

УДК 004.93

DOI: [10.26102/2310-6018/2026.58.7.016](https://doi.org/10.26102/2310-6018/2026.58.7.016)

Определение геолокации по кадрам видеорегистратора с использованием ResNet

А.Р. Басханов✉

*Финансовый университет при Правительстве Российской Федерации, Москва,
Российская Федерация*

Резюме. Актуальность исследования обусловлена ростом объемов уличных снимков и необходимостью их автоматической географической привязки без использования GPS-метаданных. В связи с этим статья направлена на оценку возможности определения координат фотографий в пределах одного города с помощью глубоких нейросетей. Ведущим методом исследования является применение сверточных нейросетей ResNet18 и ResNet50 для регрессии широты и долготы по кадрам размером 512×512 пикселей. Для сбора данных использована открытая платформа Mapillary, а для равномерного покрытия территории Уфы разработана программа с разбиением города на сетку из 10 000 ячеек. Всего собрано 88 798 изображений с точной привязкой к координатам. В статье представлены количественные результаты обучения моделей. Модель на базе ResNet50 достигает медианной ошибки локализации около 3,1 км после 13 эпох, при этом 92,6 % предсказаний попадают в радиус 10 км от истинного положения. Методом Smooth Grad-CAM визуализированы области, на которые модель опирается при определении координат; показано, что модель фокусируется на фасадах зданий, перекрестках и дорожной разметке. Материалы статьи представляют практическую ценность для систем автоматической геопривязки больших массивов уличных фотографий. Работа также предлагает воспроизводимый пайплайн для городской геолокации и анализирует поведенческое внимание нейросетей в данной задаче.

Ключевые слова: геолокация изображений, сверточные нейросети, ResNet, анализ городских сцен, глубокая регрессия, визуализация внимания, Smooth Grad-CAM, Mapillary.

Для цитирования: Басханов А.Р. Определение геолокации по кадрам видеорегистратора с использованием ResNet. *Моделирование, оптимизация и информационные технологии.* 2026;14(7). URL: <https://moitvvt.ru/ru/journal/article?id=2399> DOI: 10.26102/2310-6018/2026.58.7.016

Geolocation determination from dashcam footage using ResNet

A.R. Baskhanov✉

*Financial University under the Government of the Russian Federation, Moscow,
the Russian Federation*

Abstract. The relevance of this study is driven by the growing volume of street-level imagery and the need for automatic georeferencing without relying on GPS metadata. Accordingly, this paper aims to assess the feasibility of determining photograph coordinates within a single city using deep neural networks. The leading research method is the application of convolutional neural networks ResNet18 and ResNet50 for latitude and longitude regression from 512×512-pixel frames. Data were collected using the open Mapillary platform, and a custom program with a 10,000-cell grid was developed to ensure uniform coverage of Ufa. A total of 88,798 geotagged images were collected. The paper presents quantitative results of model training. The ResNet50-based model achieves a median localization error of approximately 3.1 km after 13 epochs, with 92.6 % of predictions falling within a 10 km radius of the ground truth. Smooth Grad-CAM visualizations reveal that the model focuses on building facades, intersections, road markings, and other infrastructure elements – semantically meaningful landmarks for urban geolocation. The materials of the article are of practical value for automatic georeferencing

systems for large collections of street-level imagery and for the development of visual localization methods for autonomous vehicles. The work also provides a reproducible pipeline for city-level image geolocation and analyses the behavioral attention of convolutional neural networks in this task.

Keywords: image geolocation, convolutional neural networks, ResNet, urban scene analysis, deep regression, attention visualization, Smooth Grad-CAM, Mapillary.

For citation: Baskhanov A.R. Geolocation determination from dashcam footage using ResNet. *Modeling, Optimization and Information Technology*. 2026;14(7). (In Russ.). URL: <https://moitvvt.ru/journal/article?id=2399> DOI: 10.26102/2310-6018/2026.58.7.016

Введение

В последние годы методы компьютерного зрения и глубокого обучения активно применяются для анализа изображений и извлечения из них пространственной информации. Одной из интересных задач в этой области является геолокация изображений – определение географических координат фотографии по ее содержанию. Это может быть полезно для картографических сервисов, городского анализа, навигации и автоматической обработки больших коллекций снимков.

Современные нейронные сети способны распознавать сложные визуальные признаки городских сцен, такие как архитектура зданий, дорожная инфраструктура, ландшафтные особенности и другие элементы городской среды. Использование этих признаков позволяет моделям машинного обучения определять местоположение изображения даже при отсутствии прямой географической метаданных.

