Известия Гомельского государственного университета
имени Ф. Скорины, № 3 (126), 2021

УДК 004.932

Повышение эффективности обнаружения объектов небольших размеров на 8К изображениях при использовании сверточных нейронных сетей

Р.П. Богуш[1], С.В. Абламейко[2,3], С.А. Игнатьева[1], Е.Р. Адамовский[1]

Рассмотрено применение методики обнаружения объектов на изображениях формата 8К, которая предполагает пирамидальное представление изображения и блочную обработку с перекрытием на каждом уровне с использованием сверточной нейронной сети (СНС). В качестве СНС при обработке блоков применяется YOLOv4. Представлен анализ архитектуры данной СНС и описаны основные особенности, которые позволяют обеспечивать высокую результативность ее работы. Для проведения экспериментов по оценке эффективности предложенного подхода подготовлена база данных с размеченными объектами на изображениях формата 8К двух классов «человек» и «транспортное средство». Для оценки качества работы вычислялась величина mAP для различных сочетаний таких параметров, как степень пороговой уверенности YOLOv4 и процент взаимного пересечения блоков при иерархическом представлении 8К изображения. Приведены результаты исследований.
Ключевые слова: блочная обработка, многомасштабное представление изображения, детектирование объектов, архитектура YOLOv4.

The application of the method of object detection in 8K images, which assumes a pyramidal representation of the image and block processing with overlap at each level using a convolutional neural network (CNN) is considered. YOLOv4 is used as a CNN for blocks processing. The analysis of the architecture of this SNS is presented and the main features are described that allow ensuring high efficiency of its work. To conduct experiments to assess the effectiveness of the proposed approach, a database with marked objects in 8K images of two classes «person» and «vehicle» was prepared.
To assess the quality of work, the mAP value was calculated for various combinations of such parameters as the degree of threshold confidence YOLOv4 and the percentage of mutual intersection of blocks in the hierarchical representation of an 8K image. Experimental results are presented.
Keywords: block processing, multi-scale image representation, object detection, YOLOv4 architecture.

Введение. Непрерывное развитие видеокамер приводит к повышению качества получаемых цифровых изображений и видеопоследовательностей, в том числе и за счет увеличения разрешения. В последнее время достаточно широко используется формат 4К, развитие которого предполагает увеличения размеров изображений до 8К. В этом случае размер большей стороны, как правило, горизонтальной, составляет около 8000 пикселей. Соответственно, на таких изображениях в системах компьютерного зрения предоставляется возможность более точного обнаружения объектов, в том числе и небольших размеров. В настоящее время для детектирования объектов применяются сверточные нейронные сети, которые обладают наилучшей обобщающей способностью среди известных методов и позволяют наиболее эффективно решать задачу автоматического обнаружения и локализации объектов. Однако возможность их практического применения в значительной мере определяется аппаратными средствами, а именно такими основными характеристиками графических процессоров, как быстродействие и доступный объем памяти. В связи с этим на первом шаге обработки с использованием СНС размеры изображения формата 8К будут значительно уменьшены, что приведет к потере информативности и уменьшению точности обнаружения объектов малых размеров. Данная статья посвящена решению задачи повышения эффективности обнаружения объектов небольших размеров на изображениях формата 8К при использовании СНС.

Теоретический анализ. Повышение точности обнаружения объектов небольших размеров на изображениях с высоким разрешением рассматривается во многих работах, но наиболее эффективными являются алгоритмы, предполагающие блочное разбиение изображения и применении СНС для каждого блока при обработке изображений формата 4К [1], [2]. При этом результативность детектирования объектов в значительной степени определяется приемами объединения результатов обнаружения фрагментов объектов на границах блоков и применяе-

мой СНС. Среди существующих эффективной является методика обнаружения объектов на основе их пирамидально блочной обработки с использованием СНС YOLOv3, результативность которой на 4К изображениях показана в [1]. Однако YOLOv3 хотя и обладает достаточно высокой точностью, скорость работы значительно медленнее по сравнению с предыдущими версиями данной СНС. Поэтому для обработки 8К изображений, имеющих значительно большие размеры по сравнению с 4К, желательно использовать СНС, требующую меньших временных затрат, но не ухудшающую точность обнаружения. Для такой задачи перспективной является новая архитектура СНС YOLOv4, которая направлена на увеличение скорости работы и оптимизацию параллельных вычислений [3]. Общая схема YOLOv4 представлена на рисунке 1.

