

Оптический журнал

Иконика — наука об изображении
Iconics — the science of image

DOI: 10.17586/1023-5086-2023-90-10-48-66

УДК 004.93'12

Обучение нейронных сетей для классификации тепловизионных изображений на основе изображений видимого спектра

Маргарита Кареновна Ермаченкова^{1✉}, Роман Олегович Малашин²,
Арина Андреевна Бойко^{3✉}

¹Университет ИТМО, Санкт-Петербург, Россия

^{2, 3}Институт физиологии им. И. П. Павлова РАН, Санкт-Петербург, Россия

^{2, 3}Санкт-Петербургский государственный университет аэрокосмического приборостроения,
Санкт-Петербург, Россия

¹ermachenkova.rita@mail.ru <https://orcid.org/0009-0005-6760-759X>

²malashinroman@mail.ru <https://orcid.org/0000-0002-2493-839X>

³boikoaa@infran.ru <https://orcid.org/0000-0001-7520-0056>

Аннотация

Предмет исследования. Были рассмотрены методы аугментации изображений видимого спектра в задачах классификации тепловизионных изображений. **Цель работы.** Исследовать способы повышения обобщающей способности нейронных сетей, обученных на изображениях видимого спектра, для распознавания тепловизионных изображений. **Метод.** Существующие наборы тепловизионных изображений имеют ограниченный размер, и для получения таких данных требуется дорогостоящее оборудование. В то же время, классификаторы, обученные на данных видимого спектра, показывают низкую точность классификации на данных других оптических диапазонов. Существуют различные методы обогащения наборов тепловизионных данных для решения задачи распознавания объектов, например, с использованием синтезированных изображений, однако, такие подходы требуют использования тепловизионных изображений в той или иной форме, что накладывает ограничение на возможности их применения. Между тем, существуют художественные методы моделирования сцен дальней инфракрасной области спектра на основе изображений видимого спектра, с помощью которых визуальное сходство достигается, например, за счёт коррекции контраста и преобразования значений цветовых каналов. Нами был предложен и исследован метод предварительного преобразования изображений, чтобы определить, способна ли нейронная сеть извлекать из модифицированных изображений видимого спектра признаки, достаточные для обобщения на тепловизионные данные. **Основные результаты.** Благодаря подобранному методу аугментации и подготовки данных видимого спектра, уровень ошибок классификации сократился с 17% до 6%. **Практическая значимость.** Наше исследование показывает, что предложенный метод обучения позволил повысить точность классификации тепловизионных данных без использования изображений соответствующего спектра в обучающей выборке. Такой подход может быть использован как метод обогащения данных, например, если имеющиеся ресурсы для получения тепловизионных данных ограничены.

Ключевые слова: классификация тепловизионных изображений, методы аугментации данных, тепловизионный инфракрасный диапазон, обучение нейронных сетей

Благодарность: работа выполнена при поддержке Госпрограммы 47 ГП «Научно-технологическое развитие Российской Федерации» (2019–2030), тема 0134-2019-0005.

Ссылка для цитирования: Ермаченкова М.К., Малашин Р.О., Бойко А.А. Обучение нейронных сетей для классификации тепловизионных изображений на основе изображений видимого спектра // Оптический журнал. 2023. Т. 90. № 10. С. 48–66. <http://doi.org/10.17586/1023-5086-2023-90-10-48-66>

Коды OCIS: 150.1135, 100.4996.

Neural network training for thermal image classification based on visible spectrum images

MARGARITA ERMACHENKOVA¹✉, ROMAN MALASHIN², ARINA BOIKO³✉

¹ITMO University, Saint-Petersburg, Russia

^{2, 3}Pavlov Institute of Physiology, Russian Academy of Sciences, Saint-Petersburg, Russia

^{2, 3}Saint-Petersburg State University of Aerospace Instrumentation, Saint-Petersburg, Russia

¹ermachenkova.rita@mail.ru <https://orcid.org/0009-0005-6760-759X>

²malashinroman@mail.ru <https://orcid.org/0000-0002-2493-839X>

³boikoa@infran.ru <https://orcid.org/0000-0001-7520-0056>

Abstract

Subject of the study. Methods of visible spectrum images augmentation in the tasks of thermal images classification were considered. The **aim of the study** is to investigate the ways to improve the generalization ability of neural networks trained on visible spectrum images to recognize the thermal images. **Method.** Existing sets of the thermal images have limited size, and obtaining such data requires expensive equipment. At the same time, the classifiers trained on visible spectrum data show low classification accuracy on data of different optical spectra. There are various methods of enriching the thermal datasets to solve the problem of object recognition, for example, the use of synthesized images, however these approaches require the use of thermal images in this or that form, which imposes restriction on the possibilities of their application. Meanwhile, there are artistic methods of modeling far-infrared scenes based on visible spectrum images that allow to achieve visual similarity, for example, by means of contrast correction and transformation of color channel values. We have proposed and investigated a preliminary image transformation method to determine whether the classifying neural network is capable of extracting features from modified visible spectrum images sufficient to generalize to thermal data. **Main results.** Owing to the developed method of augmentation and preparation of the visible spectrum data, the level of classification errors was reduced from 17% to 6%. **Practical Significance.** Our study shows that the proposed method of training made it possible to improve the classification accuracy of the thermal imaging data without using the images of the appropriate spectrum in the training sample. This approach can be used as a method of data enrichment, for example, if the available resources for obtaining thermal imagery data are limited.

Keywords: thermal image classification, data augmentation methods, thermal infrared images, neural network training

Acknowledgment: the work was carried out with the support of the State Program 47 GP "Scientific and Technological development of the Russian Federation" (2019–2030), topic 0134-2019-0005.

For citation: Ermachenkova M.K., Malashin R.O., Boiko A.A. Neural network training for thermal image classification based on visible spectrum images [In Russian] // Opticheskii Zhurnal. 2023. V. 90. № 10. P. 48–66. <http://doi.org/10.17586/1023-5086-2023-90-10-48-66>

OCIS codes: 150.1135, 100.4996.

ВВЕДЕНИЕ

В исследованиях, связанных с тепловизионными (далее ТПВ) изображениями, очень актуальны идеи переиспользования и адаптации данных видимого (далее ТВ) спектра с целью

обогащения наборов данных тепловизионных изображений. Существующие в открытом доступе базы данных с попарными фотографиями видимого и тепловизионного спектра ограничены в размерах [1, 2–5], для получения

таких данных требуется дорогостоящее оборудование [2, 4], и при этом они не лишены недостатков (низкая контрастность и зернистость на тепловизионных изображениях, вспышки, шумы, размытие на изображениях видимого спектра [1, 6, 7]). При этом известно, что классификаторы, обученные на данных видимого спектра, демонстрируют низкую точность классификации на данных, зарегистрированных в ином оптическом диапазоне, например, в ближнем инфракрасном спектре [6, 8]. В данном исследовании мы рассмотрим проблемы конфигурации и модификации обучающей выборки, состоящей из изображений видимого спектра для решения задачи классификации тепловизионных изображений, методы аугментации для повышения качества распознавания.

Существуют различные методы обогащения наборов тепловизионных данных для решения задач распознавания объектов, например, за счёт изображений, синтезированных с помощью генеративных состязательных сетей [7–9], и с помощью архитектуры кодировщик/декодировщик, обучаемой с использованием ошибки реконструкции [10–12, 14]. При этом, для достижения высокой достоверности полученных данных требуется пары изображений, снятых в видимом и инфракрасном диапазонах (иногда не идентичных по содержанию сцены [9]). Также представлены работы, исследующие возможности распознавания тепловизионных изображений на основе переноса признаков (transfer learning) [12, 13] (сеть, обученная на видимом спектре, подстраивается к ТПВ данным). Для всех представленных подходов требуется использование тепловизионных изображений в том или ином виде, что накладывает ограничение на возможности их использования. При этом существуют художественные техники по моделированию изображений ближнего инфракрасного света на основе изображений видимого спектра, с которыми за счёт корректировки контрастности и преобразований значений цветовых каналов достигается видимое сходство [15]. Такие методы можно использовать в качестве аугментации данных [16].

Целью нашей работы было исследовать способы повышения обобщающей способности нейронных сетей, обученных на изображениях видимого спектра, для распознавания тепло-

визионных изображений. В частности, мы исследовали вопросы использования цветовой и геометрической аугментации, а также особенности формирования обучающей выборки.

1. БАЗЫ ДАННЫХ, СОДЕРЖАЩИЕ ТЕПЛОВИЗИОННЫЕ ИЗОБРАЖЕНИЯ

В этой работе мы решили сконцентрироваться на задаче классификации изображений, поскольку она является наиболее исследованной и устоявшейся областью применения нейронных сетей глубокого обучения. Известно, что нейронные сети при наличии достаточной обучающей выборки могут быть крайне эффективными при решении этой задачи [14].

Для проведения экспериментов требовалось найти такой набор данных, который будет соответствовать следующим параметрам:

1. База данных включает тепловизионные изображения и изображения видимого спектра с одинаковыми классами, разрешением изображений, количеством объектов в выборках;

2. База данных будет пригодна для задачи классификации изображений.

По результатам поиска мы выделили четыре базы данных (таблица 1).

Thermal Image dataset for object classification [3] включает в себя более семи тысяч изображений, снятых на две различные тепловизионные камеры. Снимки поделены на три класса: «cat», «car», «man». Пример такого снимка изображён на рис. 1а.

Набор данных Chips Thermal Face [4] содержит более 1200 тепловизионных изображений мужчин и женщин с трёх разных континентов, возрастом от 18 до 23 лет. Этот набор данных предназначен для создания точной тепловой классификации лиц и систем теплового распознавания лиц. Из описания следует, что снимки были сделаны в разных условиях — различные локации (дом, улица), рядом с объектами с высокой температурой (батареи, кухонная техника) или без них, на разных расстояниях от камеры, к каждому сценарию применялось 5 различных цветовых режимов, участники могли быть в аксессуарах или без них, в группе или поодиночке, а также с поворотом головы лицом к камере на разный диапазон градусов (пример снимка на рис. 1б).