Ранние работы по геолокации изображений были сосредоточены на глобальном предсказании с использованием созданных вручную признаков и кластеризации визуальных слов. Более современные подходы используют глубокие сверточные нейросети. Например, архитектура PlaNet (Weyand и соавт., 2016) делит Землю на сетку и обучает классификатор предсказывать ячейку, содержащую изображение, достигая точности на уровне континентов [1]. Аналогично, система IM2GPS (Hays и Efros, 2008) использовала поиск ближайшего соседа по большой базе геопривязанных изображений [2]. Однако эти методы предназначены для локализации в масштабах всего мира или страны и часто не способны различить места внутри одного города, где визуальные различия более тонкие.

Несколько исследований были посвящены геолокации на уровне города. (Workman и соавт., 2015) использовали комбинацию уличных и аэрофотоснимков [3]. Более поздние исследования применяют метрическое обучение или сиамские сети для встраивания изображений в пространство, где евклидово расстояние коррелирует с географической близостью [4]. Тем не менее, большинство этих подходов либо требуют больших объемов размеченных данных для конкретного города, либо полагаются на специализированные сенсоры (например, GPS-привязанные панорамы улиц). Кроме того, мало работ систематически сравнивают различные архитектуры ResNet для регрессии координат и предоставляют визуальные объяснения того, чему обучается сеть [5].

Несмотря на прогресс, остаются следующие пробелы: ограниченная оценка километровой точности в пределах одного города с использованием стандартных кадров видеорегистраторов (которые содержат повторяющиеся дорожные виды и менее различимые ориентиры); отсутствие детального анализа того, как глубина сети и разрешение входных изображений влияют на точность локализации в таких сложных условиях; отсутствие визуализации внимания для понимания того, на какие городские элементы модель опирается при геолокации.

Целью данного исследования является изучение возможности определения координат фотографий в пределах одного города с помощью сверточных нейросетей. Для достижения поставленной цели были сформулированы следующие задачи:

1. Собрать и подготовить набор данных уличных геопривязанных изображений Уфы, обеспечив равномерное пространственное покрытие.
2. Обучить регрессионные модели на архитектурах ResNet18 и ResNet50 для предсказания широты и долготы по кадрам размером 512×512 пикселей.
3. Количественно оценить точность предсказаний, используя медианную ошибку и распределение ошибок.
4. Визуализировать области изображений, влияющие на решения модели, с помощью Smooth Grad-CAM, и определить семантически значимые ориентиры [6].

Материалы и методы

Для сбора набора данных уличных фотографий Уфы была выполнена регистрация на платформе Mapillary и получен Access Token, который позволил использовать API Mapillary для доступа к фотографиям и их метаданным. Mapillary – это онлайн-платформа и сервис для хранения и обмена геопривязанными уличными фотографиями, предоставляющий API для работы с изображениями и их метаданными, включая координаты, ориентацию камеры и время съемки. Все фотографии на Mapillary публикуются под лицензией CC-BY-SA¹, которая разрешает свободное использование для обучения моделей при условии указания авторства и сохранения той же лицензии при перераспределении данных.

В отличие от таких сервисов, как Yandex.Maps или Google Maps, Mapillary ориентирована именно на открытый обмен данными, предоставляя изображения и метаданные под лицензией, разрешающей обучение моделей. Напротив, изображения с Яндекс или Google защищены авторским правом и лицензиями, запрещающими их свободное использование для подобных целей.

Была разработана программа для автоматического сбора фотографий, сделанных по всему городу, и сохранения их координат. Поскольку фотографии размещались на CDN Facebook², программа использовала HTTP-прокси для обеспечения надежного доступа к данным с российского IP-адреса.

Географические границы города были определены с помощью GeoJSON-файла, полученного из открытых источников через сервис PaintMaps³. Для сбора фотографий карта города сначала была разделена на тайлы (TILE_Z = 14), которые определяют области, из которых Mapillary возвращает изображения. При таком уровне детализации каждый тайл покрывает примерно 150–300 метров в ширину. Поскольку фотографии собираются с движущихся автомобилей, все снимки привязаны к дорогам. Однако внутри одного тайла разные дороги могут быть охвачены неравномерно: некоторые дороги представлены большим количеством фотографий, тогда как другие дороги в том же тайле могут вообще не иметь ни одного снимка – не потому, что по ним мало ездят, а просто потому, что при случайном отборе фотографий эти дороги не попали в выборку.

Чтобы сгладить этот дисбаланс и гарантировать, что в выборку попадут снимки со всех дорог, поверх тайлов была наложена дополнительная сетка размером 100 на 100

¹ Mapillary. CC-BY-SA license for open data. URL: <https://help.mapillary.com/hc/en-us/articles/115001770409-CC-BY-SA-license-for-open-data> (дата обращения: 03.05.2026).