Структуру YOLOv4 можно разделить на три основных блока: блок извлечения признаков, который служит для выявления характерных особенностей объектов на входном изображении; блок сбора карт признаков с разных слоев, который собирает и передает карты признаков с различных уровней нейронной сети на блок обнаружения и классификации, который, в свою очередь, формирует выходные карты признаков, для разных масштабов, что позволяет предсказывать координаты ограничительных рамок искомых объектов и классифицировать содержимое каждой ячейки на входном изображении. На рисунке 1a блок извлечения признаков представляет собой СНС CSPDarknet-53, в основе которой находится Darknet-53, состоящая из 53 сверточных слоев. Отличие заключается в использовании межэтапных соединений (CSP – Cross-Stage-Partial connection) [4] для снижения вычислительной сложности. На рисунке 1b представлена структура блока с CSP-соединением. В данном блоке карты признаков разделяются на две части, одна из них проходит через группу Res-блоков (Residual blocks), другая поступает на сверточный слой, после чего выходные данные объединяются. Res-блоки представлены в группах по 1, 2, 4 или 8 блоков (рисунок 1c), каждый из которых состоит из сверточных слоев 1 × 1 и 3 × 3 с замыкающим соединением (shortcut connection) (рисунок 1d). Использование такого соединения обеспечивает альтернативный путь для градиента, что приводит к лучшей сходимости модели. Другим отличием CSPDarknet-53 является использование функции активации Mish [5], которая представляет собой комбинацию из функции идентичности, гиперболического тангенса, softplus и характеризуется достаточно низкой вычислительной сложностью и определяется как:

$$f(x) = x\tanh\left(softplus(x)\right) = x\tanh\left(\ln\left(1+e^{x}\right)\right).$$

Кроме этого, Mish является немонотонной, гладкой и непрерывной функцией, неограниченной сверху, но ограниченной снизу. Отсутствие верхней границы позволяет избегать насыщения, которое может приводить к замедлению обучения, а значит позволяет ускорить процесс обучения. Наличие нижней границы обеспечивает эффект регуляризации. Немонотонность сохраняет небольшие отрицательные значения, что стабилизирует градиентный поток, а гладкость и непрерывность эффективны при обобщении и оптимизации результатов.

Блок сбора карт признаков с разных слоев включает модифицированный модуль SPP (SPP-Spatial Pyramid Pooling) (рисунок 1e) для увеличения рецептивного поля и модифицированный модуль PAN для создания пирамиды признаков [6]. Первый из них выполняет операцию maxpool по картам признаков $\frac{N}{32}\times\frac{N}{32}\times 255$ с разным размером ядра $k=\{1,5,9,13\}$, но идентичным заполнением, которое используется для сохранения пространственного размера. Затем четыре комплекта соответствующих карт признаков объединяются в один, размером $\frac{N}{32}\times\frac{N}{32}\times 2048$. Это увеличивает область анализа, улучшая таким образом точность модели, при этом без существенных вычислительных затрат.

Для агрегирования характеристик объектов в YOLOv4 используется модифицированная версия PAN (Path Aggregation Networks), использующая пирамиды признаков, которые служат для объединения карт признаков нижних и верхних уровней сети [7], то есть сочетают семантическую информацию с верхних уровней и более точную с нижних. В этой модификации PAN вместо операции суммирования результатов соседних слоев, используется конкатенации, что позволяет увеличить точность прогнозов.

