

Исследование композиции алгоритмов компьютерного зрения, базирующихся на машинном обучении, для решения задачи локализации и классификации объектов*

А.В. ШЕШКУС^{I,II}, Д.П. МАТАЛОВ^{II,III}, В.В. АРЛАЗАРОВ^{I,II,III,IV}, Д.П. НИКОЛАЕВ^{II,III,IV}

^I Институт системного анализа Федерального исследовательского центра «Информатика и управление» РАН, Москва, Россия

^{II} ООО «Смарт Энджинс Сервис», Москва, Россия

^{III} Московский физико-технический институт, г. Долгопрудный, Россия

^{IV} Институт проблем передачи информации им. А.А. Харкевича РАН, Москва, Россия

Аннотация. В данной работе исследуется подход к построению композиции алгоритмов машинного обучения для решения задачи локализации и классификации объектов на изображении. Приводятся подробные описания и анализ таких алгоритмов машинного обучения как метод Виолы-Джонса и метод классификации изображений с помощью сверточных нейронных сетей. Анализ экспериментальных данных показывает применимость описываемого подхода для решения сложных задач распознавания объектов на изображении.

Ключевые слова: глубокое обучение, метод Виолы-Джонса, машинное зрение, локализация объектов, классификация объектов, сверточные нейронные сети, композиция алгоритмов машинного обучения.

DOI: 10.14357/20790279190103

Введение

Одной из важнейших задач компьютерного зрения является проблема локализации и классификации объектов на изображении. В обзорной работе по детектированию объектов [1] авторы разделяют методы их локализации на изображениях на 4 обширных категории: 1) методы, базирующиеся на сопоставлении с шаблоном локализуемого объекта (для каждого целевого объекта генерируется шаблон, который затем используется для вычисления некоторого функционала «похожести» для каждого возможного положения локализуемого объекта на изображении) [2, 3]; 2) методы, основанные на некоторой априорной информации о целевых объектах (рассматривают задачу детектирования объектов как задачу проверки множества гипотез, построенных по целевому объекту) [4]; 3) object-based image analysis, который получил широкое применение в задачах анализа спутниковых снимков (большинство таких алгоритмов содержат 2 шага: сегментация изображения и класси-

фикация сегментов) [5]; 4) методы, построенные на машинном обучении, большинство которых рассматривают детектирование объектов как задачу классификации регионов изображения [6-9]. В настоящее время внимание научного сообщества сконцентрировано именно на методах, базирующихся на алгоритмах машинного обучения. Привлекательность этого подхода очевидна: требуемый результат может быть достигнут при достаточном количестве обучающих прецедентов без разработки узкоспециализированных алгоритмов.

Несмотря на то, что существует огромное количество различных методов машинного обучения, в промышленных распознающих системах наиболее часто встречаются следующие подходы: метод поиска объектов Виолы-Джонса [10] и сверточные нейронные сети [11]. Столь широкое применение данные алгоритмы получили благодаря универсальности и способности решать задачи широкой предметной области.

В данной работе исследуется подход к построению распознающих алгоритмов, представляющих собой композиции классификаторов Ви-

* Работа выполнена при частичной поддержке РФФИ, проекты № 18-07-01384, 18-07-01387.

олы-Джонса и сверточных нейронных сетей для решения сложных задач распознавания объектов на изображении.

1. Метод Виолы-Джонса

Метод Виолы и Джонса [10] представляет из себя схему построения детекторов ригидных объектов статистическим образом (опираясь на прецеденты). Изначально метод был применен для построения детектора человеческого лица в анфас.

В качестве признакового пространства в оригинальной версии метода используются признаки Хаара, значение которых вычисляется как разность сумм яркостей пикселей областей изображения внутри черных и белых прямоугольников. Для эффективного вычисления значений признаков используется интегральное представление изображения.

