

DOI: 10.24937/2542-2324-2020-3-393-91-96  
УДК 004.032.26:[004.9:629.5]

В.А. Вялов , А.Ю. Андреев  
ФГУП «Крыловский государственный научный центр», Санкт-Петербург, Россия

## ИСПОЛЬЗОВАНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ РАСПОЗНАВАНИЯ ОПТИЧЕСКИХ ИЗОБРАЖЕНИЙ КОРАБЛЕЙ

**Объект и цель научной работы.** Объектом исследования является проблема классификации кораблей по их изображениям в видимом, инфракрасном и радиолокационном диапазонах. Цель статьи – оценка возможности использования для классификации основных современных алгоритмов и подходов, применяемых для решения задач компьютерного зрения.

**Материалы и методы.** Проанализирована эволюция компьютеров и программного обеспечения, позволяющая использовать машинное зрение и самообучающиеся компьютерные сети для распознавания кораблей по их изображениям. Изложены основные подходы развития корабельных радиолокационных систем в различных странах. Отражены два подхода при создании моделей классификаторов с использованием «готовых» сетей, обученных для классификации изображений.

**Основные результаты.** На основе анализа развития сверточных нейронных сетей выявлены основные тенденции развития машинного зрения. Отмечены наиболее перспективные методы их использования для классификации изображений кораблей. Разработаны две модели классификаторов оптических портретов кораблей и судов. Для них получены количественные оценки достигнутых точностей распознавания.

**Заключение.** Представленные материалы позволяют определить тенденции развития классификаторов портретов кораблей. Полученные результаты могут быть использованы при распознавании тепловых и радиолокационных портретов кораблей.

**Ключевые слова:** нейронная сеть, машинное зрение, обработка изображений, распознавание объектов, датасет, компьютерное зрение.

*Авторы заявляют об отсутствии возможных конфликтов интересов.*

DOI: 10.24937/2542-2324-2020-3-393-91-96  
UDC 004.032.26:[004.9:629.5]

V. Vyalov , A. Andreev  
Krylov State Research Centre, St. Petersburg, Russia

## CONVOLUTIONAL NEURAL NETWORKS FOR OPTICAL IMAGE RECOGNITION OF SHIPS

**Object and purpose of research.** This paper discusses classification of ships based on their optical, infrared and radar images to estimate the potential of modern computer-based vision algorithms and approaches towards the accomplishment of this task.

**Materials and methods.** The paper analyses the evolution of computer hardware and software that make machine vision and machine learning solutions suitable for image-based recognition of ships. It describes main approaches to shipborne

*Для цитирования:* Вялов В.А., Андреев А.Ю. Использование сверточных нейронных сетей для распознавания оптических изображений кораблей. Труды Крыловского государственного научного центра. 2020; 3(393): 91–96.

*For citations:* Vyalov V., Andreev A. Convolutional neural networks for optical image recognition of ships. Transactions of the Krylov State Research Centre. 2020; 3(393): 91–96 (in Russian).

radar developments in various countries, outlining the two ways of making the classification models based on networks “trained” for image classification.

**Main results.** Analysing the progress in convolutional neural networks, this paper outlines the main trends in machine vision developments, points out their most promising applications in ship image classification and suggests two classification models for optical ship images with quantitative estimates of recognition accuracy.

**Conclusion.** The findings of this study outline the trends in ship image classification technology and could be helpful in infrared and radar imagery recognition.

