

В.В. Ковалев**АЛГОРИТМ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИЗОБРАЖЕНИЙ
ДЛЯ СНИЖЕНИЯ ВЕРОЯТНОСТИ ПЕРЕОБУЧЕНИЯ СВЁРТОЧНЫХ
НЕЙРОННЫХ СЕТЕЙ НА НЕЙРОННОМ УСКОРИТЕЛЕ**

Основной объём требований в системах раннего обнаружения объектов предъявляется к производительности алгоритмов цифровой обработки изображений, которые реализуются на встраиваемых устройствах с ограниченным вычислительным ресурсом. В задаче раннего обнаружения объекты на изображениях представлены малым количеством пикселей. Поэтому чтобы обеспечить требуемые характеристики точности алгоритмов поиска и распознавания объектов на изображениях применяют алгоритмы предварительной обработки последовательности видеокадров для расширения исходного признакового пространства. Обработка изображений высокого разрешения алгоритмами предварительной обработки изображений приводит к неприемлемой временной задержке выполнения алгоритма и является «узким местом» всего алгоритма. В работе предложен алгоритм предварительной обработки последовательности видеокадров для нейронного ускорителя с целью расширения признакового пространства, который позволяет увеличить скорость обработки данных. Это достигается за счет слияний алгоритма предварительной обработки изображений с экстрактором признаков свёрточной нейронной сети и переносом выполнения нового экстрактора признаков на вычислительные мощности нейронного ускорителя. Произведена апробация разработанного алгоритма путём проведения вычислительного эксперимента. На вычислительных устройствах NVIDIA Jetson и Rockchip реализован алгоритм предварительной обработки дважды на центральном процессоре и нейронном ускорителе, согласно разработанному алгоритму. Получены оценки времени выполнения алгоритмов, которые показывают, что предложенный алгоритм предварительной обработки изображений для нейронного ускорителя позволяет увеличить скорость обработки данных в 1.4–4 раз в зависимости от типа рядности вычислений. Однако, переход к целочисленному типу вычислений модели СНС с модифицированным экстрактором признаков приводит к снижению метрики Mean Average Precision на 5–19.4%, характеризующей интегральную среднюю точность поиска и распознавания объектов на изображениях.

Свёрточные нейронные сети; вычислительные устройства; распознавание образов; вычислительные устройства.

V.V. Kovalev**IMAGE PREPROCESSING ALGORITHM TO REDUCE THE PROBABILITY
OF OVERFITTING OF CONVOLUTIONAL NEURAL NETWORKS ON A NEURAL
ACCELERATOR**

The main volume of requirements in early detection systems are imposed on the performance of digital image processing algorithms that are implemented on embedded devices with limited computing resources. In the early detection problem, objects in images are represented by a small number of pixels. Therefore, in order to ensure the required accuracy characteristics of the algorithms for searching and recognizing objects in images, algorithms for preliminary processing of a sequence of video frames are used to expand the original feature space. Processing high-resolution images by image preliminary processing algorithms leads to an unacceptable time delay in the execution of the algorithm and is a "bottleneck" of the entire algorithm. In this paper, an algorithm for preliminary processing of a sequence of video frames for a neural accelerator is proposed in order to expand the feature space, which allows increasing the speed of data processing. This is achieved by merging the image preliminary processing algorithm with the feature extractor of a convolutional neural network and transferring the execution of a new feature extractor to the computing power of the neural accelerator. The developed algorithm was tested by conducting a computational experiment. On NVIDIA Jetson and Rockchip computing devices, the algorithm of preliminary processing is implemented twice on the central processor and neural accelerator, according to the developed algorithm. Estimates of the execution time of the algorithms are obtained, which show that the proposed algorithm of preliminary image processing for the neural accelerator al-

lows increasing the data processing speed by 1.4 - 4 times depending on the type of bit depth of calculations. However, the transition to the integer type of calculations of the CNN model with a modified feature extractor leads to a decrease in the Mean Average Precision metric by 5–19.4%, characterizing the integral average accuracy of searching and recognizing objects in images.

