

УДК 004.89

РАЗРАБОТКА АЛГОРИТМОВ НАКОПЛЕНИЯ ДАННЫХ ПОСРЕДСТВОМ СТЕРЕОПАРЫ И ДЕТЕКТИРОВАНИЯ ДЕФЕКТОВ ДОРОЖНОГО ПОЛОТНА

Полянцева К.А.

*Ордена Трудового Красного Знамени федеральное государственное бюджетное образовательное учреждение высшего образования «Московский технический университет связи и информатики»,
Москва, e-mail: k.a.poliantseva@mtuci.ru*

В данной статье представлена разработка алгоритма накопления данных посредством установки со стереопарой, состоящей из двух камер GoPro Hero7 Black с синхронизированным видеопотоком, микрокомпьютером Jetson Nano для захвата видеопотока и специально сконструированного крепления для того, чтобы получившуюся систему стереозрения можно было закрепить на автомобиле. Поскольку для решения поставленной задачи обнаружения будет выбрана обучаемая модель, ей необходимы примеры для обучения, однако готовых наборов данных, которые бы полностью удовлетворяли условиям задачи, не существует, опытным путем было собрано около 20 000 изображений, вследствие применения методов аугментации к начальному набору данных используемый датасет увеличился до 57 137 изображений. Также описывается разрабатываемый алгоритм детектирования дефектов дорожного полотна Prophetam-DD, основанный на архитектуре модели MC-CNN (так как среди изученных ранее архитектур MC-CNN, SGM-Nets, GC-Net, SegStereo, MC-CNN показывает наилучшую точность) и модернизированный посредством добавления слоев расширенной свертки вместо традиционной операции свертки для того, чтобы увеличить поле восприятия. По результатам обучения на 33 эпохах Prophetam-DD показал наибольшую производительность и точность – 96.78%.

Ключевые слова: машинное обучение, распознавание образов, нейронные сети, стереопара, детектирование дефектов

DEVELOPMENT OF ALGORITHMS FOR DATA ACCUMULATION BY MEANS OF STEREO PAIRS AND DETECTION OF ROADWAY DEFECTS

Polyantseva K.A.

*Order of the Red Banner of Labor Federal State Budgetary Educational Institution
of Higher Education «Moscow Technical University of Communications and Informatics»,
Moscow, e-mail: k.a.poliantseva@mtuci.ru*

This article presents the development of an algorithm for data accumulation by means of an installation with a stereo pair consisting of two GoPro Hero7 Black cameras with a synchronized video stream, a Jetson Nano microcomputer for capturing the video stream and a specially designed mount so that the resulting stereo vision system can be fixed on the car. Since a trainable model will be selected to solve the detection task, it needs examples for training, however, there are no ready-made data sets that would fully satisfy the conditions of the task, about 20,000 images were collected experimentally, as a result of the application of augmentation methods to the initial data set, the dataset used increased to 57137 images. The developed algorithm for detecting defects of the roadway Prophetam-DD is also described, based on the architecture of the MC-CNN model (since among the previously studied architectures MC-CNN, SGM-Nets, GC-Net, SegStereo, MC-CNN shows the best accuracy) and upgraded by adding layers of extended convolution instead of the traditional convolution operation in order to increase the field of perception. According to the results of training on 33 epochs, Prophetam-DD showed the highest performance and accuracy – 96.78%.

Keywords: Machine learning, pattern recognition, neural networks, stereo pairs, defect detection

Целью исследования является создание алгоритмического обеспечения решения задачи обработки изображений в задаче дефектовки дорожного полотна для работы в режиме реального времени на устройствах с малой вычислительной мощностью.

Синхронизация камер – это процесс, при котором 2 или более камер создают и передают разные потоки кадров в один момент времени для получения кадров стереоизображения. Идеальная синхронизация камер достигается в тот момент, когда разница по времени между совершением кадров равняется 0 с. На практике такую синхронизацию получить практически не-

возможно, поэтому задача синхронизации камер заключается в уменьшении времени между производством двух кадров в одной системе [1].