**Keywords:** neural network, machine vision, image processing, recognition of objects, dataset, computer-based vision.

*Authors declare lack of the possible conflicts of interests.*

Быстрое развитие и миниатюризация технических средств в последние десятилетия позволили существенно увеличить объем поступающей со спутников информации о состоянии поверхности Земли. Первые фотографии из космоса были получены на полигоне White Sands (США) в 1946 г. с высоты 104,6 км, на которую смогла подняться ракета А-4, разбившаяся при посадке. Первый отечественный радиолокатор бокового обзора с высоким разрешением был запущен в 1983 г. на аппарате «Космос-1500». В настоящее время только с радиолокаторами с синтезированной апертурой вокруг Земли летает более десятка космических аппаратов дистанционного зондирования поверхности [1]. Разрешение порядка 1 м позволяют получить такие спутниковые радары, как TerraSAR-X, COSMO-SkyMed 1, Lacrosse и др. Действующая в настоящее время система обнаружения судов VDS использует информацию, полученную с таких спутников, совместно с данными от системы автоматической идентификации судов AIS [2]. Эта система используется для обнаружения «темных» судов и пресечения нелегальной морской деятельности, что требует автоматизации распознавания как их радиолокационных портретов, так и фотографий в видимом диапазоне.

Нейронные сети используются для решения задач, для которых отсутствуют явные алгоритмы, однако человеческий мозг легко с ними справляется. К таким задачам относятся распознавание изображений, речи и тому подобное [3, 4]. За последние 10 лет в задачах компьютерного зрения широкое распространение получили алгоритмы, связанные с использованием сверточных нейронных сетей (Convolution Neuron Net, CNN). Так, в рамках проекта ILSVRC (ImageNet Large Scale Visual Recognition Challenge – кампания по широкомасштабному распознаванию образов в базе данных, включающей более 10 млн аннотированных изображений) различные программные продукты ежегодно соревнуются в классификации и распознавании объектов и сцен в базе данных ImageNet. В 2011 г. хорошим результатом считалась ошибка классифика-

ции в 25 %. В 2012 г. система глубокого обучения на основе сверточной нейронной сети смогла достичь ошибки 16 %, а в следующие годы ошибка упала до нескольких процентов [5]. В 2015 г. исследователи констатировали, что компьютерные программы в определенных задачах проекта превзошли человеческие способности [6].

Сегодня алгоритмы, основанные на использовании сверточных нейронных сетей, получили широкое распространение при решении различных задач, связанных с обработкой изображений. Среди них можно выделить четыре основные группы (рис. 1, см. вклейку):

- classification – классификация изображения по типу объекта, который оно содержит;
- semantic segmentation – разделение пикселей изображения по типу объекта, к которому они относятся. Если несколько объектов одного класса перекрываются, их пиксели никак не отделяются друг от друга;
- object detection – обнаружение всех объектов указанных классов и выделение охватывающей рамки для каждого из них;
- instance segmentation – определение пикселей для каждого из объектов, принадлежащих одному из указанных классов.

## Использование нейронных сетей для решения задач компьютерного зрения

Computer-based vision applications of neural networks

Важной подобластью задач машинного обучения являются задачи компьютерного зрения, которые состоят в извлечении различной информации из изображений. Любое изображение в цифровом виде представляет собой трехмерную матрицу значений яркости пикселей каждого из цветовых каналов. Извлечение признаков из такой матрицы в явном виде является очень сложной задачей. Один и тот же предмет может быть изображен с разных ракурсов, в разном масштабе, при разном освещении,

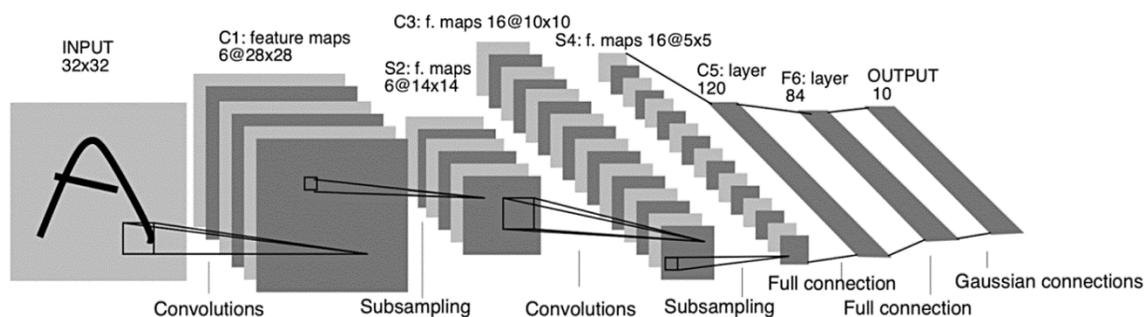


Рис. 2. Архитектура нейронной сети LeNet 5

Fig. 2. Architecture of LeNet 5 neural network

занимать разную площадь изображения или быть частично скрытым. Во всех приведенных случаях представления объекта в виде матрицы из интенсивности цветов пикселей будут очень сильно отличаться. Классический подход к решению этой задачи состоит в выделении различных признаков в явном виде (например, в виде выделения набора ребер), но он позволяет решить только весьма ограниченный круг задач.