Таблица 1. Наборы данных с тепловизионными изображениями
Table 1. Datasets with thermal images

№	Название базы данных	Наличие изображений видимого спектра	Предназначение набора данных	Количество изображений
1	Thermal Image dataset for object classification	Нет в наличии	Классификация объектов	7428
2	Chips Thermal Face Dataset	Нет в наличии	Распознавание лиц	Более 1200
3	Teledyne FLIR Free ADAS Thermal Dataset v2	Есть в наличии	Обнаружение объектов	26442
4	Pitch-In LBAM Thermal Imaging Dataset	Нет в наличии	Классификация	Не указано

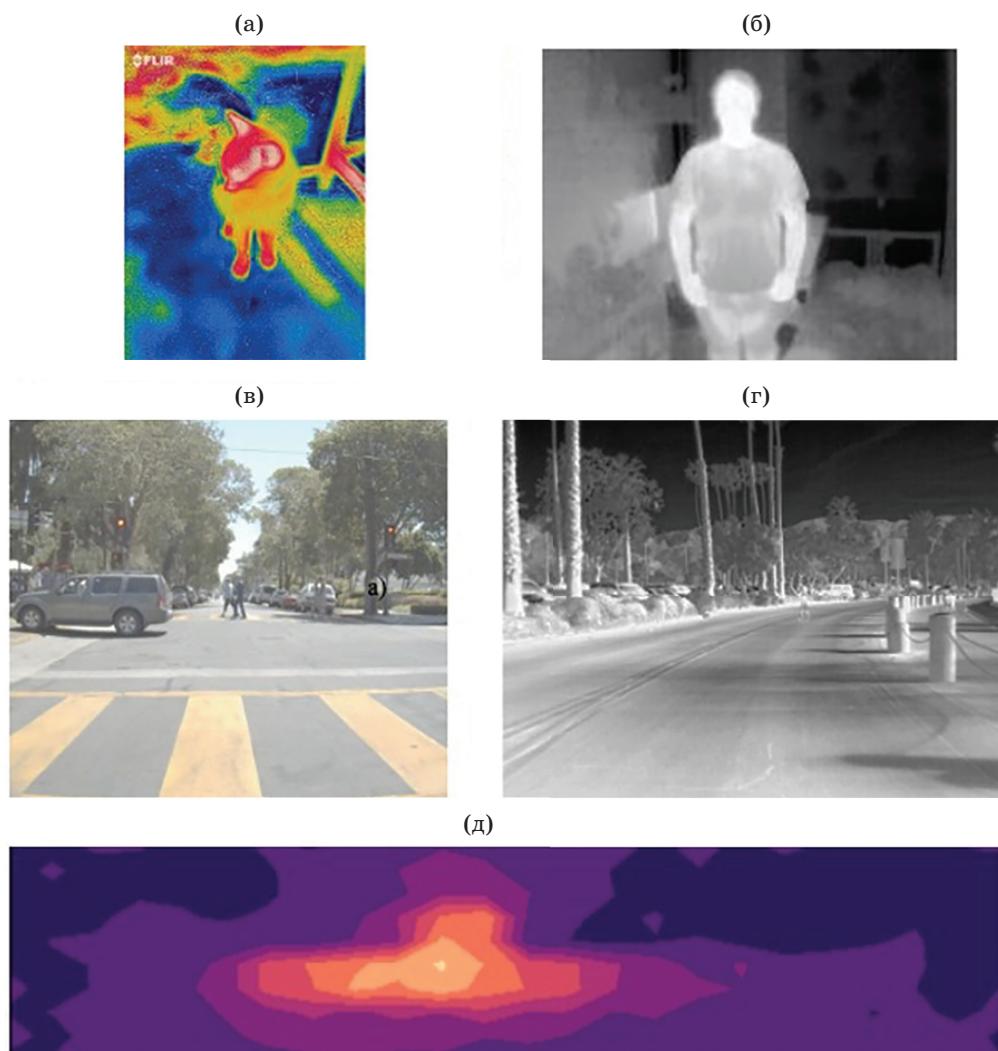


Рис. 1. Примеры для рассматриваемых наборов данных. Набор данных Thermal Image dataset for object classification (а); набор данных Chips Thermal Face (б); пример изображения видимого спектра набора данных FLIR ADAS Thermal Dataset v2 (в); пример тепловизионного изображения набора данных FLIR ADAS Thermal Dataset v2 (г); пример тепловизионного изображения набора данных Pitch-In LBAM Thermal Imaging (д)

Fig. 1. Examples for the considered datasets. (а) Thermal Image dataset for object classification; (б) Chips Thermal Face dataset; (в) example of visible image of FLIR ADAS Thermal Dataset v2; (г) example of thermal image of FLIR ADAS Thermal Dataset v2; (д) example of thermal image of Pitch-In LBAM Thermal Imaging

FLIR ADAS Thermal Dataset v2 включает в себя полностью размеченные кадры тепловизионного и видимого спектра, предназначенные для задачи обнаружения объектов [2]. В этой базе данных содержатся дневные,очные снимки и снимки в условиях сумерек/рассвета. Для тепловизионной выборки предложено 16 классов, для выборки видимого спектра — 14. Примеры изображений видимого и тепловизионного спектра представлены на рис. 1в и 1г.

Pitch-In LBAM Thermal Imaging Dataset [5] — набор данных для задачи обнаружения дефектов на машине LBAM (лазерное аддитивное производство). Пример изображения из базы данных Pitch-In LBAM приведён на рис. 1д.

Для нашего исследования помимо большого размера базы данных предпочтительно, чтобы изображения имели схожие сцены, схожее количество и качество объектов для двух выборок. Исходя из этих критерииев, для эксперимента был выбран FLIR ADAS Thermal Dataset v2, для которого съёмка проводилась одновременно на два типа камер. В связи с этим, изображения в инфракрасном спектре и в видимом спектре в нём схожи, категории совпадают, а проблему применения изображений для задачи классификации, а не обнаружения объектов, можно решить путём выделения и вырезания изображений объектов из исходной базы данных.

2. МЕТОДЫ ПОДГОТОВКИ И АУГМЕНТАЦИИ ДАННЫХ ДЛЯ ИМПАТИИ ТПВ ИЗОБРАЖЕНИЙ

Набор данных Teledyne FLIR Free ADAS Thermal Dataset v2 [2] был предназначен для задачи обнаружения объектов. При этом классы объектов не были сбалансированы по умолчанию, поэтому было принято решение использовать объекты, принадлежащие к классам с наибольшим количеством образцов — это классы «person», «car», «light», «sign». Всего база данных содержит три набора данных: обучающий (тепловизионные изображения — 175040, изображения видимого диапазона — 169174), валидационный (тепловизионные изображения — 16696, изображения видимого диапазона — 16909) и тестовый (тепловизионные изображения — 3749, изображения

Таблица 2. Среднее значение высоты и ширины описываемого прямоугольника для объекта для каждой выборки

Table 2. The average value of the height and width of the bounding box for the object for each sample

Название выборки	Среднее значение ширины (количество пикселов)	Среднее значение высоты (количество пикселов)
Обучающий набор, ТВ	66	74
Валидационный набор, ТВ	65	70
Обучающий набор, ТПВ	28	32
Валидационный набор, ТПВ	28	32
Тестовый набор, ТВ	31	32
Тестовый набор, ТПВ	22	24

видимого диапазона — 3753). В разделе, описывающем эксперименты, мы показываем, что нерелевантная геометрическая аугментация крайне негативно сказывается на точность решаемой задачи (классификации ТПВ данных при обучении на ТВ данных). При этом тестовая выборка носит косвенные признаки геометрического отклонения от валидационной и обучающей, что видно по табл. 2 — средний размер описывающих прямоугольников у тестовой выборки видимого спектра значительно отличается от тестовой выборки тепловизионного спектра, что не характерно для обучающих и валидационных выборок. По этой причине (и потому, что общее количество изображений в тестовом наборе сравнительно невелико) для экспериментов изображения тестовой выборки не использовались. В связи с этим, далее мы подробно описываем только обучающий и валидационный наборы.

Для каждого объекта класса был подсчитан описывающий прямоугольник (рис. 2). Далее описывающий прямоугольник был увеличен до размера наибольшей из сторон этого же прямоугольника, и затем так же был увеличен на 20% от новой площади для внедрения дополнительного контекста к описываемому объекту.

Тепловизионные изображения и изображения видимого спектра были отфильтрованы по следующим критериям:

1) Объекты на изображении на переднем плане ничем не загорожены (“fully_visible”), как в примере на рис. 3;



Рис. 2. Пример подготовки набора данных: чёрная рамка представляет собой описывающий прямоугольник; белая рамка — описывающий прямоугольник увеличивается до размера максимальной стороны; красная рамка — увеличенная на 20% белая рамка

Fig. 2. Example of data preparation: black box represents a bounding box; white box — the bounding box is enlarged to the size of the maximum side; red box — the white box is enlarged by 20%

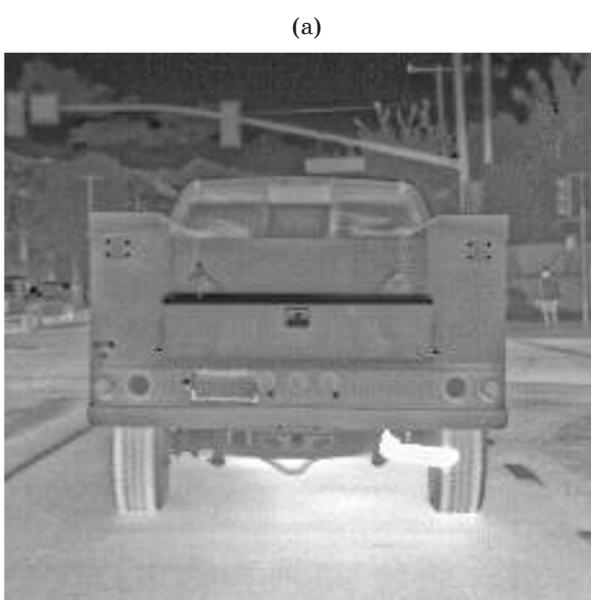
2) Диагональ описывающего прямоугольника для каждого объекта должна быть не менее 16 пикселов. Данный параметр был взят исходя из среднего размера таких описывающих прямоугольников, представленных в табл. 2. Несмотря на то, что мы берём определённого размера изображения, есть образцы, которые визуально труднораспознаваемы (рис. 4), и таким образом, полученная выборка является достаточно сложной.