Процесс получения стереоизображения напрямую зависит от синхронизации двух камер, используемых в модуле стереопары. Так как система работает при динамическом движении автомобиля, то очень важно добиться минимального времени между кадрами, чтобы получить как можно большее качество стереоизображения.

В данном случае используются 2 камеры GoPro Hero7 Black. Видеопотоки с этих камер захватываются микрокомпьютером

Jetson Nano посредством карт видеозахвата Espada EscapViHU. Каждая камера подключена к двум независимым USB 3.0-портам, что дополнительно позволяет ускорить работу модуля.

Камеры не имеют внешних триггеров для синхронизации потока кадров. Поэтому было разработано 2 алгоритма (рис. 1) для синхронизации камер, а соответственно и получения кадров с минимальной разницей во времени.

При использовании первого алгоритма кадр с первой камеры захватывается и сразу же открывается, а затем захватывается кадр со второй камеры и открывается.



Рис. 1. Блок-схема алгоритма получения кадров: а) 1 способ; б) 2 способ

При использовании первого алгоритма (кадр с камеры 1 захватывается и сразу же открывается, а затем захватывается кадр с камеры 2 и открывается) было выяснено, что разница между кадрами при 30 кадрах в секунду составляет примерно 32 миллисекунды, а при 60 кадрах в секунду составляет примерно 17 миллисекунд. При использовании второго алгоритма (камеры сначала фиксируют кадры, а после фиксации извлекаются сразу с двух камер) было выяснено, что временной интервал между кадрами

при 30 кадрах в секунду составил 11 миллисекунд, а при 60 кадрах в секунду составил 8 миллисекунд. Это существенно меньше, чем при работе первого алгоритма.

Можно сделать вывод, что задача синхронизации камер заключается в уменьшении времени между производством двух кадров в одной системе. В качестве метода синхронизации камеры был выбран второй алгоритм, так как при его использовании замечен меньший временной интервал, в сравнении с первым алгоритмом.

Методика создания системы для сбора данных заключалась в использовании двух видеокамер и специально сконструированного крепления для того, чтобы получившуюся систему стереозрения можно было закрепить на автомобиле.

В качестве видеокамер использовались две камеры GoPro 7 HERO в режиме 1080p 120 кадров в секунду. Для крепления стереопары и обрабатывающего компьютера к автомобилю была создана специальная конструкция. Для крепления камер и микрокомпьютера использовались изделия, напечатанные из нейлона на 3D-принтере для облегчения веса конструкции. Также на 3D-принтере был напечатан герметичный корпус для Jetson Nano. Данная система предполагает установку на любой автомобиль (не требуется предварительная подготовка или какое-либо дополнительное оборудование).

Основной задачей при проектировании прототипа стереокамеры является достижение стереоэффекта. Стереоэффект основан на особенностях бинокулярного зрения человека и эффекте параллакса.

Наиболее часто применяется способ достижения стереоэффекта, при котором сенсор, регистрирующий изображение, смещается параллельно на расстояние, называемое базой стереосъемки (стереобазой). Для достижения качественного стереоэффекта изображения с правой и левой камер должны быть выравнены относительно друг друга в вертикальной и горизонтальной плоскости.

В зависимости от величины стереобазы изменяется погрешность расчета карты глубины изображения. Чем больше величина стереобазы, тем выше точность расчета расстояния до наблюдаемого объекта.

Исходя из расчета, для достижения приемлемой точности измерений расстояния до наблюдаемого объекта, находящегося в пределах одного метра от стереокамеры, минимально необходимо использовать базу стереосъемки, равную 15,38 см.

Поскольку для решения поставленных целей и задач обнаружения будет выбрана

обучаемая модель, ей необходимы примеры для обучения. Обычно для популярных задач машинного обучения имеются общедоступные коллекции примеров для обучения и исследований. Однако по причине относительного новшества решаемой в данной работе задачи, и необходимости тестирования разрабатываемой системы на реальных данных, было необходимо сформировать датасет.

Обучение нейронной сети для решения поставленных задач осуществлялось на собранном посредством установки набора данных (рис. 2).