Нейронная сеть представляет собой направленный граф из различных слоев. Пример нейронной сети LeNet 5 для распознавания рукописных цифр представлен на рис. 2.

Слово «сверточная» применительно к сети является при использовании фильтров (или ядер, как на рис. 3, см. вклейку). Свертка – это процесс применения ядра к каждому участку входного изображения. Сверточный (convolutional) слой позволяет объединять значения расположенных рядом пикселей и выделять более обобщенные признаки изображения. Для этого по картинке последовательно скользят квадратным окном небольшого размера (3×3, 5×5, 7×7 пикселей и т.п.) – ядром (kernel). Каждый элемент ядра имеет свой весовой коэффициент, умножаемый на значение того пикселя изображения, на который в данный момент наложен элемент ядра. Затем полученные для всего окна числа складываются, и из этой взвешенной суммы получается значение очередного признака. Для получения матрицы («карты») признаков всего изображения ядро последовательно сдвигается по горизонтали и вертикали (рис. 3).

Изображение или карты признаков в рамках одного слоя могут сканироваться не одним, а несколькими независимыми ядрами, таким образом, давая на выход не одну карту, а несколько. Настройка весов каждого ядра происходит при помощи решения оптимизационной задачи.

Вторым после свертки (convolution) распространенным преобразованием является слой Max-Pooling (операция подвыборки), который из группы признаков выбирает максимальное значение (рис. 4, см. вклейку). В отличие от нее, MaxPooling обычно применяется к непересекающимся группам.

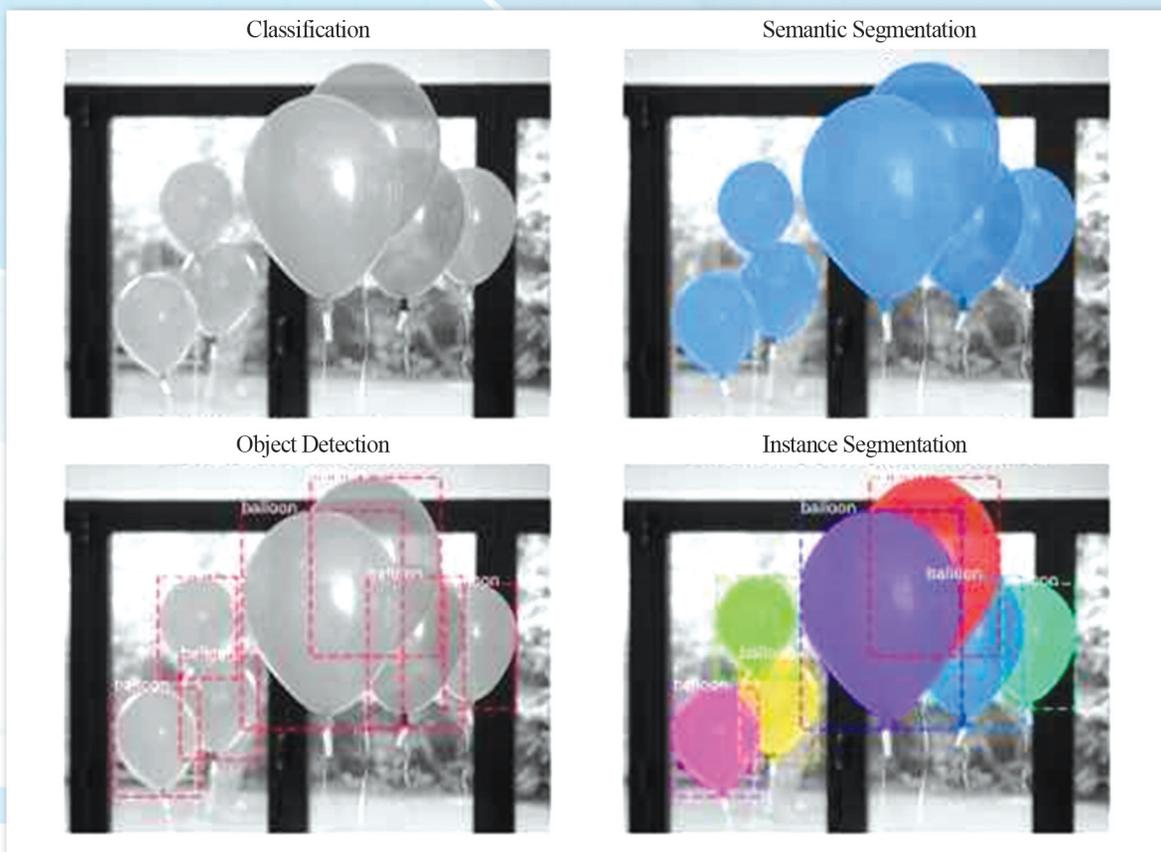
Использование сверточных нейронных сетей позволило совершить значительный скачок в точности в рамках вышеупомянутого проекта ILSVRC, для чего были предложены различные архитектуры, такие как AlexNet [7], GoogleNet, VGG, ResNet [8].