С каждым признаком связывается бинарный «слабый» классификатор $h(x): X \rightarrow \{-1, +1\}$, представленный в виде распознающего дерева с одним ветвлением. Такие классификаторы демонстрируют слабую распознающую силу, поэтому в методе Виолы и Джонса используется алгоритм AdaBoost для построения «сильного» классификатора в виде линейной комбинации слабых классификаторов:

$$S(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h(x)_t > 0 \right).$$

Для обеспечения высокой производительности детекторов объектов в методе Виолы и Джонса используется каскад сильных классификаторов, который позволяет на «ранних этапах» распознавать заведомо «пустые» участки изображения. Классический каскад сильных классификаторов представим как конъюнкция ответов сильных классификаторов:

$$\text{Cascade}(x) = \prod_{i=1}^N S(x)_i.$$

Существует множество работ, в которых авторы предлагают существенные модификации и расширения подхода Виолы-Джонса [12, 13]. В том числе были предприняты попытки по локализации объектов различных классов с помощью модификации метода Виолы-Джонса [14].

В настоящее время метод Виолы-Джонса стал одним из наиболее популярных способов локализации объектов. При этом его использование вышло далеко за пределы оригинальной области применимости [15, 16].

2. Сверточные нейронные сети

В настоящее время наиболее широко используемым методом машинного обучения для решения широкого класса задач анализа изображений является применение сверточных нейронных сетей. Существует множество работ [17, 18], авторы которых предпринимают усилия по разработке архитектур нейронных сетей способных «за один проход» вернуть параметры обрамляющего объект прямоугольника на изображении, а также метку, позволяющую идентифицировать класс объекта. Однако применение архитектур нейронных сетей, обладающих таким количеством параметров, неприменимо для решения задач в реальном времени на таких маломощных устройствах как смартфоны и портативные компьютеры. Существует большое количество работ, авторы которых предлагают различную технику для увеличения быстродействия сверточных нейронных сетей. Так, одна из них – аппроксимация 32-х битных вещественных вычислений с помощью 8-ми битной арифметики с фиксированной точкой [19]. Также, предлагались и более экзотические методы: в работе [20] авторы предлагают декомпозицию тензора сверточных слоев, а в работе [21] рассматривается применимость каскада сверточных нейронных сетей.

Хорошо известно, что метод Виолы-Джонса обеспечивает быстрое нахождение объекта на изображении, а алгоритмы, основанные на использовании сверточных нейронных сетей, отлично справляются с задачами классификации объектов на изображении. Следовательно, разумно построить такой бинарный классификатор Виолы-Джонса, который будет способен обобщить несколько похожих классов в один, а затем, с помощью сверточной нейронной сети произвести классификацию ответа детектора, которая позволит получить метку класса целевого объекта и отфильтровать ложные срабатывания детектора.

3. Постановка задачи

Необходимо локализовать и идентифицировать на изображении сложный композитный объект. В качестве ответа выступают параметры обрамляющего целевой объект прямоугольника и метка класса. В качестве задачи, удовлетворяющей описанным выше условиям, может выступить задача локализации и идентификации предъявляемой проекции танка. Танк можно разделить на 3 части: тело танка, башня и пушка. Все 3 части вращаются вокруг одной оси, но несогласованно (обычно только вращение башни согласовано с вращением пушки).

Данная особенность не присуща целевым объектам, рассмотренным в оригинальной работе Виолы и Джонса. Следует также отметить, что идентификация предъявляемой камере проекции является трудноразрешимой задачей даже для человека, в чем можно убедиться, посмотрев на рис. 2.

4. Постановка эксперимента

В данной работе рассматриваются алгоритмы, основанные на машинном обучении. Одной из ключевых проблем машинного обучения, является необходимость наличия достаточного количества обучаемых данных. Очевидно, что натуральные изображения, полученные с реальных сцен, содержащие необходимые нам объекты, нам недоступны. Поэтому, для моделирования реальных сцен был подготовлен специальный макет. На макете установлены различные объекты, моделирующие сельскую местность: характерное ландшафтное покрытие, кусты, деревья, заграждения и т.д. Изображения захватывались с помощью цифровой малоформатной камеры (рис. 1).