² Деятельность Meta Platforms Inc. (Facebook) запрещена на территории РФ решением Тверского районного суда г. Москвы от 21.03.2022 (дело № 02-2473/2022). URL: <https://mediapravo.com/wp-content/uploads/Delo-02-2473-2022.-Motivirovannoe-reshenie.-dokument-obezhlichennaya-kopiya.pdf> (дата обращения: 03.05.2026).

³ PaintMaps. Map cropping tool sample (Ufa polygon). URL: <https://tools.paintmaps.com/map-cropping/RU/4-890487017/samples> (дата обращения: 03.05.2026).

ячеек (GRID_X = 100, GRID_Y = 100). Эта сетка делит город на 10 000 равновеликих ячеек ($100 \times 100 = 10\,000$). Алгоритм отбирал фотографии так, чтобы каждая ячейка получила примерно равное количество снимков. Это гарантирует, что даже дороги, случайно пропущенные при случайном сэмплеировании, попадут в итоговую выборку.

Максимальное количество фотографий для извлечения было ограничено $MAX_PHOTOS = 100\,000$. Система формировала выборку таким образом, чтобы фотографии были равномерно распределены по всем ячейкам сетки. Однако некоторые фотографии оказались за пределами полигона (например, на границе), а в некоторых областях полигона фотографии отсутствовали из-за отсутствия дорог или невозможности проезда транспортных средств, с которых велась съемка. В итоге было собрано 88 798 фотографий. Распределение фотографий по карте города показано на Рисунке 1.

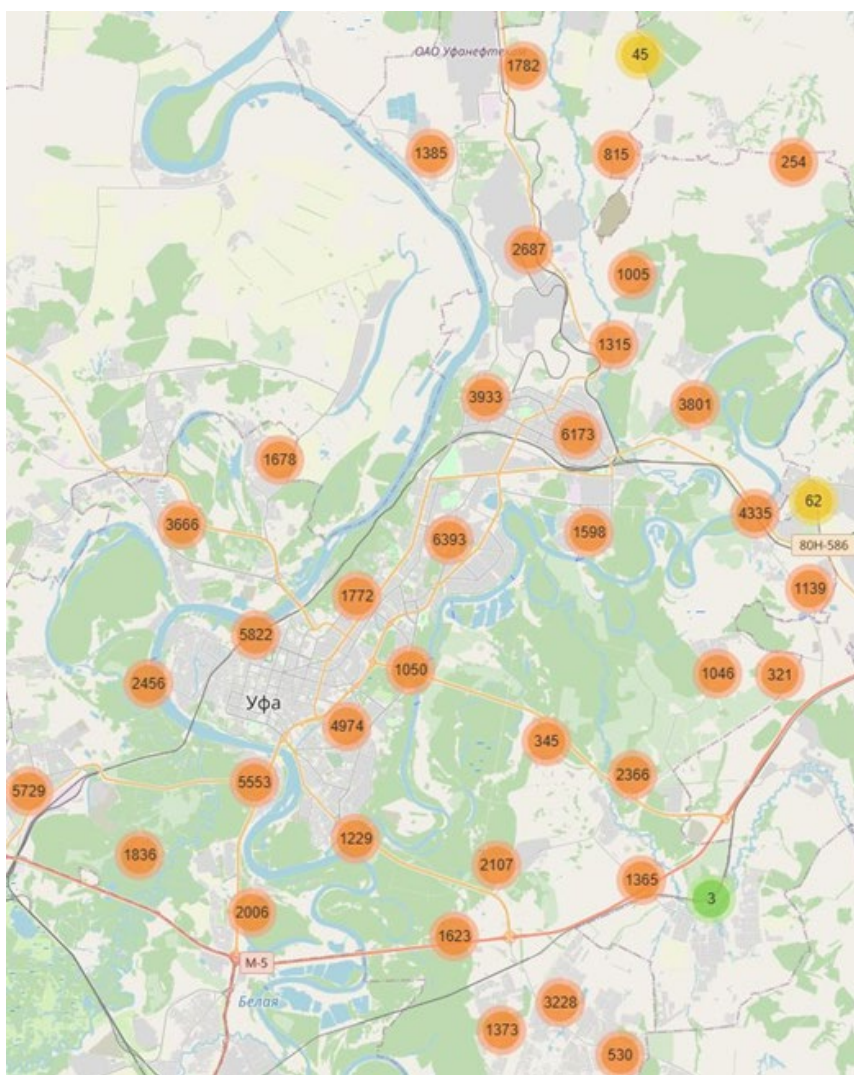


Рисунок 1 – Пространственное распределение фотографий в городе Уфа
Figure 1 – Spatial distribution of photographs in the city of Ufa

Поскольку полигон города Уфы довольно большой, значительная часть фотографий была сделана на автомагистралях и в пригородных зонах, где плотность уличных объектов невысока. Для более детального анализа производительности модели полезно отдельно рассмотреть распределение изображений в центре города (Рисунок 2).