В результате было собрано 20 000 ректифицированных стереоизображений (рис. 3), на которых отображены дефекты дорожного покрытия. Стереоизображения были разделены на 4 класса (по типам детектируемых дефектов) – трещины, выбоины, проломы, другое. На их основании были построены карты диспаратностей, которые затем были размечены на предмет выбоин при помощи утилиты LabelME [2; 3]. Наконец, результирующий набор данных был поделен на тренировочный и оценочный в соотношении 80% к 20%.

Создание наборов данных не ограничивается сбором реальных изображений,

другой подход к созданию обучающих данных – синтез искусственных изображений. Как правило, для этого используется аугментация. Процесс аугментации заключается в применении различных преобразований со случайными коэффициентами, например увеличение яркости или контраста изображения, поворот изображения по вертикали или по горизонтали и т.д. Это значительно расширяет тренировочный набор данных, что дает возможность нейронной сети лучше обобщать предсказания за счет большей устойчивости по отношению к изменению некоторых параметров входных данных.

В данной работе при обучении модели, основанной на архитектуре MC-CNN, применялись следующие случайные преобразования: горизонтальное масштабирование, поворот, изменение яркости и контраста, «наложение» погодных условий (снег, дождь), «наложение» дефектов камеры (блики, запотевания). Применялись они в основном при помощи механизма аффинных преобразований. Также каждое входное изображение стереопар было стандартизировано при помощи вычитания из каждого пикселя среднего значения всех пикселей и вычитания значения стандартного отклонения.

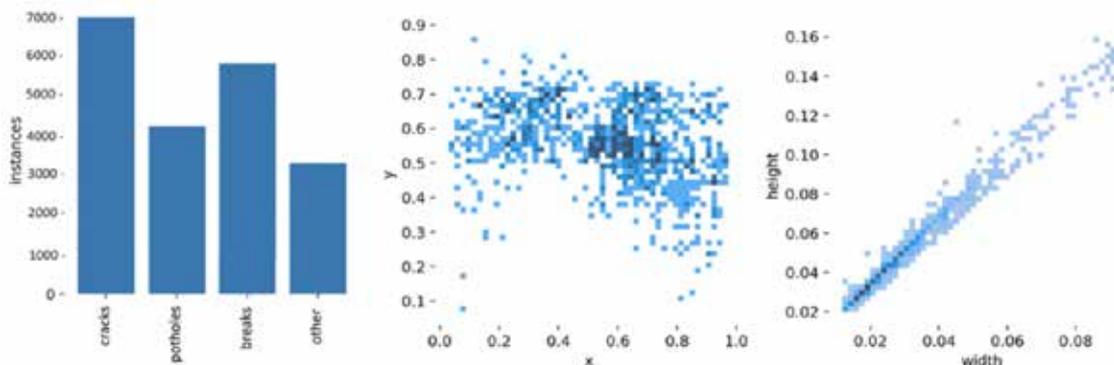


Рис. 2. Классы и количество элементов, содержащиеся в наборе данных



Рис. 3. Примеры собранных изображений

Увеличение набора данных путем многократного преобразования обучающих примеров является широко используемым методом для уменьшения ошибки. Преобразования применяются во время обучения и не влияют на производительность среды выполнения.

Вследствие применения методов аугментации к начальному набору данных, датасет увеличился до 57 137 изображений.

Параметры преобразования данных выбираются случайным образом для каждой пары патчей, и после одной эпохи обучения, когда один и тот же пример представляется сети во второй раз, выбираются новые случайные параметры. Выбираются разные параметры преобразования для левого и правого изображения: например, левый участок поворачивается на 10 градусов, а правый на 14 [4].

При анализе 4 моделей нейронных сетей и испытании их на реальных данных лучшие результаты показала MC-CNN. По результатам обучения MC-CNN показа-

ла наилучшую производительность и точность – 91.2%. При этом сети SegStereo, GC-Net, SGM-Nets также показали хорошие результаты.

Так как лучший результат среди исследуемых алгоритмов показала модель MC-CNN, то целесообразно создавать свою архитектуру на базе этой сети.