Convolutional neural networks; computing devices; pattern recognition; computing devices.

Введение. Поиск и распознавание объектов на изображениях является неотъемлемой задачей современных систем анализа и обработки видеопоследовательности. Такие системы задают высокие требования к критерию точность/быстродействие, чтобы своевременно и точно принимать решения во избежание аварийных ситуаций. Для обеспечения высокой точности поиска и распознавания объектов на изображениях в масштабе реального времени используют свёрточные нейронные сети (СНС). Чтобы обеспечить своевременное принятие решения, бортовые системы должны обнаруживать и распознавать объекты на дальних расстояниях. Чтобы изображение объекта на дальнем расстоянии от места фиксации имело не нулевое признаковое описание, физический размер объектов должен быть большим. Но всё же такие объекты представлены малой группой пикселей на изображении относительно всей площади изображения. Поиск и распознавание малоразмерных объектов на изображениях вызывает трудности даже у передовых архитектур СНС. Неудовлетворительные показатели точности поиска и распознавания малоразмерных объектов связаны с проблемой переобучения СНС. Переобучение – это негативный процесс, который возникает когда объем признакового пространства мал относительно количества степеней свободы (обучаемых параметров) СНС. Характерным проявлением переобучения модели СНС является достижение высоких показателей точности поиска и распознавания объектов на данных, которые были в обучающей выборке, а на тестовых данных, которые сеть «не видела» алгоритм показывает низкую точность поиска и распознавания. Для борьбы с переобучением применяют специальные методы, которые объединил термин регуляризация. Методы регуляризации условно можно разделить на три группы: изменение функции потерь, изменение процесса обучения и архитектуры СНС, преобразование данных. В последнее время большую популярность набирают методы преобразования данных, которые основаны на расширении признакового пространства, которые основываются на алгоритмах предварительной обработки. Однако, алгоритмы предварительной обработки в поиске и распознавании малоразмерных объектов СНС могут быть вычислительно затратными и является «узким местом» выполнения алгоритма в целом, потому что построены не центральном процессоре. Поэтому, проблемы неудовлетворительной временной задержке алгоритмов предварительной обработки изображений является актуальной задачей, которая требует решения.

Поиск и распознавание объектов на изображениях свёрточными нейронными сетями. СНС поиска и распознавания объектов на изображениях условно можно разделить на двухэтапные и одноэтапные подходы. Двухэтапные подходы осуществляют поиск и распознавание объектов в два этапа: 1) формируют множество потенциальных регионов; 2) классификация каждого региона. Одноэтапные подходы объединяют эти два этапа в один и напрямую обнаруживают и распознают объекты. В свою очередь двухэтапные подходы обеспечивают большую точность поиска и распознавания по сравнению с одноэтапными подходами. Однако, «платой» за точность поиска и распознавания является большая вычислительная сложность алгоритма.

Поэтому в распознающих системах для автономных носителей [1] используются одноэтапные СНС, которые могут обрабатывать данные в масштабе реального времени с высокой точностью распознавания на устройствах с ограниченным вычислительным ресурсом.

Входными данными для СНС могут быть изображения, последовательность изображений (видеокадров), n-мерные тензоры (батчи), пирамиды изображений. Современные одноэтапные СНС состоят из экстрактора признаков (Backbone) и блока декодирования (Head) (рис. 1). Самой вычислительно затратной составной частью алгоритма является экстрактор признаков, поэтому прямое распространение данных через экстрактор при-

знаков осуществляется на нейронном ускорителе. К числу нейронных ускорителей относятся такие устройства как: графический процессор или Graphics Processing Unit (GPU); нейронный процессор или Neural Processing Unit (NPU); тензорный процессор или Tensor Processing Unit (TPU); программируемая логическая интегральная схема (ПЛИС) или Field Programmable Gate Array (FPGA) и др.

