

Сегментация изображений трещин дорожного покрытия

И.А. Канаева¹, Ю.А. Иванова¹
iap15@tpu.ru|jbolotova@tpu.ru

¹Томский политехнический университет, г. Томск, Россия;

Вследствие развития потребностей дорожной отрасли в компьютерной обработке высококачественных видеоданных автомобильных дорог возникает необходимость в разработке алгоритма автоматической оценки качества дорожного покрытия. В работе представлен аналитический обзор по теме автоматического детектирования дефектов дорожного покрытия на изображениях. Разработан метод нейросетевого выделения трещин покрытия автомобильных дорог по изображению общего вида сцены, соответствующей виду водителя. Также представлен способ генерации синтетической обучающей выборки и оценена его применимость к данной задаче. Подчеркнута актуальность исследования по сегментации дефектов, что остается сложной задачей из-за неоднородной интенсивности пикселей, сложной топологии трещин, различного освещения и зашумленной текстуры покрытия.

Ключевые слова: синтетическая выборка, сверточные нейронные сети, пиксельная сегментация, трещины покрытия, автомобильные дороги.

Instance Segmentation of Road Pavement Cracks

I.A. Kanaeva¹, Yu.A. Ivanova¹
iap15@tpu.ru|jbolotova@tpu.ru

¹Tomsk Polytechnic University, Tomsk, Russia;

Robust automatic pavement crack detection is critical to automated road condition evaluation. Manual crack detection is extremely time-consuming. Therefore, an automatic road crack detection method is required to boost this progress. This study makes literature review of road damage detection issues. The paper considers the existing datasets for detection and segmentation distress of road and asphalt pavement. A CNN for pavement cracks instance segmentation has been developed with the use of images from the driver's seat view. A method for generating a synthetic dataset is also presented, and effectiveness of its applicability to the current problem is evaluated. The relevance of the study is emphasized by research on pixel-level automatic damage detection remains a challenging problem, due to heterogeneous pixel intensity, complex crack topology, poor illumination condition, and noisy texture background.

Keywords: synthetic dataset, CNNs, instance segmentation, pavement's crack, autoroad.

1. Введение

На сегодняшний день прогресс в компьютерном зрении в значительной степени обусловлен появлением огромного количества размеченных данных. Активно развиваются системы автономного вождения, связанные с анализом изображений окружающей среды, обнаружением и отслеживанием движущихся объектов. Для обучения используются такие наборы данных семантической сегментации, как Cityscapes [1], Wilddash [16] и KITTI [4]. Разметка таких выборок производится вручную и является дорогой и трудозатратной. Выборки в основном содержат в себе экземпляры таких классов, как дорожное полотно, пешеход, транспортное средство, небо, дорожный знак и другие характерные, общие элементы автомобильной дороги.

В настоящее время в Российской Федерации принят и активно развивается национальный проект «Безопасные и качественные автомобильные дороги», целью которого является приведение автомобильных дорог регионального значения и дорожной сети городских агломераций в нормативное состояние.

Вследствие развития потребностей дорожной отрасли в компьютерной обработке высококачественных видеоданных автомобильных дорог возникает необходимость в разработке алгоритма автоматического детектирования дефектов дорожного полотна по изображению.

Разработка эффективного алгоритма детектирования дефектов дорожного полотна на изображениях является актуальной задачей, так как его результаты могут быть использованы как в дорожных организациях, так и в беспилотных автомобилях.

2. Аналитический обзор

За последнее десятилетие был проведен ряд исследований в области автоматического обнаружения и классификации дефектов дорожного покрытия методами машинного обучения и технического зрения. Большое количество работ используют разнообразные методы анализа изображений, подробный анализ предложенных методов представлен в обзоре [9].

В работе [3] при проведении анализа существующих подходов к распознаванию трещин выделяют следующие две общие группы методов: фильтрация изображений и создание классификаторов.