Размеры получившихся выборок тепловизионного спектра представлены в табл. 3.

Такую же фильтрацию мы применили и для набора изображений видимого спектра, однако, дополнительно мы также постарались учесть критерий времени съёмки, поскольку ночные изображения видимого спектра сильным образом отличаются от дневных изображений — часто присутствуют вспышки от фонарей, автомобильных фар, светофоров (табл. 4).

Наше предположение заключалось в том, что для имитации тепловизионных изображений больше подходят изображения видимого диапазона, снятые в дневное время (рис. 5).

Всего можно выделить два типа аугментации — яркостную и геометрическую. Последняя связана с применением к изображению



(a)



(б)

Рис. 3. Сравнение снимков класса “car” с разной степенью закрытости. Ничем не закрыт, “fully_visible” (а); закрыт на 70–90% (б)

Fig. 3. Comparison of «car» class images with different degrees of occlusion. (а) Not occluded by anything, «fully_visible»; (б) occluded by 70–90%

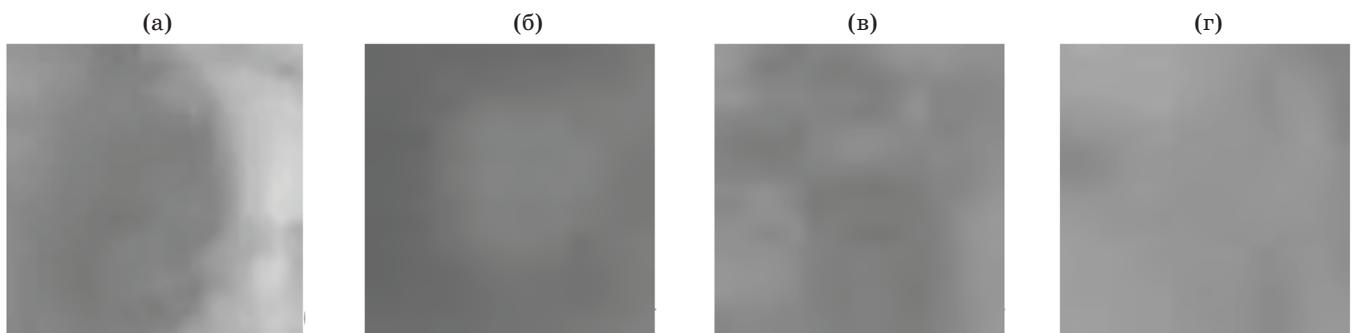


Рис. 4. Примеры нераспознаваемых изображений. Пример изображения класса “light” (а); пример изображения класса “car” (б); пример изображения класса “person” (в); пример изображения класса “sign” (г)

Fig. 4. Examples of unrecognized images. (a) Example of an image of the class «light»; (b) example of an image of the class «car»; (в) example of an image of the class «person»; (г) example of an image of the class «sign»

Таблица 3. Количество тепловизионных изображений в наборах данных, использованных в экспериментах
Table 3. Number of thermal images in the datasets used in the experiments

Имя класса	Полный набор данных		После фильтрации	
	Обучающая выборка	Валидационная выборка	Обучающая выборка	Валидационная выборка
person	50478	4470	22569	1987
car	73623	7133	21153	2789
light	16198	2005	8149	943
sign	20770	2472	9230	1222

Таблица 4. Изображения объектов в видимом спектре после фильтрации
Table 4. Images of objects in the visible spectrum after filtering

Имя класса	Полный набор данных (включающий образцы без аннотации времени суток)		День		Сумерки и рассвет		Ночь	
	Обучающая выборка	Валидационная выборка	Обучающая выборка	Валидационная выборка	Обучающая выборка	Валидационная выборка	Обучающая выборка	Валидационная выборка
person	35007	3223	1143	1135	945	0	1888	1764
car	71281	7285	1458	2067	1297	0	2235	2889
light	18640	2143	6864	906	274	0	1377	1727
sign	29531	3581	1219	1708	714	1	2002	2614

какого-нибудь пространственного преобразования, например, вращения. В настоящей работе мы исходили из того, что основные особенности подготовки и аугментации данных для решения данной задачи должны быть связаны именно с яркостными изменениями,

так как принцип получения ТПВ и ТВ изображений с точки зрения геометрии проецирования лучей на матрицу фотоприёмника практически идентичен. В связи с этим, выборки формировались таким образом, чтобы, по возможности, исключить важность геометрической

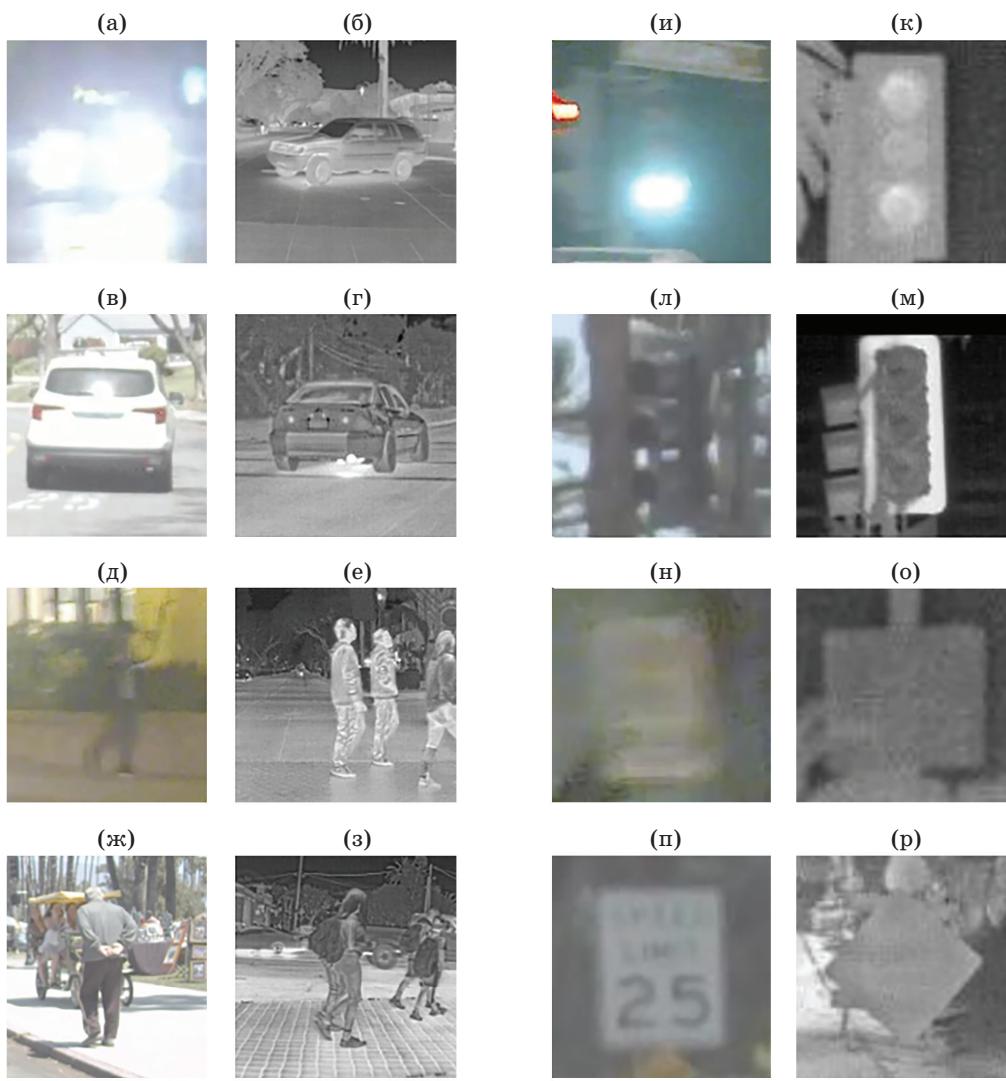


Рис. 5. Сравнение изображений разных спектров в разное время суток. Изображение видимого спектра класса “car” в ночное время (а); изображение тепловизионного спектра класса “car” в ночное время (б); изображение видимого спектра класса “car” в дневное время (в); изображение тепловизионного спектра класса “car” в дневное время (г); изображение видимого спектра класса “person” в ночное время (д); изображение тепловизионного спектра класса “person” в ночное время (е); изображение видимого спектра класса “person” в дневное время (ж); изображение тепловизионного спектра класса “person” в дневное время (з); изображение видимого спектра класса “light” в ночное время (и); изображение тепловизионного спектра класса “light” в ночное время (к); изображение видимого спектра класса “light” в дневное время (л); изображение тепловизионного спектра класса “light” в дневное время (м); изображение видимого спектра класса “sign” в ночное время (н); изображение тепловизионного спектра класса “sign” в ночное время (о); изображение видимого спектра класса “sign” в дневное время (п); изображение тепловизионного спектра класса “sign” в дневное время (р)