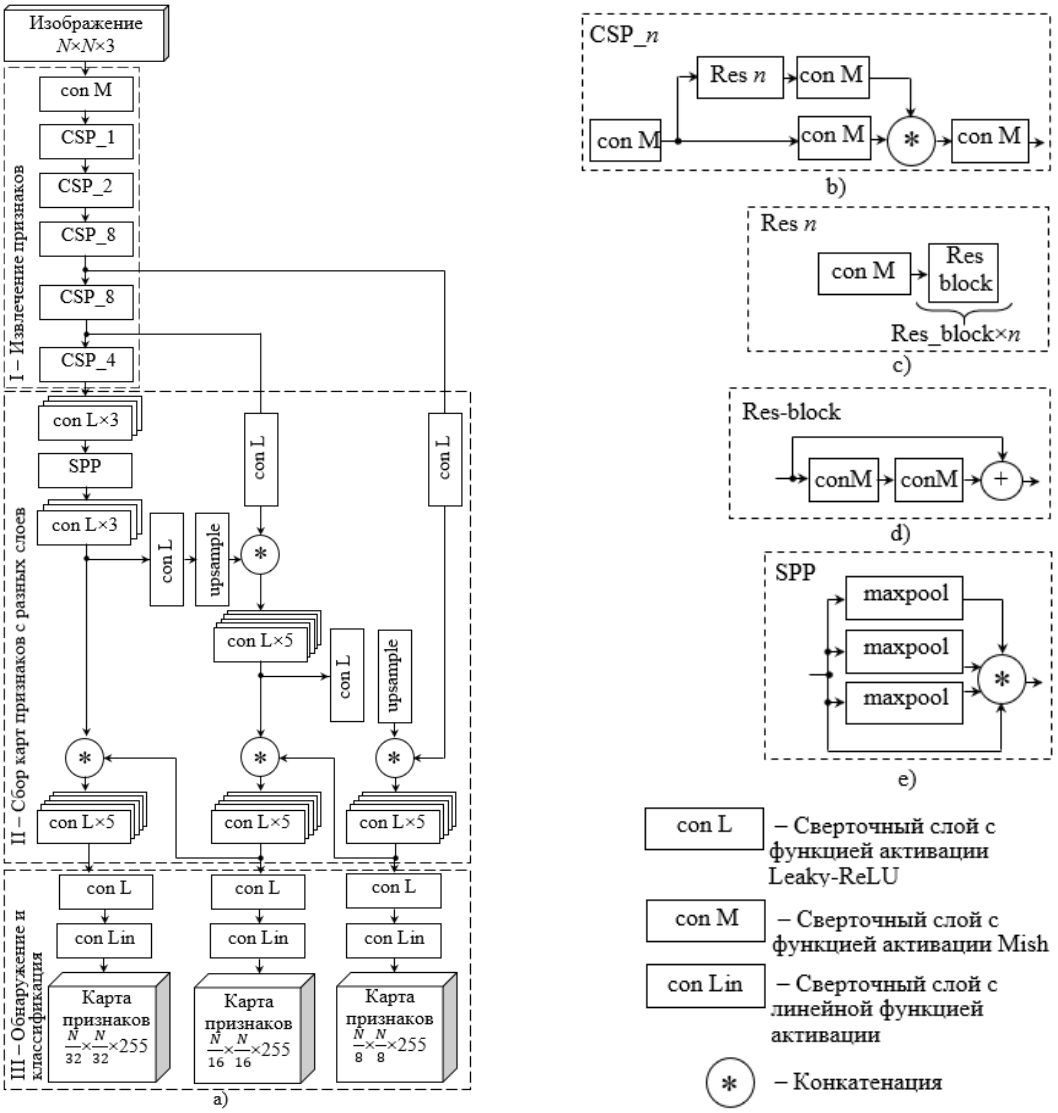

Рисунок 1 – Архитектура YOLOv4: a) общая схема; b) структура блока с CSP-соединением; c) объединение *n* Res-блоков; d) структура Res-блока; e) блок SPP

Блок обнаружения и классификации используется для нахождения ограничительной рамки и классификации содержимого в каждой ячейке. Изображение размером $N \times N \times 3$, подаваемое на вход нейронной сети, разделяется на $S \times S$ ячеек – которые представляют собой область интереса (RoI – region of interest). Количество ячеек прямо пропорционально размеру входного изображения и обратно пропорционально шагу дискретизации на каждом из масштабов. Карты признаков на выходе сети YOLOv4 используют шаги дискретизации 32, 16 и 8, которые показывают во сколько раз уменьшился размер выходного слоя относительно входного изображения. В YOLOv4 к каждой ячейке сетки для предсказания координат и размеров рамки, ограничивающей объект на изображении, используются три заранее заданных anchor-рамки, которые представляют собой прямоугольники различных размеров с разным соотношением сторон. Вектор признаков включает 85 параметров для anchor-рамки, включая смещение координат t_x и t_y, и отклонения размеров t_w и t_h, на основании которых формируется предсказанная рамка (рисунок 2), вероятность обнаружения объекта и вероятности принадлежности объекта к каждому из 80 классов. Тогда вектор признаков для каждой ячейки изображения с учетом трех anchor-рамок будет равен $1 \times 1 \times 255$. Предсказанная рамка описывается 85 параметрами, включая вероятность ее правильного определения (P_c), координаты центра $(b_x$ и $b_y)$, высоту и ширину $(b_h$ и $b_w)$ и вероятности нахождения объекта одного из 80 классов $(c_1 \ldots c_{80})$. Класс объекта, ограниченного предсказанной рамкой, определяется с использованием $c_1 \ldots c_{80}$ и сигмоидальной функции.