К первой группе методов относится обработка изображений для обнаружения текстур повреждений, которые впоследствии могут быть выделены на дороге путем пороговой фильтрации. Вначале применяются алгоритмы предварительной обработки изображения для уменьшения артефактов, вызванных освещением. Основываясь на том, что пиксели трещин на изображении покрытия сверху принадлежат локальным минимумам интенсивности, применяется пороговая фильтрация. Полученное изображение трещины дополнительно уточняется с помощью морфологических операций и поиском компонент связности. Подходы, принадлежащие к вышеупомянутой группе, представлены в работах [2,7,15]. В [10] результаты исследования реализованы в виде программного инструмента CrackIT для сегментирования трещин по изображению, снятому непосредственно над покрытием. Инструмент CrackTree [18] основан на построении карты вероятности принадлежности к трещине по предварительному очищенному от теней изображению покрытия. В статье [7] предложен метод детектирования трещин на серии изображений, сглаженных

разномасштабными фильтрами Гаусса, с последующим объединением результирующих масок. В работе [12] используются фильтры Габора для поиска кандидатов-трещин. Российские ученые в работе [19] реализовали интерактивный алгоритм для выделения дефекта на изображении покрытия, в основе которого лежит использование метода активных контуров.

Алгоритмы второй группы применяют различные типы классификаторов к участкам изображения для определения, относится ли участок к повреждениям покрытия или нет. Одна часть исследователей [5,14] первоначально выделяет некоторый вектор признаков с рассматриваемого участка, а затем его подает на вход классификатору. В данном подходе достоинством является то, что размер выделяемого участка не является четко фиксированным, но при этом первоначально нужно выделить и разбить исходное изображение на регионы-кандидаты. В работе [14] области определяются с помощью суперпиксельной сегментации, а в качестве бинарного классификатора использовались машины опорных векторов. Несмотря на то, что метод не обладает высокой точностью, он позволяет легко рассчитать соотношение поврежденного и неповрежденного дорожного покрытия и может быть дополнен для выделения разных дефектов: выбоин, люков, разметки.

С появлением свободно распространяемых наборов изображений дефектов автомобильных дорог, таких как GAPs [3] и CRACK500 [17], многие исследователи использовали глубокое обучение для детектирования дефектов дорожного полотна. Например, для извлечения вектора признаков из входного изображения в работе [5] используется усеченная СНС VGG16, а в качестве классификатора обучают нейронную сеть с одним скрытым слоем из 256 нейронов.

3. Наборы данных и типы задач

Рассмотрим наиболее известные доступные наборы данных, связанные с дефектами дорог:

1. Набор данных GAPs [3]: 1969 изображений покрытия в оттенках серого из трех городов Германии разрешением 1920×1080 пикселей, разделенные на фрагменты 64×64 пикселя, которые имеют бинарный признак наличия трещин.
2. CRACK500 [15]: 500 изображений RGB асфальтового покрытия с трещинами, разрешением примерно 2000 x 1500 пикселей, полученные с помощью смартфона на территории кампуса Темплского университета. Каждое изображение снабжено попиксельной бинарной маской принадлежности к трещине.
3. Набор данных CrackTree200 [18]: 206 изображений покрытия размером 800×600 пикселей с различными типами трещин, содержащий не только однородную текстуру фона, но и тени. Каждое изображение снабжено попиксельной разметкой.
4. Набор данных CFD [13]: 118 изображений трещин размером 480×320, семантически сегментированных, снятых по виду сверху на городских дорогах Пекина. Имеют тени, масляные пятна и водяные разводы.
5. Набор данных RoadDamageDataset [8]: 9053 изображений со смартфона, закрепленного у лобового стекла автомобиля и направленного на съемку общего вида впереди автомобиля. Данный набор имеет восемь типов разрушений дорожного покрытия, выделенных ограничивающими рамками прямоугольной формы. Всего выделено 15457 экземпляров разрушений, а сам набор данных имеет структуру PASCAL VOC.

Набор данных RoadDamageDataset записывался в семи городах Японии, и включает восемь типов разрушений дорожного покрытия: пять классов для трещин, два класса

для износа разметки и один класс для выбоин и просадок. Набор данных имеет структуру PASCAL VOC и был представлен на форуме «IEEE Big Data Cup» в 2018 году.

Набор данных японских ученых возродил интерес к решению задачи автоматического детектирования дефектов с помощью методов машинного обучения и, в частности, к использованию сверточных нейронных сетей. Достоинством выборки является её солидный размер, а также наличие других типов разрушений покрытия, а не только трещин. Недостатком же является способ выделения дефектов – ограничивающая рамка, т.к. ввиду разнообразных форм и размеров дефектов с помощью ограничивающие рамки можно лишь судить о его наличии на изображении. Для целей оценки качества автомобильной дороги наилучшим вариантом является попиксельное выделение с помощью маски, что позволяет не только точно локализовать дефект, но и оценить его площадь.