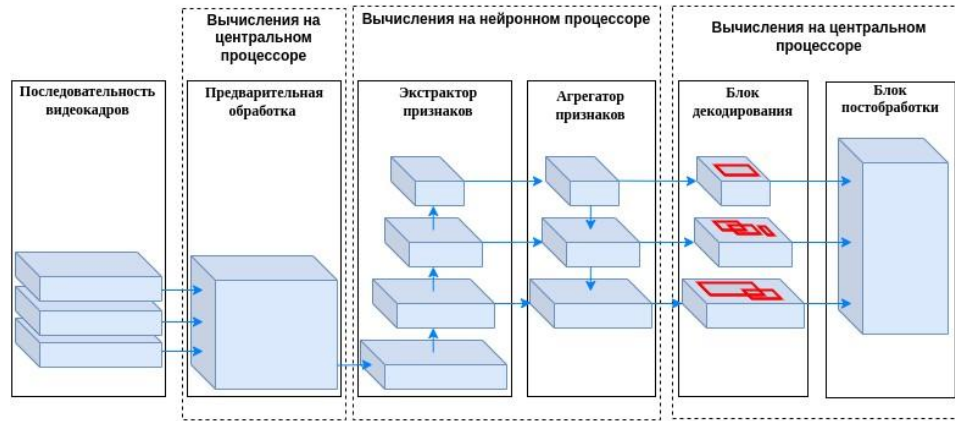


Рис. 1. Структурная схема алгоритма поиска и распознавания объектов на изображениях для устройств с ограниченным вычислительным ресурсом

Современные архитектуры СНС используют промежуточные слои между экстрактором признаков и блоком декодирования для того, чтобы собрать семантические информации с различных масштабов пирамиды признаков. Такое слияние признаков называют агрегированием или Neck. Агрегирование признаков может быть вычислительно затратным, поэтому этот этап алгоритма тоже выполняется на нейронном ускорителе. Декодирование агрегированных признаков осуществляется одноэтапными алгоритмами таким как: YOLO [2–5], SSD [6], CenterNet [7], CornerNet [8], FCOS [9] и др. Одной из проблем алгоритмов поиска и распознавания объектов является предсказание двойных обрамляющих прямоугольников для одного объекта на изображении. Для устранения данной проблем применяют алгоритм постобработки Non Maximum Suppression (NMS) [10]. Декодирование и постобработка не являются вычислительно затратными этапами алгоритма, поэтому выполняется на центральном процессоре.

Поиск и распознавания малоразмерных объектов на изображениях вызывает трудности даже у передовых архитектур СНС из-за проблемы переобучения. Поэтому применяют методы регуляризации, которые направлены на снижение вероятности переобучения СНС за счёт расширения исходного признакового пространства. Обычно, изображения видимого оптического диапазона расширяют данными различных типов (мультимодальные изображения): инфракрасные (ИК) изображения, данные с лазерных дальномеров, радиолокационные изображения, изображения с признаком движения и др. Расширение признакового пространства радиолокационными изображениями, изображениями с лазерных дальномеров и ИК изображениями имеет существенные недостатки, заключающиеся в применении дорогостоящей сенсорной аппаратуры и усложнении конструкции распознающей системы. В свою очередь, изображения с признаками движения можно получить алгоритмами предварительной обработки последовательности изображений [11–14]. Однако, обработка изображений высокого разрешения на центральном процессоре может приводить к неприемлемой временной задержке и являться «узким местом» всего алгоритма. Поэтому предложен алгоритм предварительной обработки последовательности изображений на нейронном ускорителе для снижения временной задержки.

