локализации с использованием разработанного инструмента.

Ссылка на github репозиторий реализованного инструмента – <https://github.com/prime-slam/aero-vloc>.

## Список литературы

- [1] A 2D Georeferenced Map Aided Visual-Inertial System for Precise UAV Localization / Mao Jun, Zhang Lilian, He Xiaofeng et al. // 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) / IEEE. — 2022. — P. 4455–4462.
- [2] ALTO: A Large-Scale Dataset for UAV Visual Place Recognition and Localization / Ivan Cisneros, Peng Yin, Ji Zhang et al. // arXiv preprint arXiv:2207.12317. — 2022.
- [3] Adaptive-attentive geolocalization from few queries: A hybrid approach / Gabriele Moreno Berton, Valerio Paolicelli, Carlo Masone, Barbara Caputo // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. — 2021. — P. 2918–2927.
- [4] Ali-Bey Amar, Chaib-Draa Brahim, Giguere Philippe. Mixvpr: Feature mixing for visual place recognition // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. — 2023. — P. 2998–3007.
- [5] Anyloc: Towards universal visual place recognition / Nikhil Keetha, Avneesh Mishra, Jay Karhade et al. // arXiv preprint arXiv:2308.00688. — 2023.
- [6] Arandjelovic Relja, Zisserman Andrew. All about VLAD // Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. — 2013. — P. 1578–1585.
- [7] Are Local Features All You Need for Cross-Domain Visual Place Recognition? / Giovanni Barbarani, Mohamad Mostafa, Hajali Bayramov et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2023. — P. 6154–6164.
- [8] Attentional pyramid pooling of salient visual residuals for place recognition / Guohao Peng, Jun Zhang, Heshan Li, Danwei Wang // Proceed-

ings of the IEEE/CVF International Conference on Computer Vision. — 2021. — P. 885–894.

- [9] Automatic UAV image geo-registration by matching UAV images to georeferenced image data / Xiangyu Zhuo, Tobias Koch, Franz Kurz et al. // Remote Sensing. — 2017. — Vol. 9, no. 4. — P. 376.
- [10] Berton Gabriele, Masone Carlo, Caputo Barbara. Rethinking visual geo-localization for large-scale applications // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2022. — P. 4878–4888.
- [11] Bradski Gary. The openCV library. // Dr. Dobb's Journal: Software Tools for the Professional Programmer. — 2000. — Vol. 25, no. 11. — P. 120–123.
- [12] Bui Duc Viet, Kubo Masao, Sato Hiroshi. Attention-based neural network with Generalized Mean Pooling for cross-view geo-localization between UAV and satellite // Artificial Life and Robotics. — 2023. — P. 1–11.
- [13] Choi Junho, Myung Hyun. BRM localization: UAV localization in GNSS-denied environments based on matching of numerical map and UAV images // 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) / IEEE. — 2020. — P. 4537–4544.
- [14] Coarse-to-Fine Visual Place Recognition / Junkun Qi, Rui Wang, Chuan Wang, Xiaochun Cao // International Conference on Neural Information Processing / Springer. — 2021. — P. 28–39.
- [15] CurriculumLoc: Enhancing Cross-Domain Geolocalization through Multi-Stage Refinement / Boni Hu, Lin Chen, Runjian Chen et al. // IEEE Transactions on Geoscience and Remote Sensing. — 2024.
- [16] DeTone Daniel, Malisiewicz Tomasz, Rabinovich Andrew. Superpoint: Self-supervised interest point detection and description // Proceedings

of the IEEE conference on computer vision and pattern recognition workshops. — 2018. — P. 224–236.

