УДК 004.93

МЕТОД ВИЗУАЛЬНОГО РАСПОЗНАВАНИЯ МЕСТНОСТИ NetVLAD ДЛЯ ЛОКАЛИЗАЦИИ ЛОКОМОТИВА

МАЩЕНКО
Павел Евгеньевич,
ООО «ЛокоТех-Сигнал», заместитель генерального директора, Россия, Москва

ШИРЯЕВ
Павел Павлович,
ООО «ЛокоТех-Сигнал», специалист по компьютерному зрению, Россия, Москва

Ключевые слова: *сметод визуального распознавания местности, локализация, нейронные сети, глубокое обучение, железная дорога, NetVLAD, ВЛАД*

Аннотация. **В статье приводится краткий обзор методов визуального распознавания местности (ВРМ) – эффективного решения по локализации локомотивов. Раскрываются особенности применения таких методов в железнодорожной отрасли. Описывается структура построения систем с алгоритмом ВРМ. Рассмотрен принцип работы нейронной сети NetVLAD [1], как быстрого, точного и робастного метода по идентификации текущей локации. Демонстрируется эксперимент с реальными данными, полученными системой машинного зрения гибридного маневрового локомотива ТЭМ5Х. Приводятся результаты эксперимента и рекомендации по реализации системы ВРМ с NetVLAD.**

Методы визуального распознавания местности относятся к алгоритмам компьютерного зрения. Они отыскивают совпадения текущего изображения местности с этой же самой местностью, но снятой в другое время. Данные алгоритмы применяются в робототехнике, беспилотном транспорте, геолокации, инструментах дополненной реальности и мобильных приложениях [2].

Схема работы системы ВРМ представлена на рис. 1 [3]. Ее основные компоненты: модуль обработки изображений, алгоритм ВРМ и карта локаций (КЛ). На модуль обработки изображений с камеры поступает кадр, который затем проходит трансформацию: нормализацию, масштабирование, цветовую конвертацию и др. Обработка изображений подготавливает данные для алгоритма ВРМ.

Карта локаций формируется заранее и представляет собой базу данных изображений местности, с которыми будут сравниваться новые поступающие в систему изображения. Кроме того, за каждым изображением в КЛ может быть закреплена метаинформация: географические координаты, название локации и др.

Алгоритм ВРМ сравнивает текущее изображение с камеры с изображениями из КЛ и предсказывает точность их совпадений. Затем выбирается изображение с самым высоким показателем совпадения, превышающим заранее установленное пороговое значение. Таким образом, система ВРМ по изображению с объектива камеры позволяет получать информацию о локации.

Отметим, что эффективность работы системы можно повышать за счет использования позиционных данных и обновления карты. Позиционными данными могут быть географические координаты с GPS-датчика, позволяющие вводить дополнительную проверку предсказания алгоритма ВРМ и улучшать точность системы. Обновление КЛ новыми изображениями, поступающими с камеры, значительно повышают робастность метода ВРМ.

Алгоритм определения совпадения изображений чувствителен к визуальной составляющей кадра. Погодные условия, время года и суток, освещение, окклюзии и прочие составляющие могут ухудшать качество изображения местности, затрудняя работу системы ВРМ [4]. Поэтому формирование КЛ множеством вариативных изображений для каждой локации, является одним из ключевых факторов успешной работы системы.

В железнодорожной отрасли методы ВРМ применимы на локомотивах, где требуется его локализация. Однако их применение ограничено нестабильной GPS-связью на некоторых участках железных дорог.

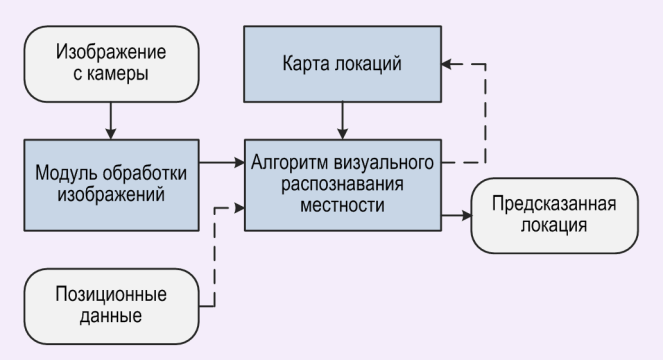

РИС. 1

РИС. 2

Таким примером могут выступать маневровые локомотивы на предприятиях и заводах, где различные металлоконструкции вдоль железнодорожных путей приводят к экранированию и ослаблению сигнала.