Алгоритм предварительной обработки последовательности изображений для расширения признакового пространства. Алгоритм предварительной обработки последовательности видеокладов направлен на расширения исходного признакового про-

странства признаками движения, полученными в процессе длительного наблюдения за объектом. На вход алгоритма поступает последовательность изображений, которая накапливается в линии задержке с отводами. Период выдачи кадров камерой определяется временной задержкой T . Накопленная последовательность видеокадров поступает в блок идентификации движения. С учётом компромисса между точностью и вычислительной сложностью, алгоритм идентификации движения основан на построении разностных изображений между опорным изображением и задержанными изображениями [15]:

$$d_i(x, y) = |f(x, y, t_0) - f(x, y, t_{i+1})|, \quad (1)$$

где i есть индекс разностного изображения, который находится в диапазон от 0 до $k - 1$, где k длина очереди видеокадров; $f(x, y, t_0)$ и $f(x, y, t_{i+1})$ – дискретные изображения, полученные в моменты времени t_0, t_{i+1} ; x, y – пространственные координаты. Изображение, полученное в момент времени t_0 , является опорным изображением, а в момент времени t_{i+1} является задержанным. Разностные изображения d_i формируются путем сравнения опорного изображения с задержанным изображением. Таким образом, разностное изображение, полученное в начальные моменты времени, отображает наличие медленного движения в видеокадре, а в позднее моменты времени отображает быстрое наличие движения. Взятие абсолютных значений разностных изображений позволяет привести значения яркостей изображения к одному динамическому диапазону от нуля до единицы во всех каналах комплексированного изображения.

Результирующее псевдоцветное изображение формируется на основе поканального объединения на уровне пикселей опорного изображения и k разностных изображений для заданных моментов времени. Таким образом, количество каналов разностных изображений и время задержки кадров не фиксированы и задаются разработчиком. Конечным результатом разработанного алгоритма предварительной обработки изображений является комплексированное изображение характерное «цветными» локальными областями, означающими признак движения.

Алгоритм предварительной обработки последовательности изображений для расширения признакового пространства на нейронном ускорителе. Концепция построения алгоритмов СНС на устройствах с ограниченным вычислительным ресурсом заключается в переносе вычислительно сложных этапов алгоритма с центрального процессора на нейронный ускоритель. Это достигается за счет объединения вычислений предварительной обработки изображений с экстрактором признаков СНС и выполнения на нейронном ускорителе. Структурная схема алгоритма поиска и распознавания объектов на изображениях с учётом применения алгоритма предварительной обработки последовательности изображений на нейронном ускорителе измениться согласно структурной схеме, представленной на рис. 2.

Согласно структурной схеме представленной на рис. 2, необходимо объединить последовательность операций, которые включает алгоритм предварительной обработки с последовательностью операций экстрактора признаков. Для реализации данного подхода необходимо представить вычисления алгоритма предварительной обработки в рамках операций поддерживаемых нейронными ускорителями. К числу операций, которые включает алгоритм предварительной обработки изображений для расширения признакового пространства, относятся: поэлементная (попиксельная) матричная разность и вычисление абсолютных значений каждого элемента матрицы.

Алгоритм использует опорное изображение и два изображения, задержанных на T и $4T$. С учётом выбранных задержек, для получения комплексированного изображения необходимо накопить последовательность из пяти изображений. На основе накопленной последовательности формируется тензор из изображений, полученных в настоящий момент времени, T и $4T$. Формирование опорного и двух разностных изображений реализовано с помощью операции двумерная свёртка (2D Convolution) размерностью фильтров $3 \times 3 \times 1 \times 1$, со значениями 0, 1, –1. Единичные значения фильтров формируют опорное изображение, отрицательные значения формируют изображение с отрицательными значениями, нулевые значения удаляют все признаки изображения. Фильтр с очередностью

значений 1, 0, 0 формирует опорное изображение. Очередность значений ядер фильтра 1, -1, 0 формирует разностное изображение, характеризующее медленное наличие движения, а фильтр с значениями 1, 0, -1 формирует разностное изображение, характеризующее быстрое наличие движения. Таким образом, на выходе свёртки формируется тензор равный по размерности входному тензору.