- [17] Deep residual learning for image recognition / Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — P. 770–778.
- [18] Deep visual geo-localization benchmark / Gabriele Berton, Riccardo Mereu, Gabriele Trivigno et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2022. — P. 5396–5407.
- [19] Dinov2: Learning robust visual features without supervision / Maxime Oquab, Timothée Darcet, Théo Moutakanni et al. // arXiv preprint arXiv:2304.07193. — 2023.
- [20] EigenPlaces: Training Viewpoint Robust Models for Visual Place Recognition / Gabriele Berton, Gabriele Trivigno, Barbara Caputo, Carlo Masone // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2023. — P. 11080–11090.
- [21] Fast aerial image geolocalization using the projective-invariant contour feature / Yongfei Li, Shicheng Wang, Hao He et al. // Remote Sensing. — 2021. — Vol. 13, no. 3. — P. 490.
- [22] GLONASS. — URL: [https://glonass-iac.ru/en/about\\_glonass/](https://glonass-iac.ru/en/about_glonass/) (дата обращения: 2023-12-01).
- [23] Garg Sourav, Fischer Tobias, Milford Michael. Where is your place, visual place recognition? // IJCAI. — Vol. 8. — 2021. — P. 4416–4425.
- [24] Geolocalization with aerial image sequence for UAVs / Yongfei Li, Hao He, Dongfang Yang et al. // Autonomous Robots. — 2020. — Vol. 44, no. 7. — P. 1199–1215.
- [25] The Global Positioning System. — URL: <https://www.gps.gov> (дата обращения: 2023-12-01).

- [26] Google Maps API. — URL: <https://developers.google.com/maps> (дата обращения: 2023-12-01).
- [27] Google map aided visual navigation for UAVs in GPS-denied environment / Mo Shan, Fei Wang, Feng Lin et al. // 2015 IEEE international conference on robotics and biomimetics (ROBIO) / IEEE. — 2015. — P. 114–119.
- [28] Gurgu Marius-Mihail, Queralta Jorge Peña, Westerlund Tomi. Vision-Based GNSS-Free Localization for UAVs in the Wild // 2022 7th International Conference on Mechanical Engineering and Robotics Research (ICMERR) / IEEE. — 2022. — P. 7–12.
- [29] Harvey Winthrop, Rainwater Chase, Cothren Jackson. Direct aerial visual geolocalization using deep neural networks // Remote Sensing. — 2021. — Vol. 13, no. 19. — P. 4017.
- [30] Image matching from handcrafted to deep features: A survey / Jiayi Ma, Xingyu Jiang, Aoxiang Fan et al. // International Journal of Computer Vision. — 2021. — Vol. 129, no. 1. — P. 23–79.
- [31] Implications of web Mercator and its use in online mapping / Sarah E Battersby, Michael P Finn, E Lynn Usery, Kristina H Yamamoto // Cartographica: The International Journal for Geographic Information and Geovisualization. — 2014. — Vol. 49, no. 2. — P. 85–101.
- [32] Izquierdo Sergio, Civera Javier. Optimal transport aggregation for visual place recognition // arXiv preprint arXiv:2311.15937. — 2023.
- [33] Jin Kim Hyo, Dunn Enrique, Frahm Jan-Michael. Learned contextual feature reweighting for image geo-localization // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2017. — P. 2136–2145.
- [34] Kinnari Jouko, Verdoja Francesco, Kyrki Ville. GNSS-denied geolocalization of UAVs by visual matching of onboard camera images

with orthophotos // 2021 20th International Conference on Advanced Robotics (ICAR) / IEEE. — 2021. — P. 555–562.