Рисунок 2 – Распределение изображений в центре города Уфы
Figure 2 – Distribution of images in the center of Ufa

Собранный набор данных был разделен на обучающую и тестовую выборки в соотношении 80/20. Для предотвращения утечки данных разбиение выполнялось не случайным образом по отдельным изображениям, а по авторам (участникам платформы Mapillary, выполнившим съемку). Все снимки, загруженные одним автором, целиком попадали либо в обучающую, либо в тестовую выборку.

В процессе обучения модели в качестве входных данных использовались только пиксели изображений. Метаданные Mapillary, такие как ориентация камеры (`compass_angle`) и время съемки (`captured_at`), были намеренно исключены. Идентификатор автора (`creator`) использовался исключительно для разбиения набора данных на обучающую и тестовую выборки и не подавался в модель в качестве признака. Это позволяет оценить способность сверточной сети определять геолокацию только по визуальному содержанию кадра, без использования дополнительных сенсорных или временных подсказок. В датасете присутствуют изображения, сделанные в разное время суток, включая ночные кадры, что повышает разнообразие обучающей выборки.

В качестве основы использовались предобученные на ImageNet сети ResNet18 и ResNet50. Стандартный классификационный слой, предсказывающий 1000 классов, был заменен на два нейрона с линейной активацией для прямой регрессии широты и долготы:

$$(\hat{lat}, \hat{lon}) = W \cdot AvgPool(f_{\theta}(X)) + b, \quad (1)$$

где X – входное изображение, f_{θ} – сверточная основа сети, W и b – обучаемые параметры выходного слоя.

Координаты нормализованы на диапазон $[0, 1]$:

$$lat_{norm} = \frac{lat - lat_{min}}{lat_{max} - lat_{min}}, lon_{norm} = \frac{lon - lon_{min}}{lon_{max} - lon_{min}}. \quad (2)$$

Оптимизация выполнялась минимизацией среднеквадратичной ошибки (MSE):

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N [(\hat{lat}_i - lat_i)^2 + (\hat{lon}_i - lon_i)^2], \quad (3)$$

где N – размер батча.

Результаты

ResNet18 – это относительно легкая сверточная нейронная сеть с 18 слоями, что обеспечивает быстрое обучение, но ограничивает способность извлекать сложные признаки из изображений [7]. ResNet50, в свою очередь, имеет 50 слоев и больше параметров, что позволяет извлекать более детальные признаки объектов, что критично для городских сцен с мелкими ориентирами [8].

Сравнение конфигураций представлено в Таблице 1. Во всех экспериментах размер батча составлял 32, за исключением случая ResNet50 при разрешении 512×512, где из-за ограничений вычислительной мощности размер батча был уменьшен до 16. Это привело к тому, что медианная ошибка для ResNet50 при 512×512 оказалась чуть выше, чем у ResNet18 при том же разрешении. Тем не менее, итоговый выбор был сделан в пользу именно этой конфигурации по нескольким причинам. Во-первых, более глубокая архитектура на начальных этапах обучения может показывать худшие результаты не только из-за меньшего батча, но и потому, что ее потенциал раскрывается на более поздних эпохах. Во-вторых, несмотря на близкую медианную ошибку, loss у ResNet50 существенно ниже.

Это означает, что ResNet50 реже допускает грубые ошибки, что видно, например, в диапазонах 5–10 км и >10 км. Однако из-за ограничений Таблицы 1 детально эти диапазоны не раскрыты. Поясним на примере: для одного и того же изображения ResNet18 может предсказать расстояние 8 км, тогда как ResNet50 – 6 км (при истинном значении, например, 5 км). За счет меньшего отклонения потери (loss) у ResNet50 оказываются ниже, что крайне важно в задаче определения геолокации.