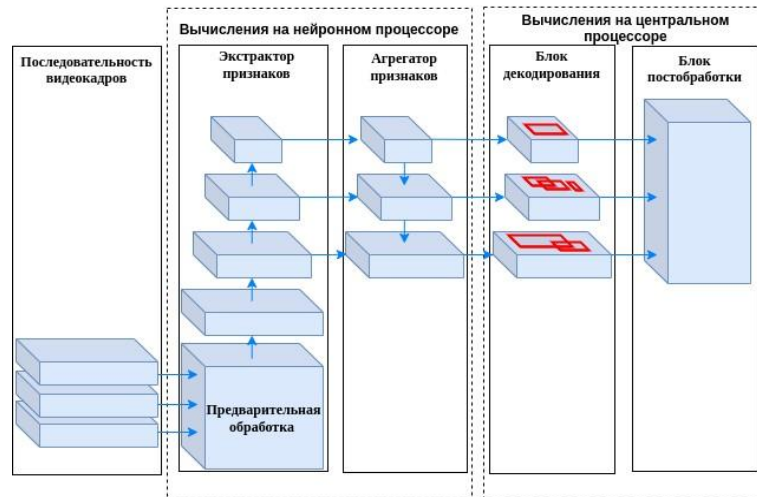


Рис. 2. Структурная схема алгоритма поиска и распознавания объектов с учётом применения алгоритма предварительной обработки последовательности изображений на нейронном ускорителе

Опорное изображение имеет только положительные значения, а значения в разностных каналах могут иметь отрицательные значения. Чтобы привести значения выходного тензора к одному динамическому диапазону от 0 до 1 во всех каналах определяются абсолютные значения всех элементов тензора. Результатом всех операций является тензор, состоящего из опорного и двух разностных изображений. Если представить тензор в цветовом пространстве Red Green Blue (RGB), то опорное изображение находится в цветовом канале Red, а разностные изображения в цветовых каналах Green и Blue.

Экспериментальное исследование алгоритма предварительной обработки последовательности изображений для расширения признакового пространства на нейронном ускорителе. В экспериментальной части данной главы произведено сравнение эффективности предложенного алгоритма путем сравнения времени выполнения алгоритма на нейронном ускорителе и центральном процессоре. Для проведения эксперимента сформировано множество вычислительных устройств с аппаратной поддержкой нейронного ускорителя: NVIDIA Jetson Nano, NVIDIA Jetson TX2, NVIDIA Jetson AGX Xavier и Rockchip RK3588.

Построение алгоритма на вычислительных устройствах NVIDIA Jetson.

Построение алгоритма на вычислительных устройствах NVIDIA Jetson. Для построения алгоритма комплексирования информации о движении на устройствах линейки Jetson используется среда разработки JetPack. Для выполнения алгоритма расширения признакового пространства на центральном процессоре используется фреймворк Pytorch, который реализует CPU-модель алгоритма. GPU-модели алгоритма построены с помощью фреймворка TensorRT.

Построение алгоритма на вычислительном устройстве Rockchip.

В необходимый набор программных средств для построения NPU-модели алгоритма предварительной обработки изображений для расширения признакового пространства на вычислительном устройстве Rockchip RK3588 входят ОС Linux Ubuntu SDK (Software Development Kit) Rockchip Neural Network-Toolkit2 (RKNN-Toolkit2) и RKNN-Toolkit-

Lite2. RKNN-Toolkit2 — это комплект средств разработки программного обеспечения, позволяющий выполнять преобразование моделей, прямой вывод и оценку производительности на персональных компьютерах и вычислительных устройствах Rockchip. RKNN-Toolkit-Lite2 предоставляет интерфейсы программирования Python для платформы Rockchip, помогающие строить NPU-модели и ускорять выполнение.

Оценка времени выполнения алгоритма предварительной обработки изображений на устройствах производилась в режиме максимально допустимой вычислительной мощности для всех устройств: устройства линейки Jetson в режиме *jetson_clocks*, устройства Rockchip RK3588 режим загрузки всех вычислительных ядер нейронного процессора.

Оценка времени выполнения Pytorch CPU-моделей алгоритма предварительной обработки изображений для всех вычислительных устройств оценивается с помощью профилировщика *profile* входящего в фреймворк Pytorch.