Fig. 5. Comparison of images of different spectrums at different times of the day. (a) Image of visible spectrum of «car» class at night time; (б) image of thermal spectrum of «car» class at night time; (в) image of visible spectrum of «car» class at day time; (г) image of thermal spectrum of «car» class at day time; (д) image of the visible spectrum of the «person» class at night; (е) image of the thermal spectrum of the «person» class at night; (ж) image of the visible spectrum of the «person» class at daytime; (з) image of the thermal spectrum of the «person» class at daytime; (и) image of the visible spectrum of the «light» class at night; (к) image of the thermal spectrum of the «light» class at night; (л) image of the visible spectrum of the «light» class at daytime; (м) image of the thermal spectrum of the «light» class at daytime; (н) photo of the visible spectrum of the «sign» class at night; (о) photo of the thermal spectrum of the «sign» class at night; (п) photo of the visible spectrum of the «sign» class at daytime; (р) photo of the thermal spectrum of the «sign» class at daytime

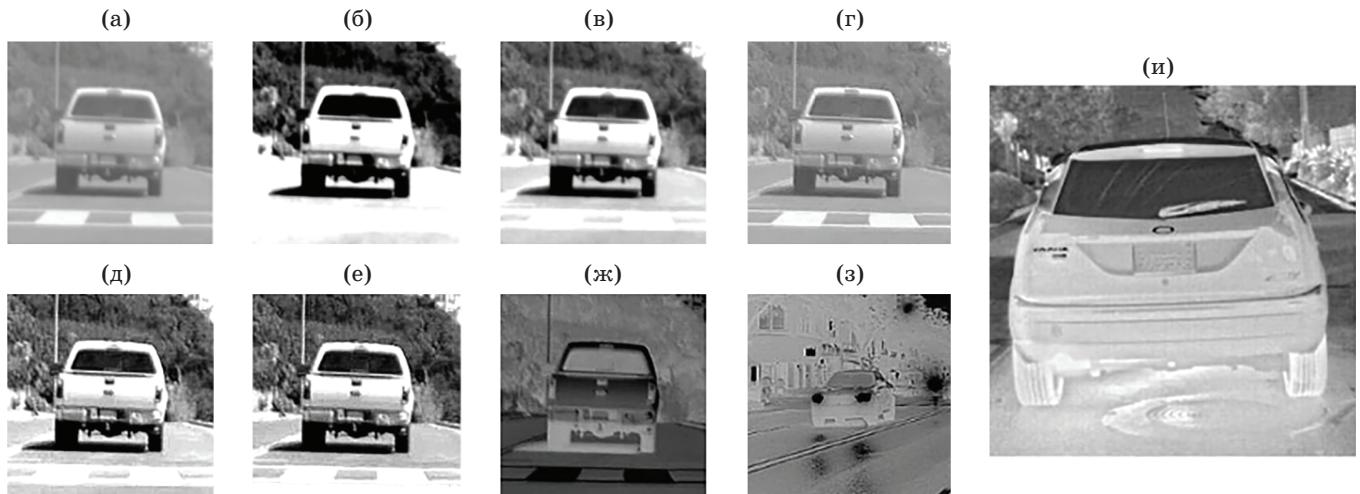


Рис. 6. Примеры аугментаций. Изображение видимого спектра без аугментаций (а); изображение видимого спектра с использованием функции с увеличением контрастности (б); изображение видимого спектра с использованием функции с уменьшением контрастности (в); изображение видимого спектра с использованием функции увеличения резкости с коэффициентом $\alpha_s = 10,0$ [19] (г); изображение видимого спектра с использованием функций с увеличением резкости и контрастности (д); изображение видимого спектра с использованием функций увеличения яркости (е); изображение видимого спектра с использованием функции инвертирования в негативный снимок (пример на автомобиле белого цвета) (ж); изображение видимого спектра с использованием функции инвертирования в негативный снимок (пример на автомобиле чёрного цвета) (з); пример изображения тепловизионного спектра (и)

Fig. 6. Examples of augmentations. (a) image of the visible spectrum without augmentations; (b) image of the visible spectrum using a function to increase contrast; (c) image of the visible spectrum using a function to reduce contrast; (g) image of the visible spectrum using a function to increase sharpness with parameter $\alpha_s = 10.0$ [19]; (d) image of the visible spectrum using functions to increase contrast and sharpness; (e) image of the visible spectrum using functions to increase brightness; (ж) image of the visible spectrum using the function of inversion into negative (as illustrated by a white vehicle); (з) image of the visible spectrum using the function of inversion into negative (as illustrated by a black vehicle); (и) example of image of the thermal imaging spectrum

аугментации — все объекты интереса располагались по центру изображения, а изображения масштабировались к одному размеру.

Далее рассмотрим способы повышения визуального сходства ТВ и ТПВ данных. В качестве примера для рис. 6 были взяты образцы одного класса. Так, на рис. 6а показано оригинальное изображение видимого спектра в чёрно-белом формате и изображение тепловизионного спектра на рис. 6з.

На тепловизионном изображении видно, что границы объектов чёткие, нет размытия, и выше контраст по сравнению с соответствующим изображением видимого спектра. Для компенсации этих эффектов на видимом изображении были применены следующие аугментации [17–19]:

1) Увеличение контрастности с помощью линейной комбинации (blending) изображе-

ния со смещённым значением его средней яркости (рис. 6б) в соответствии с формулой (2.1) [17]:

$$I_c(g, k) = \alpha_c I(g, k) + (1 - \alpha_c) \left(\frac{1}{n} \sum_{g, k} I(g, k) + 0,5 \right), \quad (2.1)$$

где $I(g, k)$ и $I_c(g, k)$ — яркость пикселя с координатами g и k исходного и контрастного изображения соответственно, n — количество всех элементов изображения, α_c — коэффициент контрастности;

2) Нормализация яркости (histogram equalization) [18] (рис. 6в);

3) Увеличение резкости снимка, как показано на рис. 6г, для прибавления чёткости краям объекта с помощью маскирования размытия (unsharp masking) [19].

Выходное изображение также можно представить в виде линейной комбинации исходного изображения и результата его свёртки с известным ядром F , сглаживающим изображение [17] (2.2):

$$I_s(g, k) = \alpha_s I(g, k) + \\ + (1 - \alpha_s) \left(\sum_i^S \sum_j^S F(i, j) I(g - i, k - j) \right), \quad (2.2)$$

где $I(g, k)$ и $I_s(g, k)$ — яркость пикселя с координатами g и k исходного и резкого изображения соответственно, S — размер ядра, α_s — коэффициент резкости;

4) Линейное увеличение яркости в α_b раз, как на рис. 6е [17];

5) Инвертирование яркостей для преобразования в негативный снимок [18]. На рис. 6 образцы под буквами (ж) и (з) представлены для визуального сравнения аугментированного изображения видимого спектра с изображениями тепловизионного спектра.

Таким образом, мы исходим из того, что комбинация следующих преобразований может приводить к повышению визуального сходства ТВ с ТПВ:

- увеличение яркости,
- увеличение контраста,
- инвертирование,
- повышения резкости изображений.

В пакетах для аугментации изображения вместо отдельных манипуляций с яркостью и контрастом используют обобщение этих

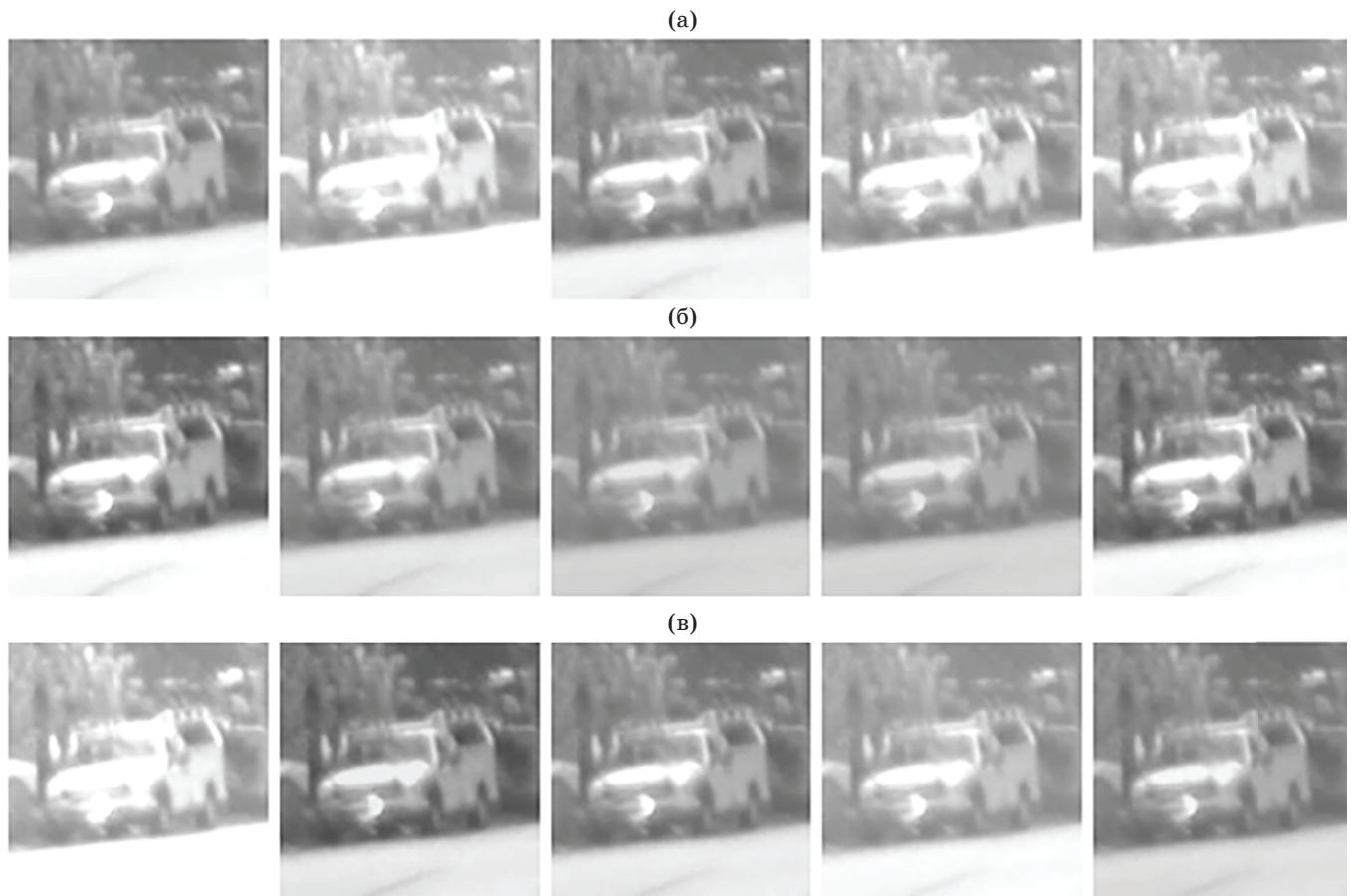


Рис. 7. Применение произвольного изменения яркости и контрастности изображения. Коэффициент изменения яркости α_b выбирается случайным образом от 0,6 до 1,4 (а); коэффициент для изменения контрастности α_c выбирается случайным образом от 0,6 до 1,4 (б); коэффициенты для изменения яркости и контрастности одновременно выбираются случайным образом от 0,6 до 1,4 (в)