Таблица 1 – Сравнение архитектур ResNet18 и ResNet50 при разных разрешениях (5 эпох)
Table 1 – Comparison of ResNet18 and ResNet50 architectures at different resolutions (5 epochs)

Архитектура	ResNet18		ResNet50	
	256×256	512×512	256×256	512×512
Разрешение				
0–10 m	0,00%	0,00%	0,00%	0,00%
10–50 m	0,00%	0,02%	0,01%	0,02%
50–100 m	0,01%	0,03%	0,00%	0,07%
100–250 m	0,16%	0,44%	0,16%	0,38%
250–500 m	0,61%	1,27%	0,74%	1,13%
500–1000 m	1,81%	4,76%	2,56%	4,13%
1–2 km	7,44%	14,37%	9,66%	13,59%
2–3 km	11,00%	16,16%	12,15%	14,51%
3–5 km	23,55%	25,06%	22,86%	24,32%
5–10 km	39,48%	28,22%	38,11%	29,30%
>10 km	15,96%	9,66%	13,74%	12,56%
Loss	3129 m	4921 m	4699 m	3293m
Median error	5488 m	3912 m	5172 m	4230 m

Таблица 2 – Оценка модели ResNet50 с изображениями 512×512 пикселей
Table 2 – Evaluation of the ResNet50 model with 512×512 pixel images

Epoch	5	6	7	8	9	10	11	12	13
0–10 m	0,00%	0,00%	0,00%	0,01%	0,00%	0,00%	0,00%	0,00%	0,00%
10–50 m	0,02%	0,01%	0,02%	0,01%	0,00%	0,03%	0,03%	0,02%	0,02%

Таблица 2 (продолжение)
Table 2 (continued)

50–100 m	0,07%	0,06%	0,07%	0,07%	0,10%	0,07%	0,08%	0,06%	0,11%
100–250 m	0,38%	0,48%	0,49%	0,42%	0,48%	0,56%	0,38%	0,48%	0,63%
250–500 m	1,13%	1,45%	1,65%	1,49%	1,82%	1,93%	1,91%	1,68%	2,28%
500–1000 m	4,13%	5,38%	5,96%	5,55%	6,32%	7,30%	5,96%	5,95%	8,26%
1–2 km	13,59%	15,65%	17,20%	16,91%	19,06%	18,59%	16,39%	17,15%	19,99%
2–3 km	14,51%	16,57%	16,08%	16,41%	16,75%	17,30%	16,36%	16,32%	17,06%
3–5 km	24,32%	24,45%	23,80%	24,08%	23,05%	22,68%	24,00%	22,64%	21,39%
5–10 km	29,30%	26,54%	26,28%	26,83%	24,56%	23,92%	25,80%	26,25%	22,91%
>10 km	12,56%	9,41%	8,45%	8,23%	7,85%	7,62%	9,09%	9,45%	7,36%
Loss	3293m	3167 m	2838 m	2699 m	2586 m	2469 m	2385 m	2284 m	2179 m
Median error	4230 m	3725 m	3612 m	3654 m	3398 m	3295 m	3606 m	3615 m	3123 m

Как видно из Таблицы 2, по мере обучения модель продолжала улучшаться: с каждой последующей эпохой средняя ошибка уменьшалась, а распределение ошибок смещалось в сторону меньших значений. Например, хотя медианная ошибка между 7-й и 13-й эпохами изменилась не столь значительно (с 3612 м до 3123 м), доля фотографий с ошибками более 10 км снизилась с 8,45 % до 7,36 %, а количество предсказаний с ошибками в диапазоне 500–1000 м увеличилось.

Решение остановить обучение на 13-й эпохе было принято потому, что дальнейшие улучшения замедлялись: медианная ошибка уменьшалась лишь постепенно, в то время как вычислительные затраты продолжали расти.

Вместо кросс-валидации данные были разделены на основе идентификаторов авторов снимков (участников Mapillary). Это исключает попадание изображений одного автора в обе выборки, что обеспечивает более объективную оценку обобщения модели, чем случайное разбиение. Для предотвращения переобучения использовались методы регуляризации: Dropout с вероятностью 0,3 на последнем полносвязном слое и L2-штраф на веса (weight decay, коэффициент 10^{-4}) в оптимизаторе AdamW.

Обсуждение

Для анализа производительности модели мы выбрали конкретную фотографию из набора данных и подали ее на вход обученной модели на базе ResNet50 с модифицированным регрессионным слоем для предсказания координат. Полученные предсказанные координаты затем были сравнены с фактическими координатами, и было рассчитано отклонение в метрах между ними. Кроме того, для визуальной оценки точности модели были показаны четыре фотографии, ближайшие к фактическому местоположению, и фотографии, ближайшие к предсказанному местоположению.