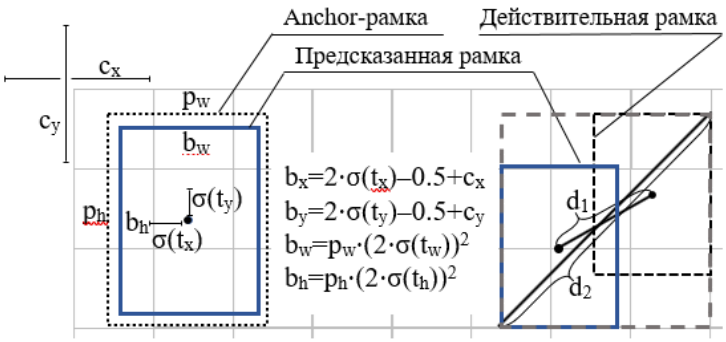

Рисунок 2 – Построение ограничительных рамок при обнаружении объектов: c_x, c_y – координаты верхней левой ячейки; p_h и p_w – размеры anchor-рамки; b_w, b_h b_x, b_y, – координаты центра и размеры предсказанной ограничительной рамки; $\sigma(t)$ – функция сигмоиды; t_x, t_y, t_w, t_h – отклонение координат и размеров предсказанной рамки; d_1 – евклидово расстояние между центральными точками действительной и предсказанной рамки; d_2 – диагональ прямоугольника, охватывающего две рамки

Для определения положения ограничительных рамок применяется функция потерь L_{CIoU} [8]:

$$L_{CIoU} = 1 - IoU + \frac{d_1^2}{d_2^2} + \alpha v,$$

где IoU – отношение между пересечением и объединением предсказанной и действительной рамок, и для расчета площадь пересечения делится на площадь объединения; d_1 – евклидово расстояние между центральными точками предполагаемой и действительной ограничительными рамками; d_2 – диагональ минимальной рамки, которая описывает предсказанную и действительную; α – параметр, зависящий от взаимного расположения предсказанной и действительной рамки, определяется как:

$$\alpha = \begin{cases} 0, \text{ если IoU}<0{,}5; \\ \dfrac{v}{(1-IoU)+v}, \text{ если IoU} \geq 0{,}5; \end{cases}$$

$v = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right)^2$ – мера согласованности сторон ограничительных рамок.

При расчете потерь учитывается площадь перекрытия, расстояние между центральными точками и соотношение сторон. L_{CIoU} позволяет увеличить площадь перекрытия действительной и предсказанной ограничительной рамки, минимизировать расстояние между центральными точками и сохранить постоянное соотношение сторон. L_{CIoU} обеспечивает лучшую скорость и точность в задаче приближения координат предсказанной ограничительной рамки к действительной по сравнению с L_{DIoU} (Distance-IoU loss) и L_{GIoU} (Generalized-IoU loss) потерями.

Подавление немаксимумов (NMS) применяется для обработки ситуации, когда для одного и того же объекта предсказано несколько рамок с высокими вероятностями соответствия. В YOLOv4 используется подход на основе greedy NMS, которая следует гипотезе о том, что обнаруженные предполагаемые рамки с большим количеством перекрытий соответствуют одному и тому же объекту. При этом рамки-кандидаты сортируются по степени достоверности классификации и на основе IoU удаляются дублирующие.

Методика проведения эксперимента. Для проведения экспериментов использована методика (рисунок 3), которая предполагает для каждого изображения следующие этапы: инициализация СНС; пирамидальное представления изображения в виде его копий с уменьшающимся масштабом (для 8К изображений число слоев пирамиды равно 4); блочное разбиение слоев пирамиды с перекрытием; обнаружение объектов в каждом блоке; объединение результатов на границах блоков; вычисление IOU для обнаруженных объектов и предварительно размеченных на входном изображении и если полученное значение IOU > 0,5 и объекты принадлежат одному классу, то объект считается обнаруженным.