Создание описанного набора данных является трудозатратным и длительным процессом, так как требует кропотливого ручного выделения дефектов на изображении. Однако существует и другой подход к данной проблеме – искусственная генерация данных. Метод генерации синтетических изображений для обучения нейросетей активно развивается в настоящее время. Он позволяет получить как труднодоступные данные, так и размножить существующие.

Наиболее простым и часто используемым способом создания синтетических данных является применение случайных преобразований к исходным изображениям: обрезка, сдвиг, поворот, масштабирование, гауссово размытие, изменение контраста и цветности.

Ввиду значительных успехов, достигнутых в компьютерной графике, одним из популярных направлений создания данных является моделирование 3D сцены и рендеринг фотореалистичных изображений. Этот подход позволяет создавать большие наборы данных для семантической сегментации, как показано в исследовании [11] с использованием Unity платформы.

В статье [20] генерация обучающих изображений дорожных знаков производилась с помощью порождающих конкурирующих нейронных сетей. Авторы показали, что порождающие сети способны генерировать реалистичные изображения, однако достигаемое качество обучения на этих данных классификатора немного уступает классификатору, обученному на реальных данных.

Еще одним способом создания данных является метод извлечения объекта из одного изображения и вставка его в другое. Данный подход продемонстрирован в данной работе для создания синтетической трещины на дороге. Для этого предлагается использовать наборы данных со съемкой сверху и попиксельной разметкой и такие общеизвестные коллекции с выделенным дорожным полотном, как KITTI [7] и Cityscapes dataset [1].

4. Алгоритм создания синтетической выборки

На данном этапе развития машинного обучения формирование множества обучающих данных имеет первостепенное значение для успешного решения задач детектирования и сегментирования. Однако кропотливая ручная разметка нескольких тысяч изображений является долгим и трудозатратным процессом, вследствие чего разработка методов получения репрезентативной синтетической выборки является актуальной задачей.

Синтетический набор данных – это набор обучающих данных, которые генерируются программным путем, т.е. такие данные нельзя получить в реальном исследовании.

Для создания синтетической выборки дефектов автомобильных дорог было решено использовать три

общедоступных набора: KITTI и Cityscapes dataset в качестве изображений дорожной сцены, CFD как источник трещин, размеченных на уровне пикселей (рис. 1).

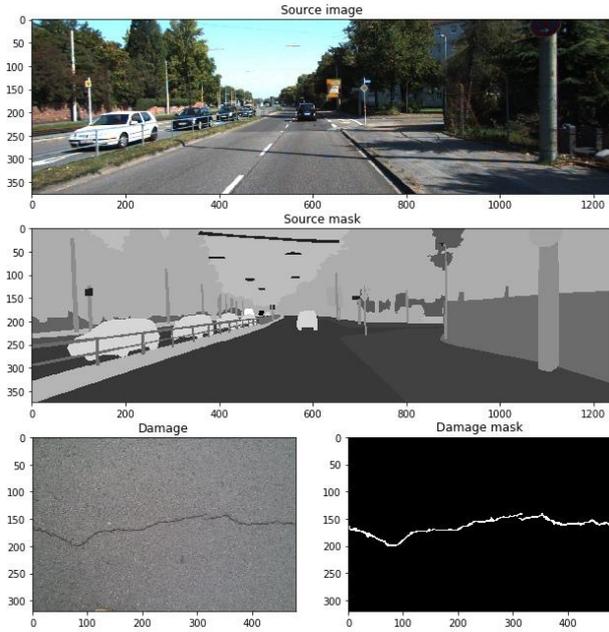


Рис. 1. Размеченные изображения дороги и трещины.

Для определения части изображения, соответствующей проезжей части, выделяются все пиксели маски дороги. Затем к полученной бинарной маске применяется алгоритм поиска 8-ми связанных областей. В результате за маску покрытия принимается область, максимальная по количеству пикселей (на рис. 2 выделена серым цветом).