## Задача классификации изображений судов

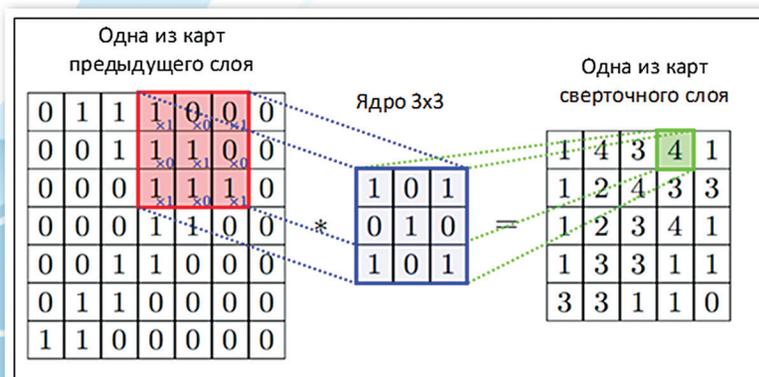
Imagery classification for commercial vessels

Важной особенностью моделей, основанных на использовании нейронных сетей, является возможность производить перенос весов ядер (фильтров) из одной задачи в другую (transfer learning). Так, сеть, обученная для классификации изображений из широко известной в интернете базы ImageNet, позволяет извлечь большое количество семантической информации об изображении. В результате появляется возможность использовать фрагменты этой сети для решения своих задач распознавания без длительной тренировки и при небольшой тренировочной выборке [5].

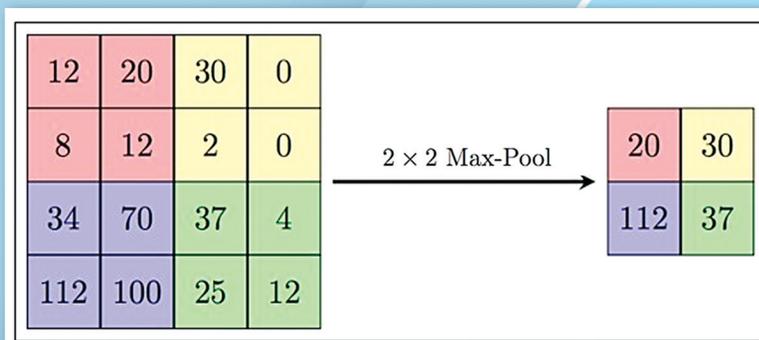
Продemonстрируем этот подход на примере решения задачи классификации кораблей по их оптическим изображениям. Взята размеченная база фотографий судов с соревнования Deep Learning Hackathon, который проводился сообществом Analytics Vidhya (Индия). Данная база состоит из 6252 изображений судов, распределенных по пяти классам: Cargo (контейнеровоз), Military (военное), Carrier (паром), Cruise (круизный лайнер), Tanker (танкер). Примеры изображений с метками представле-