Fig. 7. Application of random image brightness and contrast variation. (а) The coefficient α_b for brightness changing is chosen randomly from 0.6 to 1.4; (б) the coefficient α_c for contrast changing is chosen randomly from 0.6 to 1.4; (в) the coefficients for brightness and contrast changing are chosen randomly from 0.6 to 1.4

преобразований — рандомизированные цветовые изменения (color jitter [16]), которые включают в себя помимо указанных выше также случайное изменение насыщенности и тона. Пример аугментации изображений с помощью процедуры color jitter (без манипуляций с тоном и насыщенностью) приведён на рис. 7.

3. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ

Параметры обучения

Для проведения экспериментов мы воспользовались проектом обучения моделей компьютерного зрения, реализованным с использованием библиотеки PyTorch [20], и применили некоторые предлагаемые для обучения на ImageNet [21] параметры из этого репозитория.

Общая информация

Трудность проведения исследования заключалась в сложности отделения разных факторов, сказывающихся на точность распознавания тепловизионных изображений на основании обучения на изображениях видимого диапазона. Например, использование геометрической аугментации может по-разному сказываться на эту характеристику при малой и большой обучающей выборке, при старте обучения со случайными и неслучайными весами [22]. Нам приходилось перепроверять промежуточные выводы уже с учётом выводов, полученных к настоящему моменту, и иногда ограничено воспроизводить некоторые из них.

В тексте ниже, чтобы каждый раз не оговариваться, при предъявлении результатов мы будем ссылаться на выборки следующим образом:

- снятые днём — day;
- снятые ночью — night;
- снятые в сумерках или во время рассвета — dawn;
- весь набор (включая неаннотированное время съёмки) — whole;

Такое дополнительное описание будет использоваться как для наборов данных видимого спектра, так и для тепловизионных данных. Также мы будем ссылаться на результаты, полученные с использованием:

- случайной инициализации весов — rand-init;
- предобученных весов — imagenet-init.

Обучающие и валидационные выборки

Все обучающие и валидационные выборки были получены из базы данных FLIR ADAS Thermal Dataset v2 [2] с использованием процедуры фильтрации по размеру и центрирования, описанной в разделе 2. В каждом эксперименте мы приводим точности классификации валидационных выборок видимого и тепловизионного спектров, достигнутых в конце обучения (тестовая выборка не использовалась по причинам, описанным в разделе 2).

Для получения обучающей и валидационной выборки видимого спектра мы использовали изображения, находящиеся в каталогах images_rgb_train, images_rgb_val. При этом, если в соответствии с дизайном эксперимента к обучающей выборке применялась дополнительная фильтрация (по времени суток), то такая же фильтрация применялась и к валидационной выборке видимого спектра.

Для валидационной выборки тепловизионных изображений дополнительная фильтрация не применялась никогда (использовался набор whole), поскольку целью было исследовать влияние предлагаемых методов на распознавание тепловизионных данных, включающих ночную съёмку, когда использование невидимого глазом спектра наиболее оправдано. Кроме того, для того, чтобы повысить достоверность полученных в нашей работе выводов, для оценки точности распознавания тепловизионного набора мы использовали не валидационный, а обучающий набор (находящийся в каталоге images_thermal_train) базы данных FLIR ADAS Thermal Dataset v2, поскольку он содержит на порядок больше примеров.

Архитектура нейронной сети

Мы провели несколько предварительных экспериментов с популярными архитектурами нейронных сетей, включая ResNet [23] и EfficientNet [24], обучая их с нуля или брёя в качестве инициализации веса, предобученные на ImageNet. Для распознавания валидационной ТВ выборки (особенно с использованием предобученных весов) чуть более эффективными оказывались решения с большим количеством настраиваемых параметров (resnet-152 и efficientnet-b6), однако, для задачи обобщения на тепловизионный спектр

размер сети не оказывал ключевого влияния, при этом затраты на обучение были заметно выше, чем при обучении меньших по размеру моделей.

В связи с этим, для проведения экспериментов мы решили использовать архитектуру ResNet-18 [23], поскольку ResNet архитектуры де-факто являются стандартом при проведении экспериментов с классификаторами, а кроме того:

– ResNet-18 является небольшой по глубине сетью и может быть обучена сравнительно быстро;

– она состоит из однородных блоков, характерных и для больших моделей ResNet.

Гиперпараметры оптимизации

Мы исследовали несколько наборов гиперпараметров для оптимизаторов SGD [25] и RMSPROP [26], которые позволяют эффективно обучать различные архитектуры нейронных сетей на базе данных ImageNet. Во всех случаях мы масштабировали изображение к размеру 224×224 при подаче на вход сети. Для обучения с нуля с помощью SGD мы использовали косинусный алгоритм изменения шага обучения с пятью эпохами «прогрева», 240 эпох, пакет изображений размером 192. Для RMSPROP использовались 600 эпох, плавно меняющийся шаг обучения, бегущие средние веса, пакет 64 и дополнительный шум в шаге градиентного спуска.

SGD считается одним из наиболее подходящих оптимизаторов для обучения классификаторов, однако в нашем случае эксперименты показали, что RMSPROP с указанными выше параметрами обеспечивал более стабильную сходимость нейронных сетей (с точки зрения точности распознавания ТПВ изображений) и меньшую зависимость от случайной инициализации весов при примерно той же итоговой точности, поэтому в приведённых ниже экспериментах со случайной инициализацией использовался именно он.

Дообучение нейронных сетей в наших экспериментах происходило в течение 200 эпох с использованием SGD и скачкообразного изменения шага обучения. Во всех экспериментах мы сообщаем точность распознавания валидационных выборок в конце обучения.

Геометрическая аугментация

В качестве опорного метода мы рассмотрели сочетание двух эффективных методов аугментации:

– RandomCropResize и flip — извлечение фрагмента исходного изображения с последующим масштабированием к заданному размеру (RandomCropResize), отражение фрагмента по вертикали с вероятностью 0,5 (flip).

– RandAugment [27] — модификация алгоритма AutoAugment [28, 29], который осуществляет геометрическую и яркостную аугментацию индивидуально к каждому примеру стратегиями, сочетающими до двух различных преобразований (инвертирования, вращения). RandAugment, в отличие от AutoAugment, вместо набора фиксированных стратегий (пар преобразований) каждый раз использует случайно сгенерированную стратегию. В список используемых алгоритмом RandAugment аугментаций входят такие преобразования, как авто- и ручное контрастирование, инвертирование, вращение, изменение насыщенности, резкости, яркости и нормализация яркости, постеризация, эффект выцветания, проективные преобразования и смещения. В наших экспериментах оказалось, что RandAugment превосходит AutoAugment (со стратегиями преобразований, выученными для ImageNet) по точности распознавания ТПВ изображений. В табл. 5 приведены результаты экспериментов для RandAugment.

В условиях небольшой обучающей выборки нейронные сети характеризуются ограниченными обобщающими способностями к наличию нерелевантных образов и геометрических преобразований [22], т.к. это повышает сложность задачи и, таким образом, снижает

Таблица 5. Точность распознавания ТВ и ТПВ изображений при наличии геометрической и яркостной аугментации ImageNet (rand-init, day)

Table 5. Accuracy of TV and thermal images in the presence of ImageNet geometric and brightness augmentation (rand-init, day)

Random-crop-resize	RandAugment	ТВ	ТПВ
+	стандартн.	98,0	82,8
–	стандартн.	98,2	90,5
–	без геом.	97,5	89,3

точность распознавания для всей тестовой выборки.

Поскольку база данных строилась таким образом, чтобы исключить важность геометрических преобразований, можно было бы ожидать, что их вклад будет либо негативным, либо нейтральным, если они не будут согласованы с подготовкой базы данных. Такая несогласованность характерна для процедуры RandomCropResize.

Важно, что снижение точности распознавания валидационной выборки при использовании геометрической аугментации оказалось незначительным (0,2%) и намного больше для распознавания тепловых изображений (8%). Это можно объяснить тем, что тепловизионные изображения имеют больше сходства с изображениями видимого спектра именно в общих очертаниях образов, а не в отдельных деталях. Кроме того, некоторые исходные изображения обучающего набора уже имели низкое разрешение, которое при этом в среднем ниже для ТПВ (размеры изображений были ранее приведены в табл. 2).