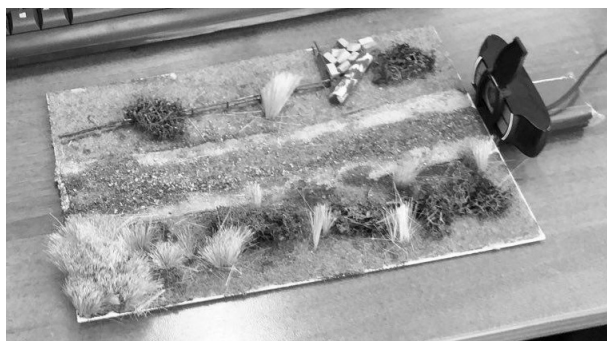


Рис. 1. Макет и регистрирующее устройство

В процессе захвата изображений существенно менялся задний план сцены для обеспечения большей устойчивости алгоритмов к изменениям фона. Объектами локализации и классификации выступили 4 модели боевых танков: Т-90 (Россия), М1А2 Абрамс (США), Т-14 (Россия), Меркава III (Израиль). На рис. 2 представлены примеры полученных изображений.

Для постановки эксперимента были подготовлены два набора изображений. Наборы изображений и идеальные ответы системы на каждое из изображений были отобраны и подготовлены экспертом. Для каждого изображения в данных наборах представлен идеальный ответ алгоритма – параметры обрамляющего объекта прямоугольника, а также его класс. Также известно, что на изображении представлено не более одного объекта



Рис. 2. Примеры полученных изображений

интереса, и если он присутствует, то полностью находится в плоскости изображения. На выходе алгоритма ожидаются параметры обрамляющего прямоугольника (координаты x и y на плоскости изображения, а также его ширина и высота) и метка класса. Расположение объекта считается корректно найденным, если индекс Жаккарда (отношение площади пересечения прямоугольника идеального ответа системы и выхода алгоритма) превосходит 0.5. Классификация считается корректно произведенной, если метки классов идеального ответа системы и выхода алгоритма совпадают. Метка классов однозначно кодирует название танка, а также представленную системе проекцию (фронтальную или заднюю).

Подготовленные ранее наборы изображений были разделены на обучающий и тестовые. Обучающий набор состоял из 17000 изображений, тестовый – из 1700.

5. Экспериментальные результаты

Было произведено 2 эксперимента, в обоих есть общая часть – локализация пока еще неизвестного объекта с помощью обученного классификатора Виолы-Джонса. Для обучения «детектирующего» классификатора в качестве положительной выборки выступили изображения всех танков в обоих проекциях. В качестве негативной выборки подавались на вход регионы изображения, не содержащие целевые объекты полностью. Для расширения обучающей выборки использовалась аугментация с помощью гамма-коррекции. В первом эксперименте классификация ответа детектора производилась множеством классификаторов Виолы-Джонса. Каждый из классификаторов обучался по принципу «один против всех», т.е. в качестве положительной выборки выступает множество изображений, принадлежащих целевому классу (конкретная модель танка, конкретная проекция), а в каче-

Табл. 1

Архитектура классифицирующей сети

Номер слоя	Тип	Входная размерность	Описание
1	Convolution	77x53x1	Filter size: 5x5, filters count: 6, no padding
2	Pooling	73x49x6	Max, window size: 3x3, stride: 2x2, with zero padding
3	Neuron	37x25x6	Bounded ReLU: [-1, 1]
4	Convolution	37x25x6	Filter size: 3x3, filters count: 6, with zero padding
5	Pooling	37x25x6	Max, window size: 3x3, stride: 2x2, with zero padding
6	Neuron	19x13x6	Bounded ReLU: [-1, 1]
7	Convolution	19x13x6	Filter size: 5x5, filters count: 6, with zero padding
8	Neuron	19x13x6	Bounded ReLU: [-1, 1]
9	Convolution	19x13x6	Filter size: 3x3, filters count: 4, no padding
10	Neuron	17x11x4	Bounded ReLU: [-1, 1]
11	Fully connected	17x11x4	Outputs: 8
12	Softmax	1x1x8	-