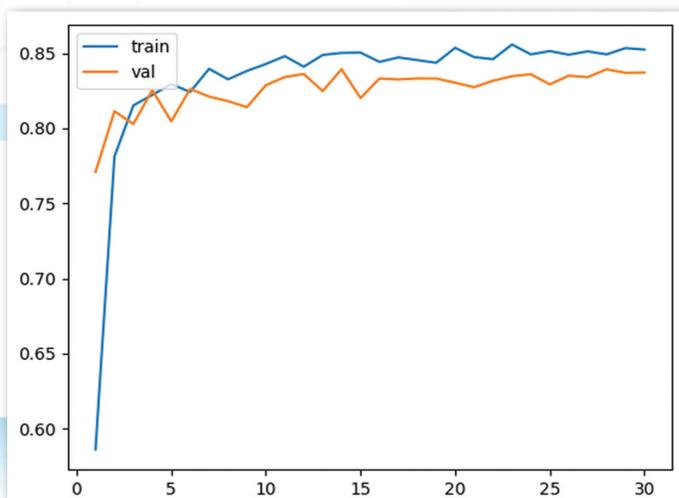
**Рис. 1.** Пример решения различных задач компьютерного зрения  
**Fig. 1.** An example of various computer-based vision applications



**Рис. 3.** Иллюстрация работы сверточного слоя  
**Fig. 3.** Illustration of convolutional layer operation

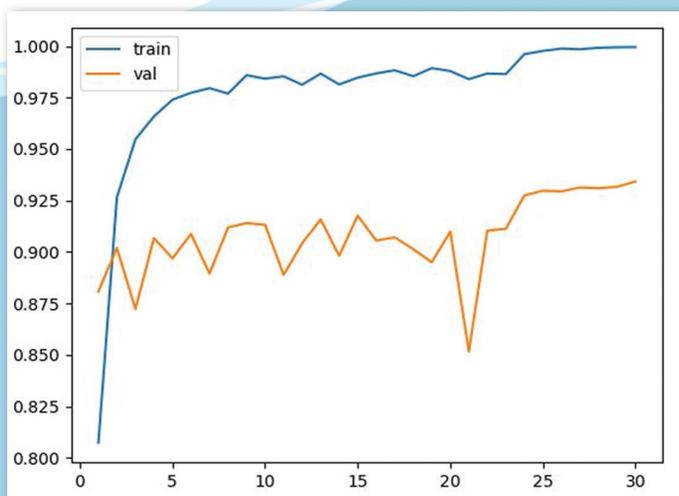


**Рис. 4.** Иллюстрация работы слоя MaxPooling  
**Fig. 4.** Illustration of MaxPooling layer operation



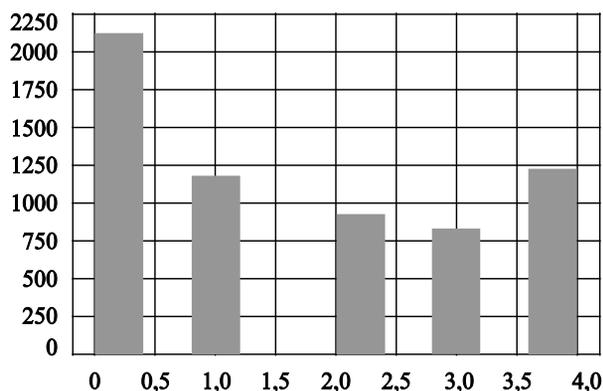
**Рис. 8.** Кривые обучения модели.  
 По горизонтальной оси отложен номер эпохи,  
 по вертикальной – доля правильных предсказаний  
 (train – тренировочный датасет;  
 val – тестовый датасет)