Для анализируемой фотографии была построена карта важности с использованием метода Smooth Grad-CAM, примененного на уровне последнего сверточного блока модели [9]. Для повышения устойчивости визуализации было сгенерировано 30 слегка возмущенных версий фотографии с добавлением гауссовского шума ($\sigma = 0,05$ от диапазона значений пикселей), после чего градиенты были усреднены. Результирующая карта была подвергнута перцентильной нормализации и наложена на исходную фотографию с использованием цветовой схемы Jet и альфа-смешивания, что позволило визуально идентифицировать области сцены, которые наиболее сильно повлияли на предсказания координат (Рисунок 3).

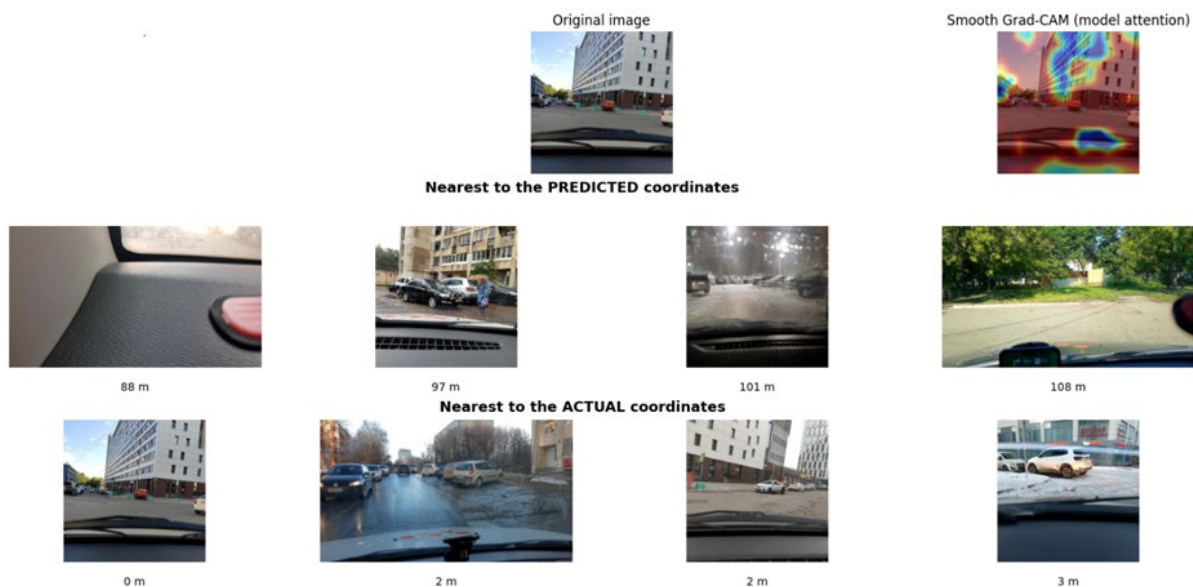


Рисунок 3 – Визуализация внимания модели Smooth Grad-CAM
Figure 3 – Smooth Grad-CAM attention visualization

Для анализа внимания модели были выбраны пять случайных фотографий из центра города, содержащих характерные городские объекты (такие как здания, перекрестки и ограждения). Каждая фотография была изменена до размера 512×512 пикселей и нормализована так же, как и во время обучения. Для визуализации использовался метод Smooth Grad-CAM: для каждого изображения было сгенерировано 30 слегка возмущенных версий с $\sigma = 0,05$, после чего была построена усредненная карта значимости. Для исключения ярких областей неба и фона была применена маска. Результирующая карта значимости была наложена на исходную фотографию с использованием цветовой схемы Jet [10]. Визуализация показала, что модель фокусируется на ориентирах, значимых для географической локализации (Рисунок 4).