Кроме того, данная система способна определять участки пути, находящиеся на расстоянии от объектов, подъезд к которым требует от машиниста повышенного внимания. Примером таких объектов может быть въезд в цех, ворота депо, платформа, стрелочный перевод и др. В случае движения маневрового локомотива по криволинейным участкам пути, данные объекты могут быть скрыты из поля видимости машиниста. Система ВРМ позволяет заранее учитывать такие ситуации и определять по изображению с камеры приближение к опасным объектам, сигнализировать об этом машинисту, тем самым повышая шанс предотвращения аварийной ситуации.

Ключевая составляющая работы системы ВРМ – алгоритм сопоставления текущего кадра с изображениями из КЛ. Для его выполнения применяются методы поиска особых точек, а точнее дескрипторов (описаний) этих точек, которые являются общими для обоих изображений. Особые точки изображения – это пиксели, по которым можно выделить уникальные черты данного изображения. Как правило, такими элементами выступают угловые точки, где присутствует резкая смена цвета, яркости (угол здания на фоне неба и др.).

Далее по сформированным дескрипторам для обоих изображений производится поиск пар. При количестве найденных пар выше установленного заранее порогового значения принимается решение о принадлежности этих изображений одной и той же локации (рис. 2).

Существует множество методов поиска особых точек [3, 5, 6]. Каждый из них обладает некоторыми преимуществами по скорости работы, точности, требуемой вычислительной мощности и др. Рассмотрим метод NetVLAD, который является робастным алгоритмом по извлечению дескрипторов изображений местности, изменяющейся во времени [6].

В основе работы NetVLAD лежит сверточная нейронная сеть. В классических методах поиска особых точек принцип их выбора заранее предопределен алгоритмом работы и назначенными коэффициентами. Модель нейронной сети NetVLAD обучается находить именно те дескрипторы, которые наиболее точно и однозначно описывают конкретную локацию. Также на вход нейронной сети могут поступать изображения одной и той же локации, но в разные временные моменты. Такая вариативность позволяет модели научиться точно определять местность при разных погодных и внешних условиях, повышая надежность работы всей системы ВРМ. Архитектура нейронной сети NetVLAD состоит из двух частей: сверточных слоев и слоя NetVLAD (рис. 3).

Задача сверточных слоев – выработка необходимой последовательности карт признаков, где маловажные детали отфильтрованы, а существенные выделены. Сверточными слоями нейронной сети могут выступать слои из архитектур классификационных нейронных сетей, которые позволяют осуществлять гибкую настройку модели NetVLAD. Существует множество архитектур классификационных сетей, отличающихся по своей глубине, количеству параметров и другим показателям [7]. Разработчик имеет возможность самостоятельно выбрать сверточные слои в архитектуре NetVLAD, генерация признаков которых удовлетворит его требования по точности и быстродействию работы нейронной сети. Выходными данными последнего сверточного слоя является карта признаков размерностью HxWxD, которую можно рассматривать как набор D-мерных дескрипторов, извлеченных в пространстве HxW.

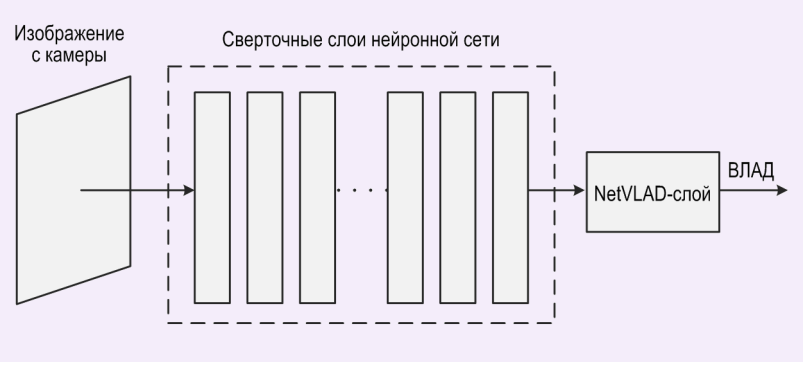

РИС. 3

NetVLAD-слой отвечает за объединение извлеченных дескрипторов в фиксированное представление изображения, а именно в вектор локально агрегированных дескрипторов (ВЛАД) [8]. Данная кодировка позволяет собирать статистическую информацию о локально агрегированных дескрипторах по всему изображению. ВЛАД хранит сумму остатков (вектор разницы между дескриптором и его соответствующим центром кластера) для каждого визуального слова (рис. 4) [1]. Концепция визуальных слов адаптирована из информационного поиска, где подсчитывается присутствие каждого слова, встречающегося в документе. В