- [35] Lindenberger Philipp, Sarlin Paul-Edouard, Pollefeys Marc. Light-Glue: Local Feature Matching at Light Speed // arXiv preprint arXiv:2306.13643. — 2023.
- [36] Liu Cuiyin, Xu Jishang, Wang Feng. A review of keypoints' detection and feature description in image registration // Scientific programming. — 2021. — Vol. 2021. — P. 1–25.
- [37] Liu Liu, Li Hongdong, Dai Yuchao. Stochastic attraction-repulsion embedding for large scale image localization // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2019. — P. 2570–2579.
- [38] Liu Xing. A cross-view UAV image localization method based on directional feature alignment and visual transformer // Academic Journal of Computing & Information Science. — Vol. 5, no. 12. — P. 22–28.
- [39] Lowe David G. Distinctive image features from scale-invariant keypoints // International journal of computer vision. — 2004. — Vol. 60. — P. 91–110.
- [40] Mapillary street-level sequences: A dataset for lifelong place recognition / Frederik Warburg, Soren Hauberg, Manuel Lopez-Antequera et al. // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2020. — P. 2626–2635.
- [41] Masone Carlo, Caputo Barbara. A survey on deep visual place recognition // IEEE Access. — 2021. — Vol. 9. — P. 19516–19547.
- [42] Masoud Alaa A, Saad Ahmed M, El Shafaey Osama NH. Geotechnical database building and 3D modeling of the soil in Medina, Saudi Arabia // Arabian Journal of Geosciences. — 2022. — Vol. 15, no. 6. — P. 506.

- [43] Mughal Muhammad Hamza, Khokhar Muhammad Jawad, Shahzad Muhammad. Assisting UAV localization via deep contextual image matching // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. — 2021. — Vol. 14. — P. 2445–2457.
- [44] Multiple-environment Self-adaptive Network for Aerial-view Geolocalization / Tingyu Wang, Zhedong Zheng, Yaoqi Sun et al. // Pattern Recognition. — 2024. — P. 110363.
- [45] NetVLAD: CNN architecture for weakly supervised place recognition / Relja Arandjelovic, Petr Gronat, Akihiko Torii et al. // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — P. 5297–5307.
- [46] ORB: An efficient alternative to SIFT or SURF / Ethan Rublee, Vincent Rabaud, Kurt Konolige, Gary Bradski // 2011 International conference on computer vision / Ieee. — 2011. — P. 2564–2571.
- [47] Patch-netvlad: Multi-scale fusion of locally-global descriptors for place recognition / Stephen Hausler, Sourav Garg, Ming Xu et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2021. — P. 14141–14152.
- [48] Progressive keypoint localization and refinement in image matching / Fabio Bellavia, Luca Morelli, Carlo Colombo, Fabio Remondino // International Conference on Image Analysis and Processing / Springer. — 2023. — P. 322–334.
- [49] Range–Visual–Inertial Odometry with Coarse-to-Fine Image Registration Fusion for UAV Localization / Yun Hao, Mengfan He, Yuzhen Liu et al. // Drones. — 2023. — Vol. 7, no. 8. — P. 540.
- [50] Real-time geo-localization using satellite imagery and topography for unmanned aerial vehicles / Shuxiao Chen, Xiangyu Wu, Mark W Mueller, Koushil Sreenath // 2021 IEEE/RSJ International

Conference on Intelligent Robots and Systems (IROS) / IEEE. — 2021. — P. 2275–2281.

- [51] SUES-200: A multi-height multi-scene cross-view image benchmark across drone and satellite / Runzhe Zhu, Ling Yin, Mingze Yang et al. // IEEE Transactions on Circuits and Systems for Video Technology. — 2023.
- [52] Schleiss Michael. Translating aerial images into street-map-like representations for visual self-localization of UAVs // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. — 2019. — Vol. 42. — P. 575–580.
- [53] Schleiss Michael, Rouatbi Fahmi, Cremers Daniel. VPAIR–Aerial Visual Place Recognition and Localization in Large-scale Outdoor Environments // arXiv preprint arXiv:2205.11567. — 2022.
- [54] Semantic reinforced attention learning for visual place recognition / Guohao Peng, Yufeng Yue, Jun Zhang et al. // 2021 IEEE International Conference on Robotics and Automation (ICRA) / IEEE. — 2021. — P. 13415–13422.
- [55] Sentinel. — URL: <https://s2maps.eu/> (дата обращения: 2023-12-01).
- [56] Speeded-up robust features (SURF) / Herbert Bay, Andreas Ess, Tinne Tuytelaars, Luc Van Gool // Computer vision and image understanding. — 2008. — Vol. 110, no. 3. — P. 346–359.
- [57] Superglue: Learning feature matching with graph neural networks / Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, Andrew Rabinovich // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2020. — P. 4938–4947.
- [58] Target localization based on cross-view matching between UAV and satellite / REN Kan, DING Lei, WAN Minjie et al. // Chinese Journal of Aeronautics. — 2022. — Vol. 35, no. 9. — P. 333–341.