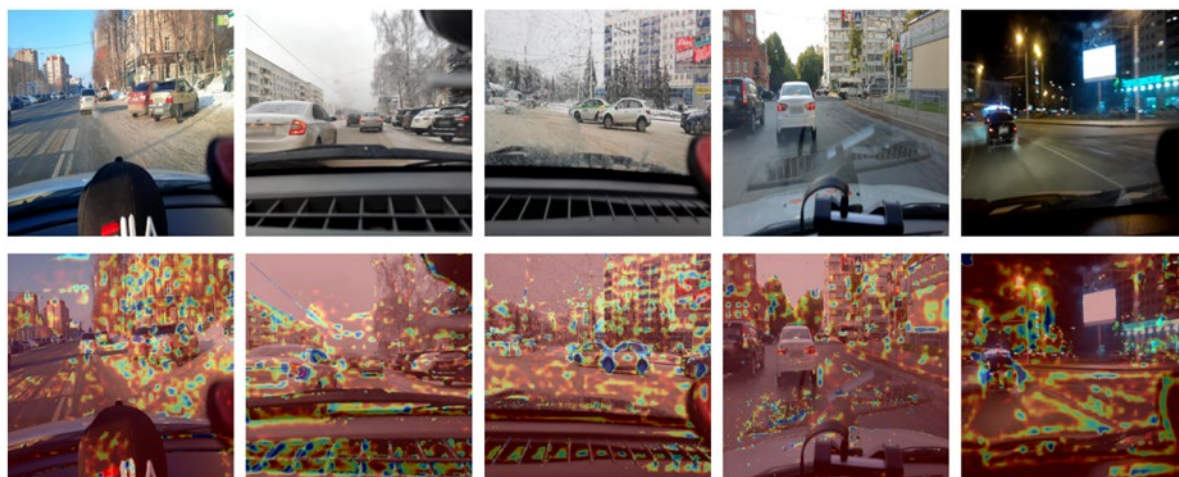


Рисунок 4 – Внимание модели на пяти снимках центра города
Figure 4 – Model attention on five city-center images

Аугментация данных не использовалась. Это позволило оценить способность модели обобщать признаки на исходных, неискаженных уличных кадрах, что соответствует условиям реальной эксплуатации видеорегистраторов. Динамические

объекты (автомобили, пешеходы) не удалялись из изображений, поскольку задача заключалась в работе с реальными данными. Предполагается, что их влияние на предсказание нивелируется за счет усреднения признаков по множеству статичных ориентиров (здания, дорожная разметка, перекрестки), на которых модель фокусируется согласно визуализациям Grad-CAM.

Для количественной оценки эффективности предложенного подхода было выполнено сравнение с ключевыми работами в области геолокации изображений (Таблица 3). Важно отметить, что прямое сравнение абсолютных значений ошибок затруднено из-за различий в датасетах (глобальный масштаб и один город) и условиях съемки (панорамы и стандартные кадры видеорегистраторов). Тем не менее, предложенная модель демонстрирует высокую конкурентоспособность, достигая километровой точности внутри одного города, в то время как глобальные методы обеспечивают лишь континентальную точность.

Таблица 3 – Сравнение с ключевыми работами по геолокации изображений

Table 3 – Comparison with key works on image geolocation

Работа	Датасет	Архитектура	Разрешение	Медианная ошибка
Предложенный подход	Уфа (Mapillary)	ResNet50	512×512	3,1 км
PlaNet (Weyand et al., 2016) [1]	Global (YFCC100M)	GoogLeNet	224×224	1131,7 км
IM2GPS (Hays & Efros, 2008) [2]	Global (6M)	KNN + SIFT	32×32 (GIST) / 1024px (база)	>1000 км (16 % фото определено в радиусе 200 км)
Müller-Budack et al., 2018 [11]	Global (Cross-View)	Hierarchical	224×224	4,2 км
Workman et al., 2015 [3]	City-scale (Panoramas)	CNN	256×256	0,5–1 км

Достигнутая медианная ошибка в 3,1 км сопоставима с результатами на уровне города, о которых сообщают Müller-Budack и соавторы [11], получившие 4,2 км на похожем наборе данных с использованием подхода, основанного на классификации. В отличие от (Workman и соавт. 2015), которые полагались на панорамы улиц, наш метод работает со стандартными кадрами видеорегистраторов. По сравнению с методами глобального масштаба, такими как PlaNet, достигающими точности на уровне континентов, наша модель специально адаптирована для внутригородской локализации, где визуальная вариативность ниже, а задача является более сложной.

Заключение

В данной работе была рассмотрена проблема определения географических координат городских фотографий с использованием методов глубокого обучения. Для проведения экспериментов на основе данных платформы Mapillary был создан набор уличных фотографий Уфы, содержащий десятки тысяч снимков с координатами съемки.

В ходе экспериментов были обучены модели на архитектурах ResNet18 и ResNet50, предназначенные для регрессионного предсказания координат по фотографиям. Результаты показали, что увеличение глубины сети и разрешения входных изображений повышает качество предсказаний и снижает медианную ошибку координат. Конфигурация ResNet50 с разрешением изображений 512×512 пикселей

оказалась наиболее эффективной, достигнув километровой точности предсказания в пределах города.

Важно отметить, что достигнутая километровая точность в пределах одного города является заслуживающей внимания. Многие фотографии в наборе данных были сделаны камерами, установленными на движущихся транспортных средствах, поэтому изображения содержат в основном виды дороги и прилегающей инфраструктуры в направлении движения. В этих условиях значительная часть визуальной информации на каждой фотографии может быть схожей в разных районах города, поскольку дорожная среда, включая проезжие части, разметку, ограждения и типовые здания, часто обладает повторяющимися признаками.