**Fig. 8.** Model learning curves.  
 X-axis – number of epoch,  
 Y-axis – share of correct prediction  
 (train – training dataset; val – validation dataset)



**Рис. 9.** Кривые обучения модели.  
 По горизонтальной оси отложен номер эпохи,  
 по вертикальной – доля правильных предсказаний  
 (train – тренировочный датасет;  
 val – тестовый датасет)

**Fig. 9.** Model learning curves.  
 X-axis – number of epoch,  
 Y-axis – share of correct prediction  
 (train – training dataset; val – validation dataset)



**Рис. 6.** Гистограмма распределения изображений по классам в следующем порядке: Cargo, Military, Carrier, Cruise, Tanker

**Fig. 6.** Histogram of ship image classification: Cargo, Military, Carrier, Cruise, Tanker

ны на рис. 5 (см. вклейку), гистограмма с распределением их по классам в исходной базе – на рис. 6.

В качестве базовой модели для классификации была использована модификация сверточной нейросети ResNet18, архитектура которой представлена на рис. 7. На вход сети подается нормализованное изображение размером  $224 \times 224$  пикселей. Первым следует сверточный слой из 64 фильтров размером  $7 \times 7$  с шагом 2 (conv1), далее идет слой MaxPooling с фильтром  $3 \times 3$  и тоже шагом 2. После этого следуют четыре блока, каждый из которых состоит из двух сверточных слоев  $3 \times 3$ , количество фильтров удваива-

Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	$7 \times 7, 64, \text{stride } 2$
		$3 \times 3 \text{ max pool, stride } 2$
conv2_x	$56 \times 56 \times 64$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	$7 \times 7 \text{ average pool}$
fully connected	1000	$512 \times 1000 \text{ fully connections}$
softmax	1000	

**Рис. 7.** Архитектура нейронной сети ResNet18

**Fig. 7.** Architecture of ResNet18 neural network

ется в каждом блоке. В результате получается тензор размером  $7 \times 7 \times 512$ . На следующем шаге каждый из слоев этого тензора усредняется, в результате чего мы получаем вектор из 512 компонент (average pool). В качестве головы сети (последнего слоя) выступает полносвязный слой, который на выходе «раскидывает» изображения по 1000 классов.

Используя вышеупомянутый подход – перенос уже подобранных весов из одной задачи в другую, – мы удалили из выбранной сети только последний слой (fully connected). Вместо него вставляется наш полносвязный слой размером  $512 \times 5$  со случайными весами (имеющий на выходе не 1000 классов картинок, как ранее, а всего пять классов судов). Веса только последнего слоя оптимизированы в процессе обучения на тренировочном датасете (train dataset), составленном из 80 % фотографий исходной базы. Оценка результатов проводилась на валидационном (тестовом) датасете – 20 %-ном остатке от исходной базы фотографий судов. В качестве метрики используется ассигасу («точность»), пояснения по которой будут даны ниже.

Для увеличения точности применялась аугментация (поворот или зеркальное отражение фотографий). Для данной задачи наиболее эффективными оказались отражение по горизонтали и изменение яркости и контраста. Для обучения был использован оптимизатор Adam с параметрами  $\text{learning\_rate} = 0,001$  (скорость обучения) и коэффициентом регуляризации 0,1. Длительность обучения – тридцать прогонов датасетов через сеть (тридцать эпох). Динамика обучения модели представлена на рис. 8 (см. вклейку). Финальная точность на тестовом множестве составила 83 %.

## Задача классификации фотографий кораблей

Imagery classification for naval ship

Рассмотрим применение модели на основе сверточных нейронных сетей к задаче классификации кораблей. Для этого в интернете была собрана база фотографий кораблей, принадлежащих разным странам, включая США, Францию, Испанию, Великобританию, Германию, Бразилию и КНР. Фотографии были разделены на шесть классов:

- aircraft carrier – авианосец;
- corvette – корвет;
- cruiser – крейсер;
- destroyer – эсминец;
- fast attack – катер;
- submarine – подводная лодка.