На каждом вычислительном устройстве построены модели алгоритма предварительной обработки изображений для расширения признакового пространства для всех поддерживаемых типов разрядности вычислений. После чего произведена оценка времени выполнения построенных моделей алгоритма для размера входных данных $1 \times 3 \times 960 \times 960$ элементов, которые отображены в табл. 1. Переход к вычислениям с пониженной разрядностью приводит к снижению точности поиска и распознавания объектов на изображениях [16, 17]. Поэтому произведём вычислительный эксперимент влияния перехода к вычислениям с пониженной разрядностью.

Для оценки точности поиска и распознавания малоразмерных объектов сформирован датасет синтетических изображений с аннотациями с помощью средств виртуального моделирования 3D графика Unreal Engine 5. Датасет состоит из аннотированных условно-реальных изображений и включает 5 классов техники. Датасет сформирован из 80 видео сюжетов видимого оптического диапазона различной местности. Из первых 50 сюжетов получена тренировочная выборка, а из остальных 30 тестовая выборка. В тренировочную выборку входят 50000, а в тестовую 30000 аннотированных изображений. На изображениях присутствуют не менее одного малоразмерного объекта.

Таблица 1

Оценка времени выполнения моделей алгоритма предварительной обработки изображений для расширения признакового пространства на вычислительных устройствах

Вычислительное устройство	Нейронный ускоритель	Точность вычислений	Время выполнения, мс
NVIDIA Jetson Nano B01	CPU	FP32	26.1
	GPU	FP32	18.2
		FP16	12.4
NVIDIA Jetson TX2 8GB	CPU	FP32	23.5
	GPU	FP32	10.9
		FP16	7.4
NVIDIA Jetson AGX Xavier	CPU	FP32	8.4
	GPU	FP32	3.1
		FP16	2.5
		INT8	2.1
Rockchip RK3588	CPU	FP32	16.4
	NPU	FP32	15.9
		FP16	11.2

Анализ научных работ в задаче обнаружения малоразмерных объектов показал, что нет общепринятого критерия малоразмерности объектов. В исследовании [18] авторы выбрали критерий малоразмерности для объектов, физический размер которых не превышал 0.3 м, а относительное перекрытие площади обрамляющего прямоугольника объекта с площадью изображения находилось в пределах от 0.08 до 0.58. В вычислительном эксперименте используется это условие только для объектов с большим физическим размером (до 6 м).

Произведена оценка точности поиска и распознавания малоразмерных объектов моделей СНС на тестовом датасете на основе метрики Mean Average Precision (mAP) [19, 20]. Метрики mAP получены для входных данных изображений размерностью 3 x 960 x 960 и представлены в табл. 2. Исходя из полученных оценок времени выполнения, можно сделать, что предложенный алгоритм позволяет увеличить скорость обработки данных предварительной обработки в 1.4–4 раз в зависимости типа нейронного ускорителя и разрядности вычислений (битности). Однако, переход к вычислениям с целочисленным типом разрядности (integer 8) приводит к снижению интегрального показателя точности поиска и распознавания Mean Average Precision на 5-19.4%.

Таблица 2

**Оценка точности поиска и распознавания малоразмерных объектов СНС
для вычислений с различной точностью**

Архитектура СНС	Нейронный ускоритель	Точность вычислений	mAP, %
YOLOv3-Tiny-M	GPU	FP32	66.67
		FP16	66.79
		INT8	49.732
	NPU	FP32	66.67
		INT8	52.8
YOLOv5s-M	GPU	FP32	73.29
		FP16	73.29
		INT8	54.15
	NPU	FP32	73.29
		INT8	63.4
YOLOv5l-M	GPU	FP32	86.77
		FP16	86.77
		INT8	80.3
	NPU	FP32	86.77
		INT8	81.7
YOLOv8s-M	GPU	FP32	73.27
		FP16	73.3
		INT8	65.1
	NPU	FP32	73.27
		INT8	64.3