В данной архитектуре (рис. 4) применяется расширенная свертка [5] вместо традиционной операции свертки для того, чтобы увеличить поле восприятия. Благодаря этому поле восприятия экспоненциально расширяется для достижения лучшего результата вычисления цены сопоставления на основе патчей изображения. Данный подход позволяет избежать ухудшения карты признаков. Также объединяются признаки из различных слоев свертки.

Проведенные исследования показали, что предлагаемая модель обеспечивает лучшую производительность, чем имеющийся алгоритм MC-CNN, а также лучше работает на участках со слабой текстурой.

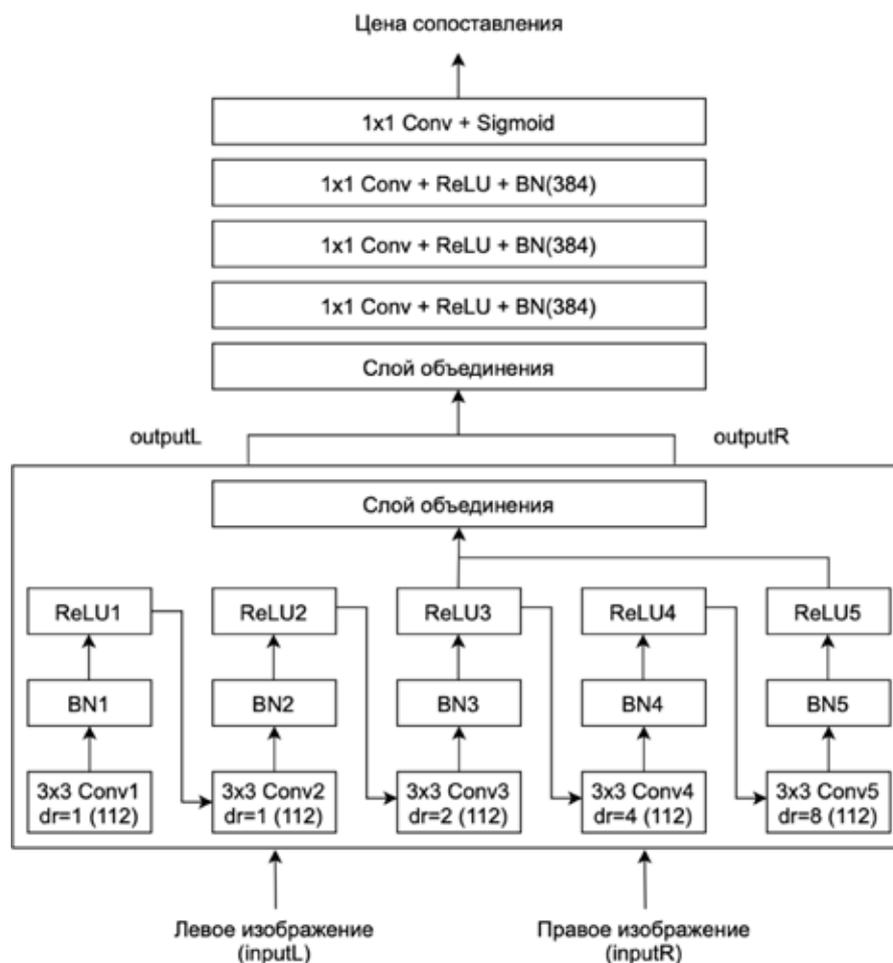


Рис. 4. Архитектура разработанной сети Prophetam-DD

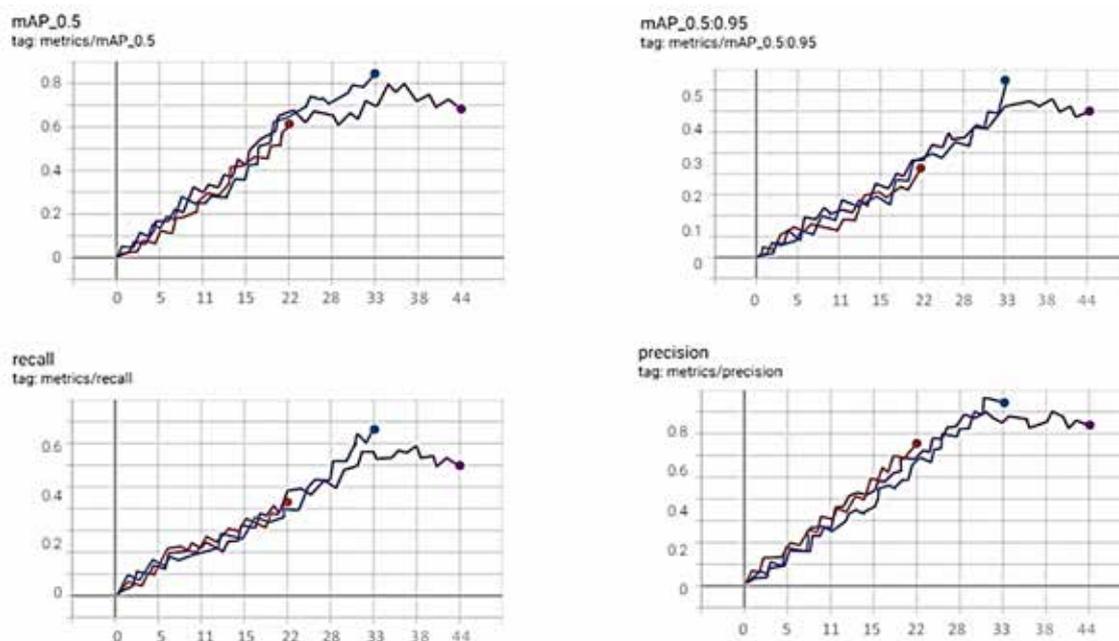


Рис. 5. Построенные системой карты диспаратности

MC-CNN состоит из двух основных частей.

Первая часть представляет собой пару сиамских сетей [6]. Сеть состоит из 5 сверточных слоев, каждый слой при этом имеет ядро 3×3 . Два интересующих нас участка изображения подаются в сеть (входные данные). Затем объекты, которые были извлечены из сиамских сетей, объединяются в качестве конечного результата.

Вторая часть состоит из трех полносвязных (FC) слоев с 384 нейронами и слоя, в котором в качестве функции активации применяется сигмоида [7]. Во избежание повторных вычислений и повторного использования модели на всем изображении полносвязные слои заменяются сверточными слоями с ядром 1×1 .