стве негативной – изображения других классов. Обычно в промышленных системах единичные положительные ответы классификатора трактуются ложными. Поэтому для большей устойчивости к данному типу ошибок регион изображения объекта масштабируется не к входному размеру классификатора, а к изоморфно расширенному. В нашем эксперименте ответ детектора расширился на 10%. Затем на расширенном регионе изображения запускается каждый классификатор методом скользящего окна. Это позволяет ввести меру уверенности ответа классификатора каждого класса: возвращается тот класс, классификатор которого вернул наибольшее количество положительных ответов. Во втором эксперименте в качестве классификатора выступает сверточная нейронная сеть. В табл. 1 представлена архитектура используемой сети.

Обучающее множество расширилось с помощью аугментации гауссовым шумом, гауссовым размытием и линейного преобразования яркости [22]. Обученная нейронная сеть имеет около 7500 параметров, и ее можно назвать «легкой», т.к. количество обучаемых параметров в ней существенно меньше, чем в других архитектурах нейронных сетей, широко используемых для задач классификации [8,9,23].

Экспериментальные данные показывают, что классификация с помощью СНС (сверточной нейронной сети) существенно медленнее, чем описанная ранее процедура классификации с помощью множества классификаторов Виолы-Джонса. Следует отметить, что в последнем эксперименте качество классификации превосходит качество ло-

кализации. Данный феномен объясняется тем, что даже если предикат корректного нахождения параметров обрамляющего прямоугольника не истинен (индекс Жакарда менее 0.5), нейронная сеть корректно классифицировала метку. Время работы алгоритма замерялось на персональном компьютере, оборудованном процессором Intel Core i7-3770 в однопоточном режиме. Суммарное время работы обоих алгоритмов практически идентично несмотря на то, что классификация с помощью нейронной сети более ресурсоемка, чем запуск 8-ми классификаторов Виолы-Джонса, однако качество классификации с помощью СНС существенно выше. В табл. 2 приведены качественные характеристики двух алгоритмов, а на рис. 3 – скриншот демонстрационной программы.

Табл. 2

Экспериментальные данные

Параметр	Время на изображение, мс.	Качество локализации	Качество классификации
Локализация с помощью Виолы-Джонса	28.2	93.9	-
Классификация с помощью семейства классификаторов Виолы-Джонса	28.8	93.9	79.1
Классификация с помощью СНС	29.6	93.9	94.8



Рис. 3. Скриншот работы демонстрационной программы

Заключение

В данной работе произведено исследование композиции из двух наиболее популярных алгоритмов компьютерного зрения, базирующихся на машинном обучении: метод Виолы-Джонса и сверточные нейронные сети. Произведенный анализ позволил построить эффективную композицию этих алгоритмов для решения задачи локализации и классификации объектов на изображении. Также произведено сравнение качественных характеристик алгоритмов многоклассовой классификации с помощью семейства классификаторов Виолы-Джонса и сверточной нейронной сети в поставленной задаче. Экспериментальные результаты свидетельствуют о том, что объекты, обладающие схожим геометрическим описанием, которые не могут быть эффективно разделены с помощью семейства классификаторов Виолы-Джонса, могут быть хорошо разделены с помощью СНС, что свидетельствует об эффективности построенной композиции. Подготовленные в рамках эксперимента изображения, а также идеальный ответ к ним приведен в [24].