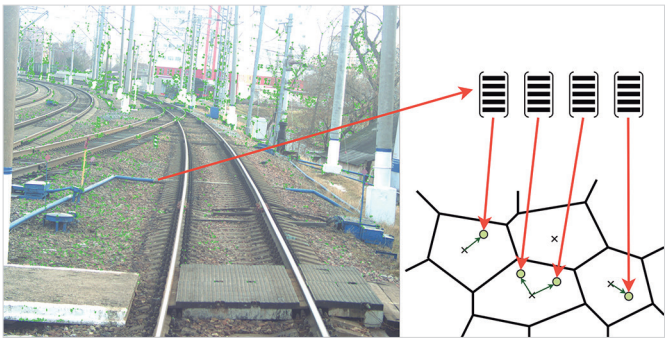

РИС. 4

случае изображения, визуальными словами являются наборы ключевых точек и дескрипторов, при помощи которых формируется представление изображения в виде частотной гистограммы объектов (рис. 5) [9]. Результат работы NetVLAD – получение частотной гистограммы наборов сумм остатков визуальных слов, т. е. ВЛАД. Далее осуществляется поиск совпадений по карте локаций текущего вектора с векторами других изображений.

Эксперимент по оценке эффективности работы ВРМ-системы проводился на данных, собранных гибридным маневровым локомотивом ТЭМ5Х (рис. 6), на экспериментальной кольцевой железной дороге ВНИИЖТ. Локомотив оснащен системой машинного зрения Ctrl@Vision, в состав которой входит набор камер. Данные собирались с января по май текущего года в ходе движения тепловоза по кольцевому пути.

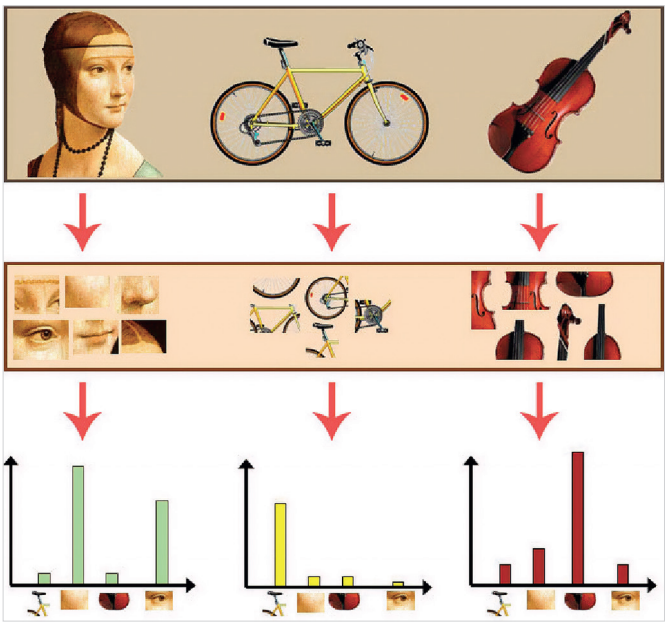

РИС. 5

РИС. 6

Такой временной промежуток был выбран специально, чтобы локации с испытательного полигона имели сильные отличия в своем внешнем виде: отсутствие или наличие снега и др.

Для эксперимента было отобрано по 600 фотографий зимнего и весеннего периода. Расстояние, которое локомотив проезжал между соседними кадрами, составило от 1 до 25 м. Из каждого периода было выбрано по 13 одинаковых локаций, которым соответствовало по пять фотографий, снятых приблизительно с одного и того же ракурса в радиусе 5–7 м. Таким образом, из каждого набора 65 фотографий имели метку от 1 до 13 класса, остальные 1070 – относились к классу 0, то есть к не отслеживаемой местности.

Для усложнения задачи было принято решение об обучении нейронной сети только на зимнем периоде с целью предсказания локализации для весенних изображений. Такое условие воспроизводит часто возникающую перед разработчиком реальную проблему, когда нет возможности иметь большое временное окно для сбора данных при выполнении проекта. Данные были разбиты по классам на выборки для обучения (80 %), валидации (20 %) и тестирования (20 %).

Нехватку данных компенсировали за счет создания дополнительных обучающих данных из уже имеющихся. В ходе обучения нейронной сети применялось выполнение аугментаций «на лету». В список аугментаций вошли: случайная яркость, случайная контрастность, размытие, различные виды шума и др.

В качестве сверточных слоев для NetVLAD была использована архитектура ResNet-18 [10]. В ходе экспериментов с разными вариантами сверточных слоев, было определено, что такая архитектура достаточна для эффективной работы NetVLAD. Небольшое количество параметров ResNet-18 в сравнение с более глубокими собратьями позволила добиться высокого быстродействия. В качестве функции потерь был выбран Triplet loss [11], широко применяемый в задачах, где требуется уменьшить расстояния между объектами одного класса и увеличить между объектами разных классов. Данное решение отлично подходит для изображений, получаемых на железных дорогах, так как разные участки пути могут быть схожи визуально при движении в открытых пространствах или в лесной местности и др. В роли алгоритма минимизации функции потерь выступал Rectified Adam [12] со скоростью обучения 0.001. Нейронная сеть обучалась на протяжении 17 эпох.