Традиционные сверточные слои были заменены расширенными сверточными слоями. На сверточных слоях с ядром 3×3 устанавливаются коэффициенты расширения (1, 1, 2, 4, 8), при этом увеличивая поле восприятия с 11×11 до 33×33 .

Вместо того чтобы использовать вывод conv5 только в качестве признаков изображения, вывод conv3, conv4 и conv5 объединяется как дескриптор.

В качестве метода оптимизации используется стохастический градиентный спуск (Stochastic gradient descent) [8]. Данный метод использует один пример из всей тренировочной серии для аппроксимации градиента, по которому затем происходит корректировка весов модели.

Итоговая модель обучалась 33 эпохи с коэффициентом обучения, равным 0.002, и моментом градиентного спуска 0.9.

Для тренировки сети использовался фреймворк PyTorch [9], для загрузки, предобработки и аугментации изображений использовалась библиотека компьютерного зрения OpenCV. Дальнейшая обработка выходных данных нейронной сети осуществляется вызовом ядер CUDA при помощи библиотеки CuPy.

При помощи ядер CUDA реализован параллельный алгоритм полуглобального сопоставления [10]. Его идея состоит в оптимизации цены сопоставления в различных направлениях и расчета на ее основании средней диспаратности. В данной работе используется вариация алгоритма с оптимизацией в 4 направлениях.

Дальнейшая постобработка осуществляется при помощи алгоритма устранения конфликтов. Этот этап используется для улучшения качества правой карты диспаратности посредством исключения конфликтующих значений между левой и правой картой.

Для оценивания точности системы используется собранный набор данных.