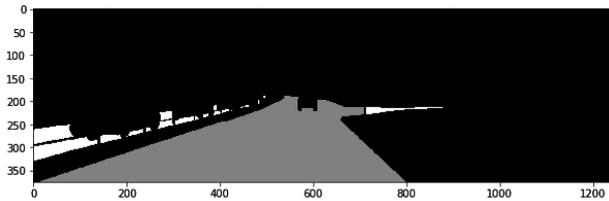


Рис. 2. Максимальная область проезжей части

Затем производится подготовка трещины – для уменьшения вычислений исходное изображение трещины и её маска обрезаются по минимальной ограничивающей прямоугольной рамке. Затем изображение трещины вместе с маской масштабируются и поворачиваются случайным образом. В результате получается изображение трещины D и его маска D^{mask} (рис. 3).

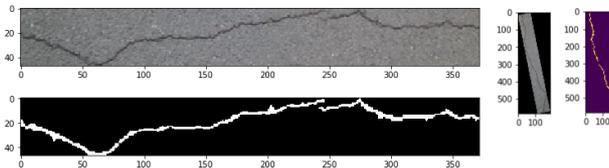


Рис. 3. Обрезка, поворот и масштабирование трещины

На следующем шаге внутри маски базовой области дороги выбирается точка, являющаяся центром накладываемой области, и вырезается область S^{mask} из маски исходного изображения, равная по размеру D^{mask} . Для смещения двух изображений в области масок рассчитываются средние значения \bar{D}_c в каждом канале c по изображению трещины, не лежащему под маской:

$$\bar{D}_c = \frac{1}{k} \sum_p D_c(p) \cdot (1 - D^{mask}(p)), c \in \{R, G, B\},$$

где $D_c(p)$ – значение пикселя p изображения трещины D в канале c , $D^{mask}(p)$ – значение пикселя p в бинарной маске

трещины, а k – количество пикселей, для которых $D^{mask}(p) = 0$.

Проведение данного вычисления возможно ввиду однородности текстуры асфальта на изображении дефекта. Затем при наложении трещины на изображение дороги рассматриваются только значения, лежащие под маской проезжей части и трещины:

$$M(p) = D^{mask}(p) \cdot S^{mask}(p)$$

Нанесение трещины осуществляется путем изменения значений пикселей исходного изображения S , находящихся под общей маской M , пропорционально отношению вычисленных средних значений асфальта к самим значениям пикселей трещины:

$$S_c(p) = S_c(p) \cdot M(p) \cdot \frac{D_c(p)}{\bar{D}_c}$$

Результат работы алгоритма представлен на рис. 4.



Рис. 4. Результат программной генерации трещин

Для увеличения информативности обучающей выборки на одном изображении генерировалось от 1 до 5 трещин, которые могли пересекаться. В результате обучающая выборка содержала 1524 изображения, а тестовая 505.

5. Модель СНС

Для одновременного решения задачи детектирования трещин и их попиксельного выделения решено было использовать современную архитектуру сверточной сети Mask R-CNN [6]. Рассмотрим ее строение и принцип работы.

Архитектура Mask R-CNN (рис. 5) исторически имеет следующий ряд предшественников, основывающихся на идее обработки небольших областей: R-CNN, Fast R-CNN, Faster R-CNN.

Архитектура Mask R-CNN имеет сложную блочную структуру. Вначале изображение подается на вход нейросети для выделения карты признаков, в качестве которой часто используются VGG16, ResNet50 и ResNet101 с исключенными слоями, отвечающими за классификацию. Одним из улучшений данной архитектуры по сравнению с предшественниками является использование подхода Feature Pyramid Network (FPN), который извлекает разномасштабные карты признаков. Последовательные слои СНС с уменьшающейся размерностью рассматриваются как иерархическая «пирамида», в которой карты нижних уровней имеют высокое разрешение, а карты верхних уровней – высокую обобщающую, семантическую способность.

Полученные карты признаков обрабатываются в блоке Region Proposal Network (RPN), задачей которого является генерация предполагаемых регионов на изображении, которые содержат объекты. Для этого по карте признаков скользят нейросеть с окном 3x3 и формируют выход на основе k анкоров – рамок заданной размерности и положения. Для каждого анкера RPN формирует предсказание наличия объекта, и уточнение координат ограничивающей рамки объекта, если он был обнаружен.