## Матрица ошибок классификации фотографий кораблей

Error matrix in ship image classification

	Авианосец	Корвет	Крейсер	Эсминец	Катер	Подводная лодка	Recall (Полнота)
Авианосец	430	3	5	0	1	3	430/442 = 97 %
Корвет	1	670	9	15	24	2	670/721 = 93 %
Крейсер	4	4	320	18	3	1	320/350 = 91 %
Эсминец	3	24	22	570	6	4	570/629 = 91 %
Катер	0	38	3	5	270	4	270/320 = 84 %
Подводная лодка	2	2	1	0	1	660	660/666 = 99 %
Precision (точность)	430/440 = 98 %	670/741 = 90 %	320/360 = 89 %	570/608 = 94 %	270/305 = 88 %	660/674 = 98 %	

Для решения этой задачи был использован подход, основанный на дообучении (finetuning) нейронной сети. В качестве основы была взята сверточная нейронная сеть ResNet50, обученная для решения задачи классификации изображений из вышеупомянутой базы данных ImageNet (1 млн картинок она «раскидала» на 1000 классов). Последний слой сети был заменен на линейный классификатор для «раскидывания» картинок на шесть классов кораблей (а не на 1000). Далее было произведено дообучение сети для задачи классификации кораблей с оптимизатором Adam со стартовым параметром  $learning\_rate = 0,0001$  и коэффициентом регуляризации  $0,001$ . Веса всех слоев оптимизировались в процессе дообучения. В процессе этого производилось уменьшение  $learning\_rate$  в случае выхода обучающей кривой на плато. Обучение производилось на протяжении пятидесяти эпох. Кривая обучения представлена на рис. 9 (см. вклейку). В результате была достигнута общая точность модели на валидационном датасете около 93 %.

Для более точной оценки результатов работы модели сети в таблице приведена матрица ошибок, в которой сведены ошибки первого и второго родов по каждому из классов.

В последнем столбце показана «полнота» анализа: так, из 442 фотографий авианосцев в тестовом датасете модель правильно распознала 430. В строке «авианосец» видно, что три фотографии авианосцев модель сети «зачислила» в корветы, пять фото – в крейсера и т.д.

В столбце «авианосец» можно видеть, что к фотографиям авианосцев модель сети ошибочно приписала одно фото корвета, четыре фото крейсеров,

три фото эсминцев, ноль – катеров и два – субмарин (итого получилось 440 фотографий). Доля фотографий «настоящих» авианосцев в итоговой классификации модели сети оценивается ранее упомянутой метрикой «точность» (97 %) внизу столбца.

Как и следовало ожидать, лучше всего распознаются подводные лодки и авианосцы.

Проведенная работа показала возможность быстрого обучения классификаторов с использованием накопленного ранее опыта. Разработанные модели сверточных нейронных сетей также планируется использовать для распознавания кораблей по их изображениям в инфракрасном и радиолокационном диапазонах.

## Библиографический список

1. Шпенст В. Радиолокационные станции дистанционного зондирования Земли космического базирования // Компоненты и технологии. 2013. № 3. С. 154–158.
2. Клименко Н.Н., Занин К.А. Новое поколение космических аппаратов для наблюдения за морской обстановкой // Воздушно-космическая сфера. 2019, № 2(99). С. 72–82.
3. Барский А.Б. Нейронные сети: распознавание, управление, принятие решений. Москва: Финансы и статистика, 2004. 174, [1] с.
4. Рашид Т. Создаем нейронную сеть. Москва; Санкт-Петербург: Диалектика: Альфа книга, 2017. 272 с.
5. Robbins M. Does an AI need to make love to Rembrandt's girlfriend to make art? [Electronic resource] // The Guardian: [site]. 2016. May 6. URL: <https://www.theguardian.com/science/2016/may/06/does-an-ai-need-to-make-love-to-rembrandts-girlfriend-to-make-art> (accessed: July 08.2020).