Кроме того, ракурс съемки ограничен ориентацией дороги, что снижает разнообразие перспектив и усложняет задачу определения точного местоположения. Учитывая эти естественные ограничения, достигнутая точность предсказания координат в несколько километров может считаться удовлетворительной, демонстрируя способность модели извлекать географически значимые признаки из городских фотографий.

Дальнейший анализ с использованием методов визуализации внимания (Smooth Grad-CAM) показал, что модель фокусируется на характерных элементах городской среды, таких как здания, перекрестки, дорожная инфраструктура и другие объекты, которые могут служить географическими ориентирами.

Эти результаты демонстрируют, что методы глубокого обучения способны извлекать географически значимую информацию из городских фотографий. Потенциальные направления будущих исследований включают увеличение размера набора данных, использование более сложных архитектур нейронных сетей и расширение задачи для определения координат в нескольких городах или регионах.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Weyand T., Kostrikov I., Philbin J. PlaNet – Photo Geolocation with Convolutional Neural Networks. In: *Proceedings of the Computer Vision – ECCV 2016: 14th European Conference, 11–14 October 2016, Amsterdam, Netherlands*. Cham: Springer; 2016. P. 37–55. https://doi.org/10.1007/978-3-319-46484-8_3
2. Hays J., Efros A.A. IM2GPS: estimating geographic information from a single image. In: *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, 23–28 June 2008, Anchorage, AK, USA*. IEEE; 2008. P. 1–8. <https://doi.org/10.1109/CVPR.2008.4587784>
3. Workman S., Jacobs N. On the location dependence of convolutional neural network features. In: *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 7–12 June 2015, Boston, MA, USA*. IEEE; 2015. P. 70–78. <https://doi.org/10.1109/CVPRW.2015.7301385>
4. Melekhov I., Kannala J., Rahtu E. Siamese network features for image matching. In: *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), 4–8 December 2016, Cancún, Mexico*. IEEE; 2016. P. 378–383. <https://doi.org/10.1109/ICPR.2016.7899663>
5. Hou Y., Quintana M., Khomiakov M., et al. Global Streetscapes – a comprehensive dataset of 10 million street-level images across 688 cities for urban science and analytics. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2024;215:216–238. <https://doi.org/10.1016/j.isprsjprs.2024.06.023>
6. Chattopadhyay A., Sarkar A., Howlader P., et al. Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks. In: *Proceedings of the*

- 2018 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 12–15 March 2018, Lake Tahoe, NV, USA. IEEE; 2018. P. 839–847. <https://doi.org/10.1109/WACV.2018.00097>
7. He K., Zhang X., Ren S., et al. Deep Residual Learning for Image Recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 27–30 June 2016, Las Vegas, NV, USA. IEEE; 2016. P. 770–778. <https://doi.org/10.1109/CVPR.2016.90>
 8. Wu Z., Shen C., Van Den Hengel A. Wider or Deeper: Revisiting the ResNet Model for Visual Recognition. *Pattern Recognition*. 2019;90:119–133. <https://doi.org/10.1016/j.patcog.2019.01.006>
 9. Selvaraju R.R., Cogswell M., Das A., et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*. 2020;128(2):336–359. <https://doi.org/10.1007/s11263-019-01228-7>
 10. Shimizu T., Nagata F., Arima K., et al. Enhancing defective region visualization in industrial products using Grad-CAM and random masking data augmentation. *Artificial Life and Robotics*. 2024;29:62–69. <https://doi.org/10.1007/s10015-023-00913-8>
 11. Müller-Budack E., Pustu-Iren K., Ewerth R. Geolocation Estimation of Photos Using a Hierarchical Model and Scene Classification. In: *Proceedings of the 15th European Conference, 8–14 September 2018, Munich, Germany*. Cham: Springer; 2018. P. 575–592. https://doi.org/10.1007/978-3-030-01258-8_35

ИНФОРМАЦИЯ ОБ АВТОРЕ / INFORMATION ABOUT THE AUTHOR

Басханов Артур Русланович, магистрант кафедры «Финансовые технологии», Финансовый университет при Правительстве Российской Федерации, Москва, Российская Федерация.

e-mail: bashanov.artur@yandex.ru

ORCID: [0009-0007-8584-6484](https://orcid.org/0009-0007-8584-6484)

Artur R. Baskhanov, Master's Degree student at the Department of Financial Technologies, Financial University under the Government of the Russian Federation, Moscow, the Russian Federation.

Статья поступила в редакцию 03.05.2026; одобрена после рецензирования 26.06.2026; принята к публикации 17.07.2026.

The article was submitted 03.05.2026; approved after reviewing 26.06.2026; accepted for publication 17.07.2026.