Оценка точности заключается в сопоставлении карты диспаратности на основании стереопар оценочного набора данных и в последующем сравнении отклонений предсказанных значений от известных истинных из набора. Если отклонение превышает установленное граничное значение,

делается вывод об ошибочности данного предсказания.

Для данной работы был выбран способ оценивания «bad2.0», в котором рассчитывается общий процент ошибочных предсказаний на всех стереопарах оценочного набора.

Для выбора оптимального срока обучения, с целью избежания ситуации недообучения или переобучения модели, было принято решение обучить модель на различном количестве эпох и сравнить полученные результаты. Сравнение обучений в 22, 33 и 44 эпохах и их визуализация осуществлены при помощи TensorFlow (рис. 5).

Оранжевый график – 22 эпохи, синий график – 33 эпохи, фиолетовый график – 44 эпохи; mAP (mean average precision) – метрика качества ранжирования; recall (полнота) – метрика, которая показывает, как много объектов класса Positive модель смогла распознать; precision (точность) – метрика, которая показывает, как много объектов, которые модель пометила классом Positive, действительно относятся к этому классу. По результатам обучения на 33 эпохах Prophetam-DD показал наибольшую производительность и точность – 96.78%.

Заключение

На основе архитектуры MC-CNN был разработан алгоритм детектирования дефектов дорожного полотна – Prophetam-DD. В искомую архитектуру были добавлены слои расширенной свертки вместо использования традиционной свертки. Это было сделано для того, чтобы расширить поле восприятия и избежать ухудшения карт признаков. В итоге разработанный нейросетевой алгоритм детектирования структурированных объектов на видеоизображении

Prophetam-DD показал высокие результаты и точность детектирования, равную 96,78%.

Список литературы

1. Gehrig S., Franke U. Stereovision for ADAS. Handbook of Driver Assistance Systems. Springer. Cham. 2016. DOI: 10.1007/978-3-319-12352-3_22.
2. Lu Y., Hu Y., Xiao J. Help LabelMe: A Fast Auxiliary Method for Labeling Image and Using It in ChangE's CCD Data. Image and Graphics. ICIIG, 2019. Lecture Notes in Computer Science. Vol 11901. Springer, Cham. DOI: 10.1007/978-3-030-34120-6_65.
3. Gorodnichev M.G., Dzhabrailov K.A., Polyantseva K.A., Gematudinov R.A. On Automated Safety Distance Monitoring Methods by Stereo Cameras. 2020 Systems of Signals Generating and Processing in the Field of on Board Communications (Moscow, 19–20 March 2020). Moscow: Institute of Electrical and Electronics Engineers Inc., 2020. P. 9078616. DOI:10.1109/IEEECONF48371.2020.9078616.
4. Zbontar J., LeCun Y. Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research. 2016. P. 1-32. DOI: 10.48550/arXiv.1510.05970.
5. Yu Fisher, Vladlen Koltun. Multi-Scale Context Aggregation by Dilated Convolutions. CoRR. 2016. DOI: 10.48550/arXiv.1511.07122.
6. Chicco D. Siamese Neural Networks: An Overview. Artificial Neural Networks. Methods in Molecular Biology. Vol 2190. Humana, New York, NY, 2020. DOI: 10.1007/978-1-0716-0826-5_3.
7. Sharma Sudhir, Chandra Pravin. An Adaptive Sigmoidal Activation Function Cascading Neural Networks. Advances in Intelligent and Soft Computing. 2011. V. 87. P. 105-116. DOI: 10.1007/978-3-642-19644-7_12.
8. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем / Пер. с англ. СПб.: ООО «Альфа-книга», 2018. 688 с.
9. Макмахан Б., Рао Д. Знакомство с PyTorch: глубокое обучение при обработке естественного языка = Natural Language Processing with PyTorch. Build Intelligent Language Applications Using Deep Learning. СПб.: Питер, 2020. 256 с.
10. Hirschmuller H. Accurate and efficient stereo processing by semi-global matching and mutual information. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). 2005. V. 2. P. 807-814. DOI: 10.1109/CVPR.2005.56.