Заключение. Произведено экспериментальное исследование времени выполнения алгоритма предварительной обработки последовательности видеок кадров для расширения признакового пространства на нейронном процессоре. На основе полученных оценок

времени выполнения алгоритма можно сделать выводы, что применение предложенного подхода позволило сократить время выполнения алгоритма в 1.4-4 раз в зависимости от типа нейронного ускорителя и разрядности (битности) вычислений по сравнению с центральным процессором. Переход к вычислениям с половинной разрядностью (float pointing 16) не изменяет точность поиска и распознавания объектов, а переход к целочисленной разрядности вычислений (integer 8) приводит к снижению критерия Mean Average Precision на 5-19.4 %, характеризующего интегральную среднюю точность.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Алпатов Б. и др. Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление. – М.: Радиотехника, 2008. – 176 с.
2. Redmon J., Farhadi A. YOLO9000: better, faster, stronger // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 7263-7271.
3. Redmon J., Farhadi A. Yolov3: An incremental improvement // arXiv preprint arXiv:1804.02767. – 2018.
4. Bochkovskiy A., Wang C.-Y., Liao H.-Y.M. Yolov4: Optimal speed and accuracy of object detection // arXiv preprint arXiv:2004.10934. – 2020.
5. Wang C.-Y., Bochkovskiy A., Liao H.-Y.M. Scaled-yolov4: Scaling cross stage partial network // Proceedings of the IEEE/cvf conference on computer vision and pattern recognition. – 2021. – P. 13029-13038.
6. Liu W., et al. SSD: Single shot multibox detector // Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. – Springer, 2016. – P. 21-37.
7. Duan K., et al. Centernet: Keypoint triplets for object detection // Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – P. 6569-6578.
8. Law H., Deng J. Cornernet: Detecting objects as paired keypoints // Proceedings of the European conference on computer vision (ECCV). – 2018. – P. 734-750.
9. Tian Z., et al. Fcos: Fully convolutional one-stage object detection // Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – P. 9627-9636.
10. Girshick R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2014. – P. 580-587.
11. Niu Y., et al. Airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform // Mathematical Problems in Engineering. – 2012. – T. 2012.
12. Liu S., Liu Z. Multi-channel CNN-based object detection for enhanced situation awareness // arXiv preprint arXiv:1712.00075. – 2017.
13. Smeelen M.A., et al. Semi-hidden target recognition in gated viewer images fused with thermal IR images // Information Fusion. – 2014. – Vol. 18. – P. 131-147.
14. Li Y., et al. Multimodal medical supervised image fusion method by CNN // Frontiers in Neuroscience. – 2021. – Vol. 15. – P. 638976.
15. Ковалев В.В. Методы решения проблемы переобучения нейронных сетей в задаче обнаружения малоразмерных объектов на изображениях // Тр. международного научно-технического конгресса «Интеллектуальные системы и информационные технологии-2023». – 2023. – P. 38-45.
16. Ковалев В.В., Сергеев Н.Е. Реализация сверточных нейронных сетей на встраиваемых устройствах с ограниченным вычислительным ресурсом // Известия ЮФУ. Технические науки. – 2021. – № 6 (223). – С. 64-72.
17. Щелкунов А.Е. и др. Ускорение прямого прохода при реализации СНС на ограниченном вычислительном ресурсе // Известия ЮФУ. Технические науки. – 2022. – № 1. – С. 289-297.
18. Chen C., et al. R-CNN for small object detection // Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13. – Springer, 2017. – P. 214-230.
19. Everingham M., et al. The pascal visual object classes (voc) challenge // International journal of computer vision. – 2010. – Vol. 88. – P. 303-338.
20. Padilla R., Netto S.L., Da Silva E.A. A survey on performance metrics for object-detection algorithms // 2020 international conference on systems, signals and image processing (IWSSIP). – IEEE, 2020. – P. 237-242.