Назначение данного этапа – выделить регионы интереса, которые могут содержать объекты. В конце дублирующиеся регионы отбрасываются благодаря операции не-максимального подавления.

Далее с помощью операции ROIAlign происходит выделение соответствующих регионам значений из карт

признаков и приведение их к одному размеру. По ним осуществляется финальные операции классификации, уточнения координат ограничивающей рамки и предсказания маски. Маска на выходе имеет сильно

уменьшенный размер, но содержит в себе вещественные значения, с помощью которых при масштабировании маски до размера выделяемого объекта удастся получить достаточную точность.

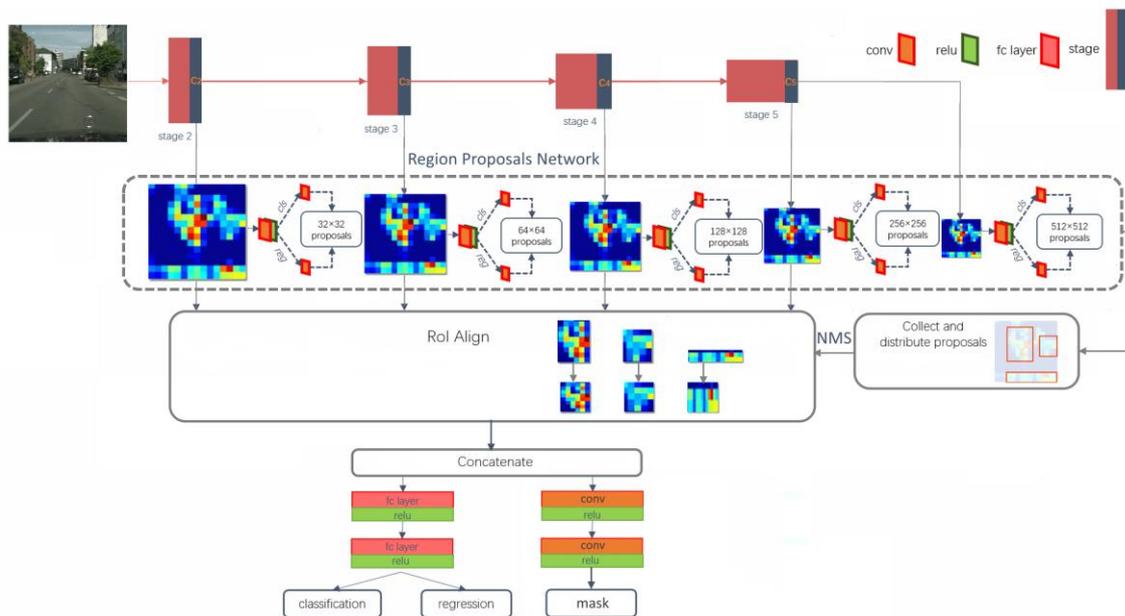


Рис. 5. Архитектура Mask R-CNN

6. Результаты обучения

На полученном синтетическом наборе изображений с попиксельно размеченными трещинами дорожного полотна производилось обучение СНС современной архитектуры Mask R-CNN. В качестве входной нейросети для построения карт признаков использовалась ResNet101. Ввиду того, что синтетический набор данных не может в полной мере отразить все характеристики реальных разрушений асфальта, использовалось технология переноса обучения (transfer learning) с помощью предварительно обученной модели ResNet101 на наборе данных MS-COCO.

На вход СНС подаются RGB изображения размера 1024×1024 , а в качестве масштабов анкоров используются следующие значения: 0.33, 0.5, 1, 2, 3.

Обучение производилось на протяжении 40 эпох по 400 итераций с использованием мини-масок размером 56×56 пикселей для оптимизации используемой памяти компьютера. Наилучшие результаты по метрике средней точности AP при значении IoU = 0.5 при тестировании на синтетическом наборе составили 78,1%.

Метрика пересечения по объединению (IoU) является методом количественного определения процентного перекрытия между целевой маской и результатами выделения. Результат обучения по IoU составил 59,17%.

Для оценки эффективности разработанного подхода к выделению дефектов был вручную подготовлен небольшой набор данных из 50 реальных изображений с трещинами на дороге. На рисунке 6 приведены результаты выделения трещин человеком вручную и нейросетевое попиксельное выделение на реальном изображении.