Литература

1. *Cheng G., Han J.* A survey on object detection in optical remote sensing images //ISPRS Journal of Photogrammetry and Remote Sensing. – 2016. – Vol. 117. – P. 11-28.
2. *Sirmacek B., Unsalan C.* Urban-area and building detection using SIFT keypoints and graph theory //IEEE Transactions on Geoscience and Remote Sensing. – 2009. – Vol. 47. – №. 4. – P. 1156-1167.
3. *Lin X. et al.* Semi-automatic road tracking by template matching and distance transform //Urban Remote Sensing Event, 2009 Joint. – IEEE, 2009. – P. 1-7.
4. *Baltsavias E.P.* Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards

operational systems //ISPRS journal of photogrammetry and remote sensing. – 2004. – Vol. 58. – №. 3-4. – P. 129-151.

5. *Blaschke T., Lang S., Hay G. (ed.)*. Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications. – Springer Science & Business Media. 2008.
6. *Felzenszwalb P.F., Girshick R.B., McAllester D.* Cascade object detection with deformable part models //Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. – IEEE, 2010. – P. 2241-2248.
7. *Sivic J., Zisserman A.* Video Google: A text retrieval approach to object matching in videos // Proceedings Ninth IEEE International Conference on Computer Vision. – IEEE, 2003. – P. 1470.
8. *Krizhevsky A., Sutskever I., Hinton G.E.* Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – P. 1097-1105.
9. *Simonyan K., Zisserman A.* Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.
10. *Viola P., Jones M.* Robust Real-time Object Detection // International Journal of Computer Vision. 2002.
11. *LeCun Y. et al.* Object recognition with gradient-based learning //Shape, contour and grouping in computer vision. – Springer, Berlin, Heidelberg, 1999. – P. 319-345.
12. *Dollár P. et al.* Integral channel features. – 2009.
13. *Kuznetsova E., Shvets E., Nikolaev D.* Viola-Jones based hybrid framework for real-time object detection in multispectral images //Eighth International Conference on Machine Vision (ICMV 2015). – International Society for Optics and Photonics, 2015. – Vol. 9875. – P. 98750N.
14. *Isukapalli R., Elgammal A., Greiner R.* Learning to detect objects of many classes using binary classifiers //Computer Vision–ECCV 2006. – Springer, Berlin, Heidelberg, 2006. – P. 352-364.
15. *Usilin S. et al.* Visual appearance based document image classification //Image Processing (ICIP), 2010 17th IEEE International Conference on. – IEEE, 2010. – P. 2133-2136.
16. *Арлазаров, В.В., Маталов Д.П., Усилин С.А.* Локализация образа печати на документе, удостоверяющем личность, методом машинного обучения // Труды ИСА РАН. Спецвыпуск. 2018. С. 158-166.
17. *Redmon J. et al.* You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 779-788.

18. *Liu W. et al.* Ssd: Single shot multibox detector //European conference on computer vision. – Springer, Cham, 2016. – P. 21-37.
19. *Ilin D. et al.* Fast integer approximations in convolutional neural networks using layer-by-layer training //Ninth International Conference on Machine Vision (ICMV 2016). – International Society for Optics and Photonics, 2017. – Vol. 10341. – P. 103410Q.
20. *Limonova E., Sheshkus A., Nikolaev D.* Computational optimization of convolutional neural networks using separated filters architecture // International Journal of Applied Engineering Research. – 2016. – Vol. 11. – №. 11. – P. 7491-7494.
21. *Li H. et al.* A convolutional neural network cascade for face detection //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2015. – P. 5325-5334.
22. *Гайер А.В., Шешкус А.В., Чернышова Ю.С.* Аугментация обучающей выборки «на лету» для обучения нейронных сетей // Труды ИСА РАН. 2018. С. 150-157.
23. *Liu W. et al.* A survey of deep neural network architectures and their applications // Neurocomputing. – 2017. – Vol. 234. – P. 11-26.
24. *Датасет локализации и распознавания танков* [Электронный ресурс] URL: ftp://vis.iitp.ru/tank_recognition (дата обращения 13.02.2019)

Шешкус Александр Владимирович. ООО «Смарт Энджинс Сервис», г. Москва, Россия. Старший научный сотрудник-программист. Количество печатных работ: 22. Область научных интересов: глубокое обучение, компьютерное зрение, проективно-инвариантная сегментация изображений.