Почти во всех экспериментах, направленных на повышение точности распознавания тепловизионных изображений, мы получали одинаково стабильные результаты распознавания изображений видимого спектра ($(98 \pm 0,5)\%$). Таким образом, точность распознавания изображений видимого спектра не может быть критерием при распознавании тепловизионных данных, даже если способ выбора фрагмента на двух снимках является одинаковым. Тем не менее, для большинства экспериментов мы всё равно приводим точность для видимых изображений, т.к. это предоставляет дополнительную информацию о том, как происходило обучение.

Алгоритм RandAugment содержит внутри себя стратегии с использованием геометрических аугментаций (например, сдвиг), поэтому мы попробовали заменить их вручную на более подходящие (с нашей точки зрения) яркостные преобразования, однако это не увеличило точность распознавания (табл. 5). Можно сделать вывод, что по крайней мере некоторые виды геометрической аугментации могут быть полезны для решения данной задачи. В дальнейших экспериментах мы исключили внешние геометрические аугментации (RandomCropResize, flip), но оста-

вили AutoAugment и RandAugment в неизменном виде, предлагаемом для обучения на ImageNet.

Удаление фрагментов (Random Erasing)

Одной из известных техник расширения данных является удаление случайных фрагментов изображения [30]. В соответствии с предположением о том, что отдельные детали изображений видимого спектра неважны для формирования нейронной сетью признаков, пригодных для распознавания снимков теплового спектра, можно предположить, что данный вид аугментации окажется важным.

Мы рассмотрели несколько способов удаления фрагментов:

- удалённый блок заполняется нулями (const);
- удалённый блок заполняется случайными пикселями (pixel).

Эксперименты показали, что эта процедура несёт незначительный вклад и лишь слегка повышает точность распознавания тепловизионного спектра при отсутствии другой аугментации (табл. 6).

Resnet-18, обученная без использования аугментации на видимом диапазоне, обеспечивает точность распознавания ТПВ изображений 66,5%, а с использованием случайного удаления фрагментов можно поднять эту точность на 2,5%.

Несмотря на незначительный вклад этой аугментации, мы её использовали во всех остальных экспериментах (pixel, размер стороны удаляемого фрагмента до 0,2), т.к. ожидали, что она может иметь положительный вклад в решение задачи вместе с другими видами аугментации.

Таблица 6. Точность распознавания ТПВ изображений при применении случайного удаления фрагментов в зависимости от размера удалённого фрагмента относительных размеров преобразуемого изображения (rand-init, day)

Table 6. Accuracy of thermal image recognition using random fragment removal (rand-init, day)

Способ удаления фрагментов	Размер стороны удалённого фрагмента 20%	Размер стороны удалённого фрагмента 40%
const	67,9	68
pixel	67,8	69,13

**Изменение естественного освещения:
исследование влияния времени суток
съёмки видимых изображений
на результаты распознавания
тепловизионных**

В разделе 2 мы привели примеры данных видимого спектра и указали на то, что дневные изображения могут быть наиболее подходящими для имитации ТПВ изображений. Для проверки этой гипотезы мы обучили ResNet-18 на четырёх разных наборах изображений. Результаты приведены в табл. 7.

Как видно из таблицы, наилучшая точность распознавания ТПВ достигается на выборке дневных изображений несмотря на то, что она самая маленькая. К удивлению, точность распознавания при использования полного

набора наименьшая. При этом, однако, тяжело судить о том, какая пропорция в полном наборе сумеречных изображений, добавление которых, как видно из таблицы, имеет сильный отрицательный вклад в точность распознавания.

Преимущество использования дневных изображений видимого диапазона для имитации ТПВ хорошо заметно на очень малых обучающих выборках. Например, точность распознавания ТПВ данных с помощью ResNet-18, обученной на 100 случайных примерах дневной съёмки, составила почти 76%. В то же время использование 100 случайных примеров без ограничения на период съёмки позволило получить лишь 69% (табл. 8). Схожий эффект достигается, если доучивать нейронную сеть без использования какой-либо аугментации.

При этом, если обучать сеть без аугментации, начиная со случайно инициализированных весов на всей базе данных, то получается строго обратное: лучше уже работает использование большей обучающей выборки, включающей все времена суток (табл. 9).

Это согласуется с гипотезой о большей пригодности именно дневных изображений видимого диапазона для имитации ТПВ изображений, поскольку при случайной инициализации и отсутствии аугментации любое

Таблица 7. Точность распознавания в зависимости от времени съёмки (random-init)

Table 7. Accuracy of recognition depending on shooting time (random-init)

Время съёмки	ТВ, %	ТПВ, %
Любое (whole)	98,0	88,9
День, ночь, сумерки (аннотированные)	98,1	90,5
День, сумерки	98,0	90,0
День	98,2	90,8

Таблица 8. Результаты при использовании 100 обучающих примеров

Table 8. Results of using 100 training examples

Время съёмки	Аугментация (RandAugment)	Инициализация сети	ТВ, %	ТПВ, %
День	+	rand-init	88,2	75,9
Любое	+	rand-init	85,6	69,2
День	–	imagenet-init	95,7	76,1
Любое	–	imagenet-init	95,6	74,1

Таблица 9. Результаты без использования аугментации данных

Table 9. Results without data augmentation

Время съёмки	Аугментация (RandAugment)	Инициализация сети	ТВ, %	ТПВ, %
Любое	–	rand-init	97,5	72,7
День, сумерки, ночь	–	rand-init	97,6	71,7
День, сумерки	–	rand-init	98	67,3
День	–	rand-init	97,9	66,5

разнообразие изображений позволяет лучше сформировать признаки общего назначения, располагающиеся в первых слоях нейронной сети, однако эти признаки неспецифичны к решаемой задаче. Предобученная сеть уже обладает такими признаками и дневные изображения в итоге обеспечивают лучший результат.

Данный вывод является крайне важным, поскольку в ходе выполнения первых экспериментов мы не заметили пользы от предобученных весов для распознавания ТПВ изображений, пока используемая аугментация не была согласована с представлениями об имитации ТПВ данных. Во всех экспериментах, описанных далее, используется выборка, состоящая только из дневных изображений.

Аугментация через имитацию ТПВ

В разделе 2 мы привели примеры яркостных преобразований, которые с нашей точки зрения обеспечивают визуальное сходство изображений видимого спектра с изображениями ТПВ спектра:

- инвертирование изображения;
- усиление резкости.

В этом разделе мы приводим результаты экспериментов, которые подтверждают эффективность использования этих преобразований. Resnet-18, обученная без использования аугментации, обеспечивает точность 66,5%. Результаты экспериментов приведены в табл. 10.

Во-первых, мы используем преобразование *color jitter*, которое обобщает преобразование увеличения контраста. Хотя по отдельности это лишь слегка повышает точность распознавания ТПВ изображений, в совокупности с другими преобразованиями оказывается достаточно эффективным. Добавление инвертирования изображения с вероятностью 0,5 уже значительно повышает точность распознавания ТПВ изображений, хотя лишь удваивает количество примеров обучающей выборки. Также мы наблюдаем положительный вклад преобразования увеличения резкости. Набор преобразований, описанных выше, уже позволяет догнать по точности методы, основанные на индивидуальном формировании стратегии аугментации к каждому изображению (*RandAugment* и *AutoAugment*). Совмещение *AutoAugment* и *RandAugment* с предлагае-

Таблица 10. Аугментация с использованием преобразований имитации ТПВ (*rand-init*)

Table 10. Augmentation using thermal image imitation transformations (*rand-init*)

Аугментация	ТВ, %	ТПВ, %
–	98,3	67,8
color jitter	97,7	70,1
color jitter + invert	97,4	89,7
color jitter + invert + sharpness (Thermal imitation)	97,2	90,6
<i>RandAugment</i> [27]	98,4	90,5
<i>AutoAugment</i> [28]	98,3	90,3
<i>RandAugment</i> + invert	98,1	90,8
<i>AutoAugment</i> + invert + sharpness	97,7	90,9
thermal imitation + <i>AutoAugment</i>	97,4	91,4
thermal imitation + random augment	97,5	92,1

мым подходом позволяет повысить итоговую точность ещё на 1,6% (*RandAugment*) и 1,1% (*AutoAugment*) соответственно. Заметим, что в приведённых в табл. 10 экспериментах мы не использовали геометрическую аугментацию, связанную со случайным отражением изображения по горизонтали. Интуитивно кажется, что такая аугментация должна повышать точность распознавания и в видимом, и в тепловизионном спектре, однако в наших экспериментах добавление этой аугментации не повлияло на результаты. Мы связываем это с особенностью решаемой задачи. Камеры располагаются на автомобиле в определённом положении относительно дороги, и, например, дорожные знаки (в видимом диапазоне) не инвариантны к зеркальному отражению. Вероятно, нечто подобное может происходить и для ТПВ изображений, хотя подробный анализ этого вопроса выходил за рамки данного исследования.

Дообучение нейронных сетей

Инициализация с помощью весов, предобученных на ImageNet, позволяет ещё увеличить точность распознавания ТПВ данных при использовании больших моделей (табл. 11).

Согласно полученным результатам (табл. 11), большие сети показывают лучшие результаты

Таблица 11. Аугментация с использованием преобразований имитации ТПВ при использовании больших предобученных моделей

Table 11. Augmentation with thermal imaging imitation transformations using large pre-trained models

Аугментация	Инициализация	Архитектура	ТПВ, %
Thermal imitation + RandAugment	rand-init	ResNet-18	92,1
Thermal imitation + RandAugment	imagenet-init	ResNet-18	91,6
Thermal imitation + RandAugment	rand-init	ResNet-101	92,3
Thermal imitation + RandAugment	imagenet-init	ResNet-50	92,9
Thermal imitation + RandAugment	imagenet-init	ResNet-101	94,0
Thermal imitation + RandAugment	imagenet-init	ResNet-152	93,4

по сравнению с Resnet-18 (точность распознавания ТВ в таблице не приведена, но она осталась на уровне чуть менее 98%). Наибольшая точность была получена с помощью Resnet-101 и составила 94,0%.