После формирования нейронной сетью ВЛАД, сравнение полученного вектора с векторами из КЛ происходит посредством библиотеки Faiss [13] – набора инструментов для эффективного поиска сходства векторов и их кластеризации. В качестве метода поиска был выбран метод на основе нахождения эвклидового расстояния между векторами.

Качество работы модели на тестовой выборке оценивалось при помощи метрики recall. Ее значение составило 1, что говорит о способности алгоритма правильно предсказать классы для всех изображений из тестовой выборки. Стоит отметить, что такой высокий результат соответствует именно данной выборке. При других изображениях, большем объеме локаций и другой выборке результат может быть ниже. Примеры локаций, совпадение которых найдено алгоритмом NetVLAD, представлено на рис. 7.

Скорость работы модели, конвертированной в формат TensorRT [14], на компьютере с видеокартой

РИС. 7

GeForce RTX 2080 Ti составила 201 кадр в секунду при разрешении входного изображения 672x736 пикселей.

■ В статье кратко проведен обзор текущего положения по применению методов визуального распознавания местности. Освещена работа системы ВРМ с алгоритмом NetVLAD. Приведены основные этапы работы данной модели. Продемонстрирован эксперимент с реальными данными, полученными от системы машинного зрения гибридного маневрового локомотива ТЭМ5Х.

Благодаря высоким показателям быстродействия и точности работы на тестовой выборке, система ВРМ с алгоритмом NetVLAD может считаться эффективным решением задачи определения локализации локомотива в реальном времени. Применение обучаемой нейронной сети позволяет обеспечить более индивидуальный подход к используемым данным, тем самым повышая робастность всей системы.

Однако стоит учитывать, что переобучение модели критически влияет на точность ее работы. При реализации NetVLAD требуется грамотный подход по формированию выборок для обучения, валидации, тестирования и правильный подбор настраиваемых параметров.

ЛИТЕРАТУРА

1. NetVLAD : CNN architecture for weakly supervised place recognition / R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV. 2016. P. 5297-5307. doi: 10.1109/CVPR.2016.572.

2. Place recognition: An overview of vision perspective / Z. Zeng, J. Zheng, X. Wang, Y. Chen, Ch. Zhu // Applied Sciences. 2018. T. 8, №. 11. C. 2257. https://doi.org/10.3390/app8112257.

3. Visual place recognition: A survey / S. Lowry, N. Sünderhauf, P. Newman, J.J. Leonard, D. Cox, P. Corke, M.J. Milford // IEEE Transactions on Robotics. 2015. T. 32, №. 1. P. 1–19. doi: 10.1109/TRO.2015.2496823.

4. Naseer T., Burgard W., Stachniss C. Robust visual localization across seasons // IEEE Transactions on Robotics. 2018. T. 34, №. 2. P. 289–302. doi: 10.1109/TRO.2017.2788045.

5. Karami E., Prasad S., Shehata M. Image matching using SIFT, SURF, BRIEF and ORB: performance comparison for distorted images // arXiv:1710.02726v2. 2017. 7 Oct. 5 p.

6. Tareen S. A. K., Saleem Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK // 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur. 2018. P. 1-10. doi: 10.1109/ICOMET.2018.8346440.

7. A survey of the recent architectures of deep convolutional neural networks / A. Khan, A. Sohail, U. Zahoora, A.S. Qureshi // Artificial Intelligence Review. 2020. Vol. 53. DOI: https://doi.org/10.1007/s10462-020-09825-6.

8. Aggregating local descriptors into a compact image representation / H. Jégou, M. Douze, C. Schmid, P. Pérez // 2010 IEEE computer society conference on computer vision and pattern recognition , San Francisco, CA. 2010. P. 3304-3311. doi: 10.1109/CVPR.2010.5540039.

9. Fei-Fei L. Recognizing and learning object categories : CVPR Short Course. 2007. URL: https://courses.cs.washington.edu/courses/cse576/08sp/lectures/CVPR2007_tutorial-Abridged.pdf.

10. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, Sun // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV. 2016. P. 770-778. doi: 10.1109/CVPR.2016.90.

11. Hermans A., Beyer L., Leibe B. In defense of the triplet loss for person re-identification // arXiv:1703.07737v2. 2017. 21 Nov.

12. On the variance of the adaptive learning rate and beyond / L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, J. Han // arXiv:1908.03265v4. 2020. 17 Apr.

13. Johnson J., Douze M., Jégou H. Billion-scale similarity search with GPUs // arXiv:1702.08734v1. 2017. 28 Feb.

14. NVIDIA TensorRT : release : сайт. URL: https://developer.nvidia.com/tensorrt (дата обращения: 24.08.2020).