E-mail: astdcall@gmail.com

Маталов Даниил Павлович. ООО «Смарт Энджинс Сервис», г. Москва. Студент магистратуры МФТИ (ГУ). Научный сотрудник-программист. ООО «Смарт Энджинс Сервис». Количество печатных работ: 1. Область научных интересов: обработка изображений, распознавание образов, машинное обучение.

E-mail: mataloff@gmail.com

Арлазаров Владимир Викторович. ИППИ РАН, г. Москва. Ведущий научный сотрудник. Кандидат технических наук. Количество печатных работ: 34. Область научных интересов: искусственный интеллект, машинное обучение, системы распознавания, информационные технологии, теория массового обслуживания. E-mail: vva777@gmail.com.

Николаев Дмитрий Петрович. Институт проблем передачи информации им. Харкевича РАН, г. Москва. Заместитель директора по научной работе. Количество печатных работ: 195. Область научных интересов: машинное зрение, быстрые алгоритмы обработки изображений, распознавание образов. E-mail: dimonstr@iitp.ru

The study of the ensemble of the computer vision algorithms based on machine learning for object detection and recognition

A.V. Sheshkus^{I,II}, D.P. Matalov^{II,III}, V.V. Arlazarov^{I,II,III,IV}, D.P. Nikolaev^{II,III,IV}

^I Federal Research Center “Computer Science and Control” RAS, Moscow, Russia;

^{II} Smart Engines Service LLC, Moscow, Russia;

^{III} Moscow Institute of Physics and Technology, Dolgoprudny, Russia;

^{IV} Institute for information transmission problems (Kharkevich Institute) RAS, Moscow, Russia.

Abstract. In this paper we study ensemble approach of the machine learning algorithms for object detection and classification problem. Detailed descriptions analysis of such machine learning algorithms as Viola-Jones method and image classification method using convolutional neural networks are given. The analysis of the experimental results shows applicability of the provided approach for solving complex image recognition tasks.

Keywords: *deep learning, Viola-Jones approach, machine vision, object detection, object classification, convolutional neural networks, ensemble of the machine learning algorithms*