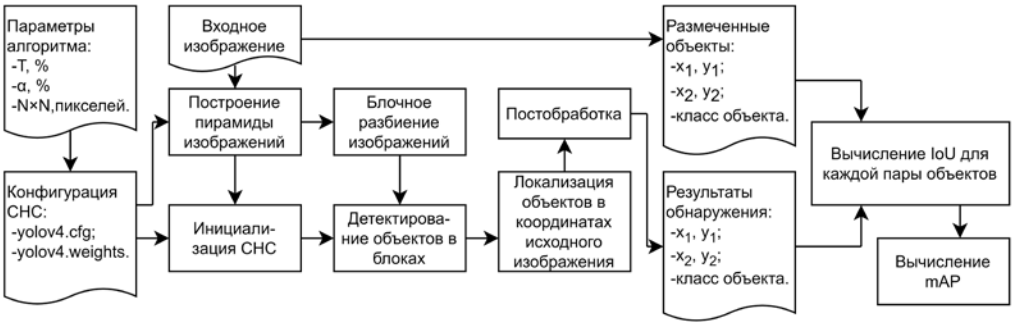

Рисунок 3 – Схема методики эксперимента

В качестве объектов использованы «человек» и «транспортное средство», причем во второй класс включены такие, как «автобус», «машина» и «грузовик». Алгоритм обнаружения объектов на 8К изображениях написан на языке Python 3.7 с использованием технологии CUDA, которая применяется для пакетной обработки блоков на основе СНС, что позволяет значительно уменьшить временные затраты. В качестве СНС использована реализация YOLOv4 на фреймворке PyTorch [9], применены файлы весовых коэффициентов yolov4.weights [10] и конфигурации yolov4.cfg, а также набор из 120 изображений формата 8К [11]–[13], на которых было размечено 4200 объектов, наименьший из них размером 38×10. Таблица 1 содержит результаты определения mAP для различных сочетаний параметров степени пороговой уверенности T СНС и процента взаимного пересечения блоков α при пирамидальном разбиении изображения.

Таблица 1 – Оценка mAP с использованием пирамидально-блочной обработки 8К изображений

		α, %								
		15	20	25	30	35	40	45	50	55
T, %	60	0,5759	0,5651	0,5638	0,5727	0,5299	0,5527	0,5305	0,5616	0,5333
	65	0,5802	0,5720	0,5806	0,5812	0,5515	0,5761	0,5577	0,5676	0,5534
	70	0,5781	0,5825	0,5905	0,6087	0,5621	0,5994	0,5703	0,5841	0,5609
	75	0,5795	0,5876	0,5973	0,5948	0,5656	0,5808	0,5836	0,5835	0,5817
	85	0,5738	0,5620	0,5679	0,5765	0,5739	0,5711	0,5624	0,5664	0,5632

На рисунке 4a представлена уменьшенная копия видеоизображения формата 8К [14], а на рисунках 4b и 4c отображены результаты обнаружения объектов для его правой верхней части на основе СНС YOLOv4 и на основе предлагаемой методики.

a) b)

c)

Рисунок 4 – Примеры обнаружения объектов на 8К изображении: a) уменьшенная копия исходного изображения; b) на основе СНС YOLOv4; c)на основе предложенной методики

Из таблицы 1 следует, что при заданных условиях наиболее оптимальным является сочетание параметров T = 70 %, α = 30 %, для которых получено наибольшее результирующее значение mAP_{max} = 0,6087, при этом размер наименьшего обнаруженного объекта 39 × 17. При использовании YOLOv4 со входным слоем 1024 × 1024 для T = 70 % mAP = 0,295, размер наименьшего обнаруженного объекта 44 × 27. Таким образом, предложенный подход на используемой базе данных 8К изображений позволяет обнаруживать объекты меньших размеров и увеличить результативность обнаружения более чем в два раза.

С использованием СНС YOLOv4 на изображении (рисунок 4b) обнаружено 50 объектов, с минимальным размером 319 × 124 пикселей, а с применением рассмотренной методики найдено 132 объекта (рисунок 4c), размер минимального из них 42 × 36 пикселей. Кроме этого, достоинством предложенной методики для 8К изображений является то, что она позволяет существенно повысить результативность обнаружения при близком расположении множества объектов с их перекрытием. В этом случае признаки объектов будут в значительной мере отличны от полученных в результате обучения СНС, а дальнейшее уменьшение изображения формата 8К к размерам входного слоя еще в большей мере уменьшает информативность таких объектов. Применяемая методика, использующая блочную и иерархическую обработку, позволяет сохранить исходные признаки каждого объекта и обеспечить их правильное обнаружение. На рисунке 5 показаны примеры детектирования близко расположенных людей на 8К изображении.

a) b)

Рисунок 5 – Результаты обнаружения: a) на основе СНС YOLOv4; b) на основе рассмотренной методики

Применение СНС YOLOv4 позволило обнаружить одного человека размером 235 × 110 (рисунок 5a), представленный подход обеспечил обнаружение 210 объектов, минимальный размер 41 × 21 (рисунок 5b).