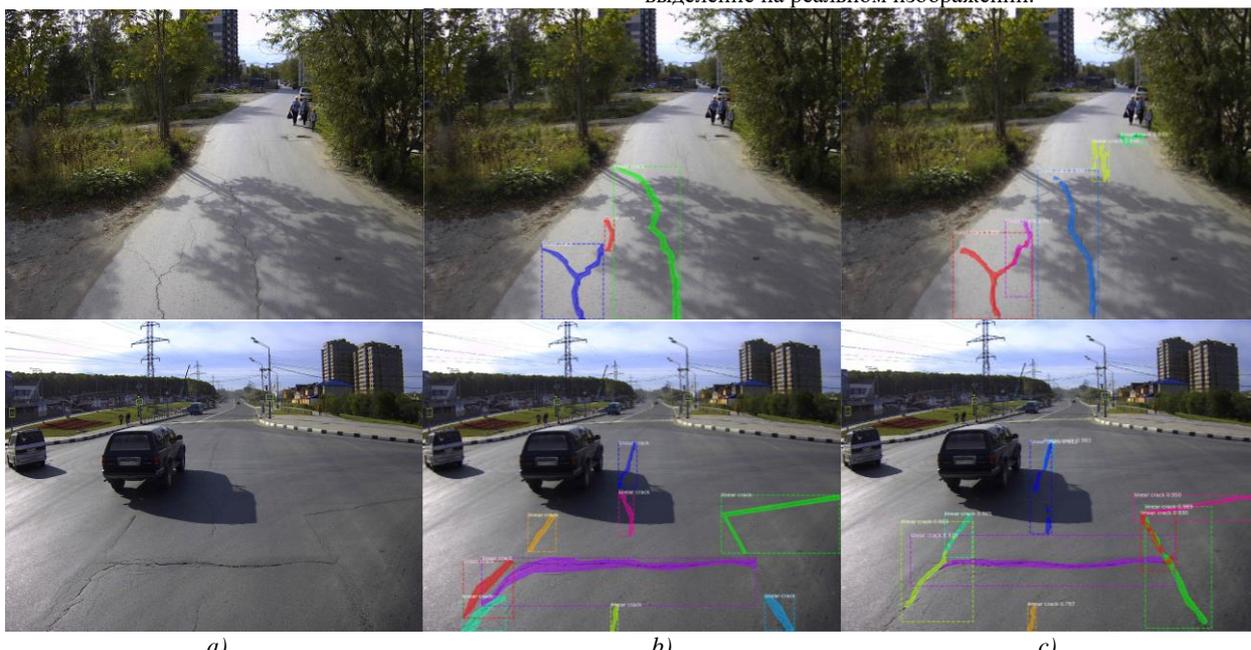


Рис. 6. Сравнение результатов: а) исходное изображение, б) ручное выделение, в) нейросетевое выделение

Ввиду сложной топологии и пересечений, объединений разрушений асфальтового покрытия возможны различные варианты сегментации по экземплярам, что приводит к низкой точности детектирования ограничивающих рамок на реальных изображениях $AP[IoU = 0.5] = 10.98\%$. Однако, оценка качества попиксельной сегментации без привязки к экземплярам составляет $IoU = 46\%$.

Анализ конкретных результатов нейросетевого выделения позволяет судить о том, что обучение на синтетических данных является применимым к данной задаче и позволяет получить приемлемые по точности результаты. Обученная предложенным способом СНС хорошо справляется с выделением небольших дефектов покрытия, мало чувствительна к теням и дорожной разметке. Улучшение результатов детектирования возможно при дополнительном обучении СНС с уделением внимания к дорожным техническим сооружениям, таким как решетки ливневых канализаций и люки колодцев.

7. Заключение

В рамках задачи автоматического детектирования дефектов дорожного полотна был представлен алгоритм генерации синтетической выборки данных. С помощью полученного набора данных происходило обучение СНС Mask R-CNN для детектирования трещин и их сегментации на уровне пикселей.

Аналитический обзор данной области показал, что исследования по обнаружению трещин являются ограниченными, так как автоматическое обнаружение трещин на уровне пикселей остается сложной задачей из-за неоднородной интенсивности пикселей, сложной топологии трещин, различного освещения и зашумленной текстуры покрытия.