ЗАКЛЮЧЕНИЕ

Мы рассмотрели способы обучения нейронных сетей распознаванию тепловизионных изображений, когда в наличии имеются только данные видимого диапазона. Мы провели эксперименты на базе данных, содержащей изображения двух спектров. Параметры обучения, используемые для ImageNet, хотя и обеспечили высокую точность распознавания видимых изображений (98%), оказались малопригодными для обобщения этих знаний на ТПВ данных (82,8%).

Нами предложен способ аугментации изображений видимого спектра на основе метода визуальной имитации ТПВ изображений. В соответствии с проведёнными экспериментами для решения поставленной задачи следует:

1. По возможности исключить при обучении аугментации, связанные с извлечением фрагментов изображений.

2. Исключить из обучающей выборки изображения, на которых присутствуют вспышки от искусственных источников освещения.

3. Использовать для аугментации данных:
– инвертирование с вероятностью 0,5,
– случайное контрастирование,
– случайное изменение яркости,
– увеличение резкости с вероятностью 0,5.

4. Использовать (по возможности) большие предобученные на ImageNet нейронные сети.

Мы показали, что каждое из указанных предложений положительным образом сказывается на точность распознавания и при этом ни одно из них (кроме последнего пункта) не может быть оценено по точности распознавания изображений видимого диапазона. В итоге нам удалось сократить количество ошибок при распознавании ТПВ изображений чуть менее, чем в три раза (с 17,2% до 6,0%), и добиться итоговой точности распознавания более 94%.

Отметим, что мы провели эксперименты по аугментации лишь с одной конкретной базой данных на ограниченном наборе классов изображений (человек, машина, дорожный знак, светофор). Можно ожидать, что при наличии других условий съёмки и другого оборудования некоторые параметры аугментации могут быть изменены для улучшения результатов. Тем не менее, мы ожидаем, что предложенный подход будет хорошо обобщаться на различные условия, поскольку при выборе способов имитации тепловизионных изображений мы исходили из общего впечатления о визуальных свойствах изображений двух спектров.

СПИСОК ИСТОЧНИКОВ

1. Mittal U., Srivastava S., Chawla P. Object detection and classification from thermal images using region based convolutional neural network // Journal of Computer Science. 2019. V. 15. № 7. P. 961–971. <https://doi.org/10.3844/jcssp.2019.961.971>

REFERENCE

1. Mittal U., Srivastava S., Chawla P. Object detection and classification from thermal images using region based convolutional neural network // Journal of Computer Science. 2019. V. 15. № 7. P. 961–971. <https://doi.org/10.3844/jcssp.2019.961.971>

2. FLIR Systems. Inc. Free FLIR thermal dataset for algorithm training. [Электронный ресурс — описание набора данных] Режим доступа: <https://www.flir.com/oem/adas/adas-dataset-agree> (дата доступа: 13.06.2022)
3. Qirat Ashfaq, Usman Akram, Roshaan Zafar. Thermal image dataset for object classification. [Электронный ресурс — описание набора данных] Mendeley Data. 2021. Ver 1. <https://doi.org/10.17632/btmrycjpj.1>
4. Cook J. Chips thermal face dataset [Электронный ресурс — описание набора данных] (дата обновления: 29.03.2021) Режим доступа: <https://www.kaggle.com/kagglechip/chips-thermal-face-dataset> (дата обращения: 30.04.2022)
5. Miller D., Boyang Song, Farnsworth M., Divya Tiwari. Pitch-In LBAM thermal imaging dataset" [Электронный ресурс, описание набора данных] (дата обновления: 14.05.2021) Режим доступа: <https://www.kaggle.com/dbmiller/pitchin-lbam-thermal-imaging-dataset> (дата обращения: 30.04.2022)
6. Farzeen Munir, Shoaib Azam, Muhammd Aasim Rafique, Ahmad Muqeem Sheri, Moongu Jeon, Witold Pedrycz. Exploring thermal images for object detection in underexposure regions for autonomous driving // Applied Soft Computing. May 2022. V. 121. № C. <https://doi.org/10.1016/j.asoc.2022.108793>.
7. Zhou Hang, Min Sun, Xiang Ren, Xiuyuan Wang. Visible-thermal image object detection via the combination of illumination conditions and temperature information // Remote Sensing. 2021. V. 13. № 18. P. 3656. <https://doi.org/10.3390/rs13183656>
8. Tsurkov V.I, Grinchuk O.V. Cyclic generative neural networks to improve face recognition in non-standard domains // Journal of Computer and Systems Sciences International. 2018. V. 57. № 4. P. 620–625. <https://doi.org/10.1134/S1064230718040093>
9. Dai Xuerui, Yuan Xue, Wei Xueye. TIRNet: Object detection in thermal infrared images for autonomous driving // Applied Intelligence. 2021. V. 51. P. 1–18. <https://doi.org/10.1007/s10489-020-01882-2>
10. Lee D.-G., Jeon M.-H., Cho Y., Kim A. Edge-guided multi-domain RGB-to-TIR image translation for training vision tasks with challenging labels // In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). May 29, 2023 — Jun 2, 2023. London, UK. P. 8291–8298. <https://doi.org/10.1109/ICRA48891.2023.10161210>
11. Akkaya B., Altinel F., Halici U. Self-training guided adversarial domain adaptation for thermal imagery // Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), online, June 19–25. 2021. Nashville, TN, USA. P. 4317–4326. <https://doi.org/10.1109/cvprw53098.2021.00488>
12. Gundogdu E., Koç A., Alatan A.A. Object classification in infrared images using deep representations // 2016 IEEE International Conference on Image Processing (ICIP). Phoenix. Arizona. USA. 25–28 Sept. 2016. P. 1066–1070. <https://doi.org/10.1109/ICIP.2016.7532521>
13. Xu J., Vázquez D., Ramos S., López A.M., Ponsa D. Adapting a Pedestrian detector by boosting LDA exemplar classifiers // 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 2013, P. 688–693. <https://doi.org/10.1109/CVPRW.2013.104>
14. Feng D., Haase-Schütz C., Rosenbaum L., Hertlein H., Glaeser C., Timm F., Wiesbeck W., Dietmayer K. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges // IEEE Transactions on Intelligent Transportation Systems. Feb 17, 2020. V. 22 № 3. P. 1341–1360. <https://doi.org/10.1109/TITS.2020.2972974>
2. FLIR Systems, Inc. Free Flir Thermal Dataset for Algorithm Training // URL: <https://www.flir.com/oem/adas/adas-dataset-agree> (accessed 13.06.2022)
3. Qirat Ashfaq, Usman Akram, Roshaan Zafar. Thermal Image dataset for object classification. Mendeley Data. 2021. Ver 1. <https://doi.org/10.17632/btmrycjpj.1>
4. James Cook. Chips Thermal Face Dataset // URL: <https://www.kaggle.com/kagglechip/chips-thermal-face-dataset> (accessed 30.04.2022)
5. Miller D., Boyang Song, Farnsworth M., Divya Tiwari. Pitch-In LBAM Thermal Imaging Dataset // (updated 14.05.2021) URL: <https://www.kaggle.com/dbmiller/pitchin-lbam-thermal-imaging-dataset> (accessed 30.04.2022)
6. Farzeen Munir, Shoaib Azam, Muhammd Aasim Rafique, Ahmad Muqeem Sheri, Moongu Jeon, Witold Pedrycz. Exploring thermal images for object detection in underexposure regions for autonomous driving // Applied Soft Computing. May 2022. V. 121. № C. <https://doi.org/10.1016/j.asoc.2022.108793>.
7. Zhou Hang, Min Sun, Xiang Ren, Xiuyuan Wang. Visible-thermal image object detection via the combination of illumination conditions and temperature information // Remote Sensing. 2021. V. 13. № 18. P. 3656. <https://doi.org/10.3390/rs13183656>
8. Tsurkov V.I, Grinchuk O.V. Cyclic generative neural networks to improve face recognition in non-standard domains // Journal of Computer and Systems Sciences International. 2018. V. 57. № 4. P. 620–625. <https://doi.org/10.1134/S1064230718040093>
9. Dai Xuerui, Yuan Xue, Wei Xueye. TIRNet: Object detection in thermal infrared images for autonomous driving // Applied Intelligence. 2021. V. 51. P. 1–18. <https://doi.org/10.1007/s10489-020-01882-2>
10. Lee D.-G., Jeon M.-H., Cho Y., Kim A. Edge-guided multi-domain RGB-to-TIR image translation for training vision tasks with challenging labels // In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA). May 29, 2023 — Jun 2, 2023. London, UK. P. 8291–8298. <https://doi.org/10.1109/ICRA48891.2023.10161210>
11. Akkaya B., Altinel F., Halici U. Self-training guided adversarial domain adaptation for thermal imagery // Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), online, June 19–25. 2021. Nashville, TN, USA. P. 4317–4326. <https://doi.org/10.1109/cvprw53098.2021.00488>
12. Gundogdu E., Koç A., Alatan A.A. Object classification in infrared images using deep representations // 2016 IEEE International Conference on Image Processing (ICIP). Phoenix. Arizona. USA. 25–28 Sept. 2016. P. 1066–1070. <https://doi.org/10.1109/ICIP.2016.7532521>
13. Xu J., Vázquez D., Ramos S., López A.M., Ponsa D. Adapting a Pedestrian detector by boosting LDA exemplar classifiers // 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 2013, P. 688–693. <https://doi.org/10.1109/CVPRW.2013.104>
14. Feng D., Haase-Schütz C., Rosenbaum L., Hertlein H., Glaeser C., Timm F., Wiesbeck W., Dietmayer K. Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges // IEEE Transactions on Intelligent Transportation Systems. Feb 17, 2020. V. 22 № 3. P. 1341–1360. <https://doi.org/10.1109/TITS.2020.2972974>