- [59] Uav's status is worth considering: A fusion representations matching method for geo-localization / Runzhe Zhu, Mingze Yang, Ling Yin et al. // Sensors. — 2023. — Vol. 23, no. 2. — P. 720.
- [60] Umar I, Yelwa N, Jibrin Babangida. 3D Geological Models in Ground Investigation: Examples from the United Kingdom // Journal of Scientific Research and Reports. — 2016. — Vol. 9, no. 3. — P. 1–6.
- [61] Van Rossum Guido, Drake Jr Fred L. Python tutorial. — Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands, 1995. — Vol. 620.
- [62] Viewpoint invariant dense matching for visual geolocalization / Gabriele Berton, Carlo Masone, Valerio Paolicelli, Barbara Caputo // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2021. — P. 12169–12178.
- [63] Vision-Based UAV Localization System in Denial Environments / Ming Dai, Jinglin Huang, Jiedong Zhuang et al. // arXiv preprint arXiv:2201.09201. — 2022.
- [64] Vision-based UAV self-positioning in low-altitude urban environments / Ming Dai, Enhui Zheng, Zhenhua Feng et al. // IEEE Transactions on Image Processing. — 2023.
- [65] Visual place recognition: A survey / Stephanie Lowry, Niko Sünderhauf, Paul Newman et al. // IEEE transactions on robotics. — 2015. — Vol. 32, no. 1. — P. 1–19.
- [66] Workman Scott, Souvenir Richard, Jacobs Nathan. Wide-area image geolocalization with aerial reference imagery // Proceedings of the IEEE International Conference on Computer Vision. — 2015. — P. 3961–3969.
- [67] Zaffar Mubariz. Visual Place Recognition for Autonomous Robots : Ph.D. thesis / Mubariz Zaffar ; University of Essex. — 2020.

- [68] Zhao Yiming, Huang Xinming, Zhang Ziming. Deep lucas-kanade homography for multimodal image alignment // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. — 2021. — P. 15950–15959.
- [69] Zheng Zhedong, Wei Yunchao, Yang Yi. University-1652: A multi-view multi-source benchmark for drone-based geo-localization // Proceedings of the 28th ACM international conference on Multimedia. — 2020. — P. 1395–1403.
- [70] The faiss library / Matthijs Douze, Alexandr Guzhva, Chengqi Deng et al. // arXiv preprint arXiv:2401.08281. — 2024.
- [71] A faster and more effective cross-view matching method of uav and satellite images for uav geolocalization / Jiedong Zhuang, Ming Dai, Xuruoyan Chen, Enhui Zheng // Remote Sensing. — 2021. — Vol. 13, no. 19. — P. 3979.
- [72] A practical cross-view image matching method between UAV and satellite for UAV-based geo-localization / Lirong Ding, Ji Zhou, Lingxuan Meng, Zhiyong Long // Remote Sensing. — 2020. — Vol. 13, no. 1. — P. 47.
- [73] A semantic vector map-based approach for aircraft positioning in GNSS/GPS denied large-scale environment / Chenguang Ouyang, Suxing Hu, Fengqi Long et al. // Defence Technology. — 2023.
- [74] A transformer-based feature segmentation and region alignment method for uav-view geo-localization / Ming Dai, Jianhong Hu, Jiedong Zhuang, Enhui Zheng // IEEE Transactions on Circuits and Systems for Video Technology. — 2021. — Vol. 32, no. 7. — P. 4376–4389.