Полученные результаты позволяют судить о применимости подхода обучения на синтетической выборке, что позволяет получить лучшие результаты по сравнению с использованием небольшого набора данных, размеченного вручную.

8. Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-08-00977 А.

9. Литература

- [1] Cordts M. [и др.]. The Cityscapes Dataset for Semantic Urban Scene Understanding.
- [2] Danilescu D. [и др.]. Road Anomalies Detection Using Basic Morphological Algorithms // Carpathian Journal of Electronic and Computer Engineering. 2015. № 2(8). С. 15-18.
- [3] Eisenbach M. [и др.]. How to get pavement distress detection ready for deep learning? A systematic approach IEEE, 2017. 2039–2047 с.
- [4] Fritsch J., Kuhn T., Geiger A. A new performance measure and evaluation benchmark for road detection algorithms // IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 2013. С. 1693–1700.
- [5] Gopalakrishnan K. [и др.]. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection // Construction and Building Materials. 2017. (157). С. 322–330.
- [6] He K. [и др.]. Mask R-CNN.
- [7] Li H. [и др.]. Automatic Pavement Crack Detection by Multi-Scale Image Fusion // IEEE Transactions on Intelligent Transportation Systems. 2018. С. 1–12.
- [8] Maeda H. [и др.]. Road Damage Detection and Classification Using Deep Neural Networks with

- Smartphone Images // Computer-Aided Civil and Infrastructure Engineering. 2018. № 12 (33). С. 1127–1141.
- [9] Mohan A., Poobal S. Crack detection using image processing: A critical review and analysis // Alexandria Engineering Journal. 2018. № 2 (57). С. 787–798.
 - [10] Oliveira H., Correia P.L. CrackIT - An image processing toolbox for crack detection and characterization // 2014 IEEE International Conference on Image Processing, ICIP. 2014. С. 798–802.
 - [11] Ros G. [и др.]. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes // IEEE Conference on CVPR, Proceedings, NV. 2016. С. 3234–3243.
 - [12] Salman M. [и др.]. Pavement crack detection using the Gabor filter // IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. 2013. № October. С. 2039–2044.
 - [13] Shi Y. [и др.]. Automatic Road Crack Detection Using Random Structured Forests // IEEE Transactions on Intelligent Transportation Systems. 2016. № 12 (17). С. 3434–3445.
 - [14] Varadharajan S. [и др.]. Vision for road inspection // 2014 IEEE Winter Conference on Applications of Computer Vision, WACV 2014. 2014. С. 115–122.
 - [15] Yang F. [и др.]. Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection 2019. С. 1–11.
 - [16] Zendel O. [и др.]. WildDash - Creating Hazard-Aware Benchmarks.
 - [17] Zhang L. [и др.]. Road crack detection using deep convolutional neural network // Proceedings - International Conference on Image Processing, ICIP. 2016. № October 2017 (2016-Augus). С. 3708–3712.
 - [18] Zou Q. [и др.]. CrackTree: Automatic crack detection from pavement images // Pattern Recognition Letters. 2012. № 3 (33). С. 227–238.
 - [19] Судаков С. [и др.]. Алгоритмы детектирования разметки и дефектов дорожного покрытия 2008.
 - [20] Шахуро, В.И. Синтез обучающих выборок для классификации дорожных знаков с помощью нейросетей / В.И. Шахуро, А.С. Коношин // Компьютерная оптика. – 2018. – Т. 42, № 1. – С. 105-112.

Об авторах

Канаева Ирина Александровна, 1994 года рождения, в 2018 окончила магистратуру Томского политехнического университета по направлению «Информатика и вычислительная техника», специализация «Компьютерный анализ и интерпретация данных». Обучается в аспирантуре Томского политехнического университета. Область научных интересов: обработка и анализ изображений и видео, рекуррентные нейронные сети. E-mail: iap15@tpu.ru.

Иванова Юлия Александровна, 1986 года рождения, в 2009 году окончила Томский политехнический университет по специальности «Программное обеспечение вычислительной техники и автоматизированных систем», к.т.н., доцент отделения информационных технологий. Область научных интересов: обработка изображений, распознавание образов, биологически-подобные алгоритмы. E-mail: jbolotova@tpu.ru.