DOI: 10.14357/20790279190103

References

1. Cheng G., Han J. A survey on object detection in optical remote sensing images //ISPRS Journal of Photogrammetry and Remote Sensing. – 2016. – Vol. 117. – P. 11-28..
2. Sirmacek B., Unsalan C. Urban-area and building detection using SIFT keypoints and graph theory //IEEE Transactions on Geoscience and Remote Sensing. – 2009. – Vol. 47. – №. 4. – P. 1156-1167.
3. Lin X. et al. Semi-automatic road tracking by template matching and distance transform //Urban Remote Sensing Event, 2009 Joint. – IEEE, 2009. – P. 1-7.
4. Baltasvias E.P. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems //ISPRS journal of photogrammetry and remote sensing. – 2004. – Vol. 58. – №. 3-4. – P. 129-151..
5. Blaschke T., Lang S., Hay G. (ed.). Object-based image analysis: spatial concepts for knowledge-driven remote sensing applications. – Springer Science & Business Media, 2008.
6. Felzenszwalb P.F., Girshick R.B., McAllester D. Cascade object detection with deformable part models //Computer vision and pattern recognition (CVPR), 2010 IEEE conference on. – IEEE, 2010. – P. 2241-2248.
7. Sivic J., Zisserman A. Video Google: A text retrieval approach to object matching in videos // Proceedings Ninth IEEE International Conference on Computer Vision. – IEEE, 2003. – P. 1470.
8. Krizhevsky A., Sutskever I., Hinton G.E. Imagenet classification with deep convolutional neural networks //Advances in neural information processing systems. – 2012. – P. 1097-1105.
9. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition //arXiv preprint arXiv:1409.1556. – 2014.
10. Viola P., Jones M. Robust Real-time Object Detection // International Journal of Computer Vision. 2002.
11. LeCun Y. et al. Object recognition with gradient-based learning //Shape, contour and grouping in computer vision. – Springer, Berlin, Heidelberg, 1999. – P. 319-345.
12. Dollár P. et al. Integral channel features. – 2009.
13. Kuznetsova E., Shvets E., Nikolaev D. Viola-Jones based hybrid framework for real-time object detection in multispectral images //Eighth International Conference on Machine Vision (ICMV 2015). – International Society for Optics and Photonics, 2015. – Vol. 9875. – P. 98750N.
14. Isukapalli R., Elgammal A., Greiner R. Learning to detect objects of many classes using binary classifiers //Computer Vision–ECCV 2006. – Springer, Berlin, Heidelberg, 2006. – P. 352-364.
15. Usilin S. et al. Visual appearance based document image classification //Image Processing (ICIP), 2010 17th IEEE International Conference on. – IEEE, 2010. – P. 2133-2136.
16. Arlazarov V.V., Matalov D.P., Usilin S.A. Lokalizatsiya obraza pechati na dokumente, udostoverayushchem lichnost', metodom mashinnogo obucheniya // Trudy ISA RAN. Spetsvypusk. 2018. P. 158-166.
17. Redmon J. et al. You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 779-788.
18. Liu W. et al. Ssd: Single shot multibox detector //European conference on computer vision. – Springer, Cham, 2016. – P. 21-37.

19. *Ilin D. et al.* Fast integer approximations in convolutional neural networks using layer-by-layer training //Ninth International Conference on Machine Vision (ICMV 2016). – International Society for Optics and Photonics, 2017. – Vol. 10341.
20. *Limonova E., Sheshkus A., Nikolaev D.* Computational optimization of convolutional neural networks using separated filters architecture // International Journal of Applied Engineering Research. – 2016. – Vol. 11. – №. 11. – P. 7491-7494.
21. *Li H. et al.* A convolutional neural network cascade for face detection //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2015. – P. 5325-5334.
22. *Gayer A.V., Sheshkus A.V., Chernyshova Y.S.* Augmentatsiya obuchayushchei vyborki «na letu» dlya obucheniya neironnykh setei // Trudy ISA RAN. Spetsvypusk. 2018. C. 150-157.
23. *Liu W. et al.* A survey of deep neural network architectures and their applications // Neurocomputing. – 2017. – Vol. 234. – P. 11-26.
24. *Dataset lokalizatsii i raspoznavaniya tankov* [Electronical Resource] URL: ftp://vis.iitp.ru/tank_recognition (accessed February 13, 2019)

Sheshkus A.V. Institute for System Analysis Federal Research Center “Computer Science and Control” of Russian Academy of Sciences. Moscow. Russia. LLC “Smart Engines Service”. Moscow. Russia. E-mail: astdcall@gmail.com

Matalov D.P. Moscow Institute of Physics and Technology (State University). Moscow. Russia. E-mail: mataloff@gmail.com

Arlazarov V.V. PhD, Lead researcher, Institute for Information Transmitting Problems (Kharkevich Institute) of Russian Academy of Science. Moscow. Russia. E-mail: vva777@gmail.com.

Nikolaev D.P. PhD, Head of the Laboratory of Vision Systems, Institute for Information Transmitting Problems (Kharkevich Institute) of Russian Academy of Science. Moscow. Russia. E-mail: dimonstr@iitp.ru