Заключение. Представлен подход для повышения точности обнаружения объектов небольших размеров на 8К изображениях при применении СНС. На основе проведенного анализа для такой задачи предложено использовать YOLOv4. Рассмотрены особенности архитектуры данной СНС. Показано, что применение предлагаемой методики детектирования объектов на изображениях формата 8К на основе их пирамидального представления и блочной обработки с перекрытием на каждом уровне обеспечивает увеличение mAP на подготовленной базе 8К изображений более, чем в два раза для объектов небольших размеров. Предложенная методика более эффективна при близком расположении объектов с перекрытием на 8К изображении.

Литература

1. Богуш, Р. П. Обнаружение объектов на изображениях с большим разрешением на основе их пирамидально-блочной обработки / Р. П. Богуш, И. Ю. Захарова, С. В. Абламейко // Информатика. – 2020. – Т. 17, № 2. – С. 7–16.

2. Ruzicka, V. Fast and accurate object detection in high resolution 4K and 8K video using GPUs / V. Ruzicka, F. Franchetti // IEEE High Performance Extreme Computing Conference, Waltham, 25–27 Sept. 2018. – Waltham, 2018. – P. 1–7.

3. Bochkovskiy, A. YOLOv4: Optimal Speed and Accuracy of Object Detection [Electronic resource] / A. Bochkovskiy, C. Wang, H. Liao // ArXiv. – Mode of access : https://arxiv.org/pdf/2004.10934.pdf. – Date of access : 12.03.2021.

4. Wang, C. CSPNet: A New Backbone that can Enhance Learning Capability of CNN / C. Wang [et. al.] // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, 14–19 June 2020. – Seattle, 2020. – P. 1571–1580.

5. Misra, D. Mish: A Self Regularized Non-Monotonic Activation Function [Electronic resource] / D. Misra // ArXiv. – Mode of access : https://arxiv.org/vc/arxiv/papers/1908/1908.08681v2.pdf. – Date of access : 12.03.2021.

6. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition / K. He [et. al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2015. – Vol. 37, iss. 9. – P. 1904–1916.

7. Path Aggregation Network for Instance Segmentation/ S. Liu [et. al.] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 18–23 June 2018, Salt Lake City, 2018. – P. 8759–8768.

8. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression [Electronic resource] / Z. Zheng [et. al.] // ArXiv. – Mode of access : https://arxiv.org/pdf/1911.08287.pdf. – Date of access : 12.03.2021.

9. Pytorch-YOLOv4 [Electronic resource]. – Mode of access : https://github.com/Tianxiaomo/pytorch-YOLOv4. – Date of access : 16.02.2021.

10. Yolov4.weights [Electronic resource]. – Mode of access : https://drive.google.com/u/0/open?id=1cewMfusmPjYWbrnuJRuKhPMwRe_b9PaT. – Date of access : 16.02.2021.

11. New York City 8K – VR 360 Drive [Electronic resource]. – Mode of access : https://www.youtube.com/watch?v=2Lq86MKesG4. – Date of access : 18.03.2021.

12. Walk in Shinjuku, Tokyo, Japan @8K 360° VR / Sep 2020 [Electronic resource]. – Mode of access : https://www.youtube.com/watch?v=YYQufxYrBiU. – Date of access : 18.03.2021.

13. Лондон, Великобритания. Виртуальное путешествие. 360 видео в 8К [Электронный ресурс]. – Режим доступа : https://www.youtube.com/watch?v=KGerjHMa90s. – Дата доступа : 18.03.2021.

14. Magic of Hong Kong. Mind-blowing cyberpunk drone video of the craziest Asia's city by Timelab.pro [Electronic resource]. – Mode of access : https://www.youtube.com/watch?v=gYO1uk7vIcc. – Date of access : 18.03.2021.

[1]Полоцкий государственный университет

[2]Белорусский государственный университет

[3]Объединенный институт проблем
информатики НАН Беларуси

Поступила в редакцию 16.04.2021