- challenges // IEEE Transactions on Intelligent Transportation Systems. Feb 17, 2020. V. 22 № 3. P. 1341–1360. <https://doi.org/10.1109/TITS.2020.2972974>
- 15. Craig Boehman. How to create an infrared effect in Photoshop. [Электронный ресурс] (published 28.06.2022) URL: <https://www.makeuseof.com/how-to-create-an-infrared-effect-in-photoshop/> (accessed 01.05.2023).
 - 16. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning // J Big Data. 2019. V. 6. № 60. P. 1–48. <https://doi.org/10.1186/s40537-019-0197-0>
 - 17. Haeberli P., Voorhies B. Image processing by linear interpolation and extrapolation // IRIS Universe Magazine. 1994. V. 28. P. 8–9.
 - 18. Gonzalez R.C., Woods R.E. Digital image processing (4th ed.). New York, NY: Pearson, 2018. P. 138–140.
 - 19. Leo Levi. Unsharp masking and related image enhancement techniques // Computer Graphics and Image Processing. 1974. V. 3. № 2. P. 163–177.
 - 20. Wightman R. PyTorch Image Models. GitHub repository. 2019. [Электронный ресурс] Режим доступа: <https://github.com/rwightman/pytorch-image-models>. (дата обращения: 30.04.2023) <https://doi.org/10.5281/zenodo.4414861> Wightman, R. PyTorch Image Models. GitHub repository. 2019. URL: <https://github.com/rwightman/pytorch-image-models> (accessed 30.04.2023). <https://doi.org/10.5281/zenodo.4414861>
 - 21. Deng J., Dong W., Socher R., Li L.-J., Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database // 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA. 2009. P. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
 - 22. Малашин Р.О., Кадыков А.Б. Исследование обобщающих способностей сверточных нейронных сетей при формировании признаков, инвариантных к вращению // Оптический журнал. 2015. Т. 82. № 8. С. 24–32.
 - 23. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. June 27 — June 30. 2016. Las Vegas, Nevada, US. P. 770–778. <https://doi.org/10.48550/arXiv.1512.03385>
 - 24. Tan M., Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks // Proceedings of the 36th International Conference on Machine Learning. PMLR. 9–15 June 2019. Long Beach. California. USA. V. 97. P. 6105–6114. <https://doi.org/10.48550/arXiv.1905.11946>
 - 25. Murphy Kevin P. Machine learning: a probabilistic perspective. Cambridge, Massachusetts, USA: MIT Press, 2012. 1067 p.
 - 26. Hinton G. Coursera neural networks for machine learning. Lecture 6. 2018. [Электронный ресурс] Режим доступа: <https://www.coursera.org/learn/neural-networks-deep-learning> (дата обращения: 30.04.2023).
 - 27. Cubuk E.D., Zoph B., Shlens J., Le Q.V. RandAugment: Practical automated data augmentation with a reduced search space // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA. June 14–19 2020. P. 3008–3017. <https://doi.org/10.1109/CVPRW50498.2020.00359>
 - 28. Cubuk E.D., Zoph B., Mané D., Vasudevan V., Le Q.V. AutoAugment: Learning augmentation strategies from data // IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. June 16–20. 2019. P. 113–123. <https://doi.org/10.1109/CVPR.2019.00020>
 - 29. Hendrycks D., Mu N., Cubuk E.D., Zoph B., Gilmer J., Lakshminarayanan B. Augmix: A simple data processing method to improve robustness and uncertainty // arXiv preprint arXiv:1912.02781 (published: 17.02.2020). URL: <https://doi.org/10.48550/arXiv.1912.02781> (accessed 30.04.2023).
 - 15. Craig Boehman. How to create an infrared effect in Photoshop (published 28.06.2022). URL: <https://www.makeuseof.com/how-to-create-an-infrared-effect-in-photoshop/> (accessed 01.05.2023).
 - 16. Shorten C., Khoshgoftaar T.M. A survey on image data augmentation for deep learning // J Big Data. 2019. V. 6. № 60. P. 1–48. <https://doi.org/10.1186/s40537-019-0197-0>
 - 17. Haeberli P., Voorhies B. Image processing by linear interpolation and extrapolation // IRIS Universe Magazine. 1994. V. 28. P. 8–9.
 - 18. Gonzalez R.C., Woods R.E. Digital image processing (4th ed.). New York, NY: Pearson, 2018. P. 138–140.
 - 19. Leo Levi. Unsharp masking and related image enhancement techniques // Computer Graphics and Image Processing. 1974. V. 3. № 2. P. 163–177.
 - 20. Wightman R. PyTorch image models. GitHub repository. 2019. URL: <https://github.com/rwightman/pytorch-image-models>. (accessed 30.04.2023) <https://doi.org/10.5281/zenodo.4414861>
 - 21. Deng J., Dong W., Socher R., Li L.-J., Kai Li, Li Fei-Fei. ImageNet: A large-scale hierarchical image database // 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. 2009. P. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
 - 22. Malashin R.O., Kadykov A.B. Investigation of the generalizing capabilities of convolutional neural networks in forming rotationinvariant attributes // Journal of Optical Technology. 2015. V. 82. № 8. P. 509–515. <https://doi.org/10.1364/JOT.82.000509>
 - 23. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition. June 27 — June 30. 2016. Las Vegas, Nevada, US. P. 770–778. <https://doi.org/10.48550/arXiv.1512.03385>
 - 24. Tan M., Le Q. EfficientNet: Rethinking model scaling for convolutional neural networks // Proceedings of the 36th International Conference on Machine Learning. PMLR. 9–15 June 2019. Long Beach. California. USA. V. 97. P. 6105–6114. <https://doi.org/10.48550/arXiv.1905.11946>
 - 25. Murphy Kevin P. Machine learning: a probabilistic perspective. Cambridge, Massachusetts, USA: MIT Press, 2012. 1067 p.
 - 26. Hinton G. Coursera neural networks for machine learning. Lecture 6. 2018. URL: <https://www.coursera.org/learn/neural-networks-deep-learning> (accessed 30.04.2023).
 - 27. Cubuk E.D., Zoph B., Shlens J., Le Q.V. RandAugment: Practical automated data augmentation with a reduced search space // IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA. June 14–19 2020. P. 3008–3017. <https://doi.org/10.1109/CVPRW50498.2020.00359>
 - 28. Cubuk E.D., Zoph B., Mané D., Vasudevan V., Le Q.V. AutoAugment: Learning augmentation strategies from data // IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA. June 16–20. 2019. P. 113–123. <https://doi.org/10.1109/CVPR.2019.00020>
 - 29. Hendrycks D., Mu N., Cubuk E.D., Zoph B., Gilmer J., Lakshminarayanan B. Augmix: A simple data processing method to improve robustness and uncertainty // arXiv preprint arXiv:1912.02781 (published: 17.02.2020). URL: <https://doi.org/10.48550/arXiv.1912.02781> (accessed 30.04.2023).

- [Электронный ресурс] arXiv preprint arXiv:1912.02781 (дата обновления: 17 февраля 2020). <https://doi.org/10.48550/arXiv.1912.02781> (дата обращения 30.04.2023).
30. Zhong Zhun, Liang Zheng, Guoliang Kang, Shaozi Li, Yi Yang. Random erasing data augmentation // Proceedings of the AAAI conference on artificial intelligence. 2020. V. 34. № 7. P. 13001–13008. <https://doi.org/10.48550/arXiv.1708.04896>
30. Zhong Zhun, Liang Zheng, Guoliang Kang, Shaozi Li, Yi Yang. Random erasing data augmentation // Proceedings of the AAAI conference on artificial intelligence. 2020. V. 34. № 7. P. 13001–13008. <https://doi.org/10.48550/arXiv.1708.04896>

АВТОРЫ

Ермаченкова Маргарита Кареновна — студент, Университет ИТМО, 197101, Санкт-Петербург, Россия; <https://orcid.org/0009-0005-6760-759X>, ermachenkova.rita@mail.ru

Роман Олегович Малашин — канд. техн. наук, старший научный сотрудник, Институт физиологии им. И.П. Павлова РАН, 199034, Санкт-Петербург, Россия; доцент Санкт-Петербургского государственного университета аэрокосмического приборостроения, 190000, Санкт-Петербург, Россия; Scopus ID: 55789083900, <https://orcid.org/0000-0002-2493-839X>, malashinroman@mail.ru

Арина Андреевна Бойко — младший научный сотрудник, Институт физиологии им. И.П. Павлова РАН, 199034, Санкт-Петербург, Россия; ассистент Санкт-Петербургского государственного университета аэрокосмического приборостроения, 190000, Санкт-Петербург, Россия; Scopus ID: 57225010175, <https://orcid.org/0000-0001-7520-0056>, boikoaa@infran.ru

Статья поступила в редакцию 04.05.2023

Одобрена после рецензирования 13.06.2023

Принята к печати 18.08.2023

AUTHORS

Margarita K. Ermachenkova — a Student, ITMO University, 197101, Saint-Petersburg, Russia; <https://orcid.org/0009-0005-6760-759X>, ermachenkova.rita@mail.ru

Roman O. Malashin — PhD, Senior Research Fellow, Pavlov Institute of Physiology, Russian Academy of Sciences, 199034, Saint-Petersburg, Russia; Associate Professor, Saint-Petersburg State University of Aerospace Instrumentation, 190000, Saint-Petersburg, Russia; Scopus ID: 55789083900, <https://orcid.org/0000-0002-2493-839X>, malashinroman@mail.ru

Arina A. Boiko — Junior Research Fellow, Pavlov Institute of Physiology, Russian Academy of Sciences, 199034, Saint-Petersburg, Russia; Assistant, Saint-Petersburg State University of Aerospace Instrumentation, 190000, Saint-Petersburg, Russia; Scopus ID: 57225010175, <https://orcid.org/0000-0001-7520-0056>, boikoaa@infran.ru

The article was submitted to the editorial office 04.05.2023

Approved after review 13.06.2023

Accepted for publication 18.08.2023