REFERENCES

1. *Alpatov B. i dr. Metody avtomaticheskogo obnaruzheniya i soprovozhdeniya ob"ektov. Obrabotka izobrazheniy i upravlenie* [Methods of automatic detection and tracking of objects. Image processing and control]. Moscow: Radiotekhnika, 2008, 176 p.
2. *Redmon J., Farhadi A. YOLO9000: better, faster, stronger, Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263-7271.
3. *Redmon J., Farhadi A. Yolov3: An incremental improvement, arXiv preprint arXiv:1804.02767*, 2018.
4. *Bochkovskiy A., Wang C.-Y., Liao H.-Y.M. Yolov4: Optimal speed and accuracy of object detection, arXiv preprint arXiv:2004.10934*, 2020.
5. *Wang C.-Y., Bochkovskiy A., Liao H.-Y.M. Scaled-yolov4: Scaling cross stage partial network, Proceedings of the IEEE/cvf conference on computer vision and pattern recognition*, 2021, pp. 13029-13038.
6. *Liu W., et al. SSD: Single shot multibox detector, Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21-37.
7. *Duan K., et al. Centernet: Keypoint triplets for object detection, Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 6569-6578.
8. *Law H., Deng J. Cornernet: Detecting objects as paired keypoints, Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 734-750.
9. *Tian Z., et al. Fcos: Fully convolutional one-stage object detection, Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 9627-9636.
10. *Girshick R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation, Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.
11. *Niu Y., et al. Airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform, Mathematical Problems in Engineering*, 2012, T. 2012.
12. *Liu S., Liu Z. Multi-channel CNN-based object detection for enhanced situation awareness, arXiv preprint arXiv:1712.00075*, 2017.
13. *Smeelen M.A., et al. Semi-hidden target recognition in gated viewer images fused with thermal IR images, Information Fusion*, 2014, Vol. 18, pp. 131-147.
14. *Li Y., et al. Multimodal medical supervised image fusion method by CNN, Frontiers in Neuroscience*, 2021, Vol. 15, pp. 638976.
15. *Kovalev V.V. Metody resheniya problemy pereobucheniya neyronnykh setey v zadache obnaruzheniya malorazmernykh ob"ektov na izobrazheniyakh* [Methods for solving the problem of overtraining neural networks in the task of detecting small-sized objects in images], *Tr. mezhdunarodnogo nauchno-tekhnicheskogo kongressa «Intellectual'nye sistemy i informatsionnye tekhnologii-2023»* [Proceedings of the international scientific and technical congress "Intelligent systems and information technologies-2023"], 2023, pp. 38-45.
16. *Kovalev V.V., Sergeev N.E. Realizatsiya svertochnykh neyronnykh setey na vstraivaemykh ustroystvakh s ogranichenym vychislitel'nym resursom* [Implementation of convolutional neural networks on embedded devices with limited computing resources], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2021, No. 6 (223), pp. 64-72.
17. *Shchelkunov A.E. i dr. Uskorenie pryamogo prokhoda pri realizatsii SNS na ogranichenom vychislitel'nom resurse* [Acceleration of the forward pass when implementing a CNN on a limited computing resource] *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 1, pp. 289-297.
18. *Chen C., et al. R-CNN for small object detection, Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13*. Springer, 2017, pp. 214-230.
19. *Everingham M., et al. The pascal visual object classes (voc) challenge, International journal of computer vision*, 2010, Vol. 88, pp. 303-338.
20. *Padilla R., Netto S.L., Da Silva E.A. A survey on performance metrics for object-detection algorithms, 2020 international conference on systems, signals and image processing (IWSSIP)*. IEEE, 2020, pp. 237-242.

Статью рекомендовал к опубликованию д.т.н., профессор Ю.А. Кравченко.

Ковалев Владислав Владимирович – Южный федеральный университет; e-mail: vlad.kovalev94@mail.ru; г. Таганрог, Россия; тел.: +79525864492; кафедра вычислительной техники; соискатель.

Kovalev Vladislav Vladimirovich – Southern Federal University; e-mail: vlad.kovalev94@mail.ru; Taganrog, Russia; phone: +79525864492; the Department of Computer Science; post-graduate student.