6. *Markoff J.* A learning Advance in Artificial Intelligence Rivals Human Abilities [Electronic resource] // The New York Times: [site]. 2015. Dec. 10. URL: <https://www.nytimes.com/2015/12/11/science/an-advance-in-artificial-intelligence-rivals-human-vision-abilities.html> (accessed: July 08.2020).
  7. *Krizhevsky A., Sutskever I., Hinton G.* ImageNet Classification with deep convolutional neural networks [Electronic resource] // Advances in neural information processing systems. 2012. № 25(2). URL: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (accessed: July 08.2020).
  8. Aggregated Residual Transformations for Deep Neural Networks / Xie S. [et al.] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, Honolulu. Piscataway, 2017. P. 1492–1500. DOI: 10.1109/CVPR.2017.634.
- 

## References

1. *V. Shpenst.* Spaceborne radar stations for remote Earth probing // Components & Technologies. 2013. No. 3. P. 154–158 (*in Russian*).
2. *N. Klimenko, K. Zanin.* New generation of satellites for maritime surveillance // Aerospace Sphere Journal. 2019. No. 2(99). P. 72–82 (*in Russian*).
3. *A. Barsky.* Neural networks: recognition, management, decision-making. Moscow: Finances & Statistics, 2004. 174, [1] p. (*in Russian*).
4. *T. Rashid.* Make Your Own Neural Network. Moscow: St. Petersburg, Dialektika, Alfa-Kniga, 2017. 272 p. (*Russian translation*).
5. *M. Robbins.* Does an AI need to make love to Rembrandt's girlfriend to make art? [Electronic resource] // The Guardian: [site]. 2016. May 6. URL: <https://www.theguardian.com/science/2016/may/06/does-an-ai-need-to-make-love-to-rembrandts-girlfriend-to-make-art> (accessed: July 08.2020).
6. *J. Markoff.* A learning Advance in Artificial Intelligence Rivals Human Abilities [Electronic resource] //

The New York Times: [site]. 2015. Dec. 10. URL: <https://www.nytimes.com/2015/12/11/science/an-advance-in-artificial-intelligence-rivals-human-vision-abilities.html> (accessed: July 08.2020).

7. *A. Krizhevsky, I. Sutskever, G. Hinton.* ImageNet Classification with deep convolutional neural networks [Electronic resource] // Advances in neural information processing systems. 2012. № 25(2). URL: <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (accessed: July 08.2020).
  8. Aggregated Residual Transformations for Deep Neural Networks / S. Xie [et al.] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21-26 July 2017, Honolulu. Piscataway, 2017. P. 1492–1500. DOI: 10.1109/CVPR.2017.634.
- 

## Сведения об авторах

*Вялов Виктор Андреевич*, к.ф.-м.н., научный сотрудник ФГУП «Крыловский государственный научный центр». Адрес: 196158, Россия, Санкт-Петербург, Московское шоссе, д. 44. Тел.: +7 (812) 415-47-41. E-mail: [vyalov@gmail.com](mailto:vyalov@gmail.com). <https://orcid.org/0000-0002-9237-2367>.

*Андреев Александр Юрьевич*, начальник сектора ФГУП «Крыловский государственный научный центр». Адрес: 196158, Россия, Санкт-Петербург, Московское шоссе, д. 44. Тел.: +7 (812) 415-47-41. E-mail: [andre-alex@yandex.ru](mailto:andre-alex@yandex.ru).

## About the authors

*Viktor A. Vyalov*, Cand. Sci. (Phys. & Math.), Researcher, Krylov State Research Centre. address: 44, Moskovskoe sh., St. Petersburg, Russia, post code 196158. Tel.: +7 (812) 415-47-41. E-mail: [vyalov@gmail.com](mailto:vyalov@gmail.com). <https://orcid.org/0000-0002-9237-2367>.

*Alexandr Yu. Andreev*, Head of Sector, Krylov State Research Centre. Address: 44, Moskovskoe sh., St. Petersburg, Russia, post code 196158. Tel.: +7 (812) 415-47-41. E-mail: [andre-alex@yandex.ru](mailto:andre-alex@yandex.ru).

Поступила / Received: 25.03.20  
Принята в печать / Accepted: 24.08.20  
© Вялов В.А., Андреев А.Ю., 2020