

DOI: 10.17586/1023-5086-2023-90-01-37-48

УДК 004.93

Категоризация объектов и сцен нейронной сетью, входы которой предварительно обучены декодированию пространственных неоднородностей текстуры

Денис Викторович Явна¹✉, Виталий Вадимович Бабенко²,
Ольга Алексеевна Горбенкова³, Игорь Владимирович Плавельский⁴,
Виктория Дмитриевна Вороная⁵, Александр Сергеевич Столетний⁶

Южный федеральный университет, Ростов-на-Дону, Россия

¹yavna@fortran.su

<https://orcid.org/0000-0003-2895-5119>

²babenko@sfnedu.ru

<https://orcid.org/0000-0002-3750-1277>

³gorbenkova@sfnedu.ru

<https://orcid.org/0000-0002-3302-563X>

⁴nilpaga@yandex.ru

<https://orcid.org/0000-0003-1594-1842>

⁵vikt.voronaya@yandex.ru

<https://orcid.org/0000-0002-9984-2245>

⁶stoletniynk@inbox.ru

<https://orcid.org/0000-0001-7955-8696>

Аннотация

Предмет исследования. Исследование возможности использования нейросетевых моделей зрительных механизмов второго порядка в качестве входов нейросетей-классификаторов. Зрительные механизмы второго порядка отвечают за обнаружение пространственных неоднородностей контраста, ориентации и пространственной частоты в изображении. Эти механизмы традиционно рассматриваются исследователями зрительной системы как одно из звеньев начальной зрительной обработки; хорошо изучена их роль в восприятии текстур. **Цель работы** — проверить, будет ли предварительное обучение входных модулей классификатора демодуляции пространственных модуляций градиентов яркости способствовать категоризации объектов и сцен. **Метод.** В качестве основного метода применялось нейросетевое моделирование. На первом этапе исследования была сгенерирована выборка текстурных изображений, используемая для обучения нейросетевых моделей зрительных механизмов второго порядка, и проведено обучение этих моделей. На втором этапе были подготовлены выборки объектов и сцен, на которых было проведено обучение сетей-классификаторов. На входе этих сетей располагались ранее обученные модели зрительных механизмов второго порядка с замороженными весами. **Основные результаты.** Информации второго порядка, представленной в виде карты мгновенных значений функции модуляции контраста, ориентации и пространственной частоты в изображении, может быть достаточно для идентификации только некоторых классов сцен. В целом, в рамках предложенных нейросетевых архитектур использование значений функции модуляции градиентов яркости для решения задачи классификации объектов оказалось малоэффективным. Таким образом, не подтвердилась гипотеза о том, что зрительные фильтры второго порядка кодируют признаки, позволяющие идентифицировать объект. Этот результат делает необходимым проверку альтернативной гипотезы, согласно которой роль фильтров второго порядка ограничивается задачей построения карт салиентности, а сами фильтры являются окнами, через которые поступает информация с выходов фильтров первого порядка. **Практическая значимость.** Проведена оценка возможностей использования моделей зрительных механизмов второго порядка в системах компьютерного зрения.

Ключевые слова: механизмы зрительной обработки, текстура, сверточная нейронная сеть, нейросеть-классификатор, машинное зрение

Благодарность: исследование выполнено при финансовой поддержке РФФИ, проект № 18-29-22001 МК «Исследование нейрокогнитивных технологий управления вниманием и формирования ментальных репрезентаций визуального интернет-контента».

Ссылка для цитирования: Явна Д.В., Бабенко В.В., Горбенкова О.А., Плавельский И.В., Вороная В.Д., Столетний А.С. Категоризация объектов и сцен нейронной сетью, входы которой предварительно обучены декодированию пространственных неоднородностей текстуры // Оптический журнал. 2023. Т. 90. № 1. С. 37–48. <http://doi.org/10.17586/1023-5086-2023-89-11-37-48>

Коды OCIS: 100.4996, 330.5370

Categorization of objects and scenes by a neural network whose input modules are pre-trained to decode spatial texture inhomogeneities

DENIS V. YAVNA¹✉, VITALY V. BABENKO²,
OLGA A. GORBENKOVA³, IGOR V. PLAVELSKY⁴,
VIKTORIA D. VORONAYA⁵, ALEKSANDR S. STOLETNIY⁶

Southern Federal University, Rostov-on-Don, Russia

¹yavna@fortran.su

²babenko@sfedu.ru

³gorbenkova@sfedu.ru

⁴nilpga@yandex.ru

⁵vikt.voronaya@yandex.ru

⁶stoletniynk@inbox.ru

Abstract

Scope of research. Investigation of the possibility of using neural network models of second-order visual mechanisms as input data for neural network classifiers. Second-order visual mechanisms make it possible to detect spatial inhomogeneities in contrast, orientation, and spatial frequency in an image. These mechanisms are traditionally considered by visual researchers as one of the stages of early visual processing; their role in the perception of textures has been well studied. **The purpose** of the work is to study whether the use of classifier input modules previously trained to demodulate spatial modulations of brightness gradients will contribute to the categorization of objects and scenes. **Method.** Neural network modeling was used as the main method. At the first stage of the study, a set of texture images was generated, which is used to train neural network models of second-order visual mechanisms, and these models were trained. At the second stage, samples of objects and scenes were prepared, on which classifier networks were trained. Previously trained models of second-order visual mechanisms with frozen weights were placed at the input of these networks. **Main results.** The second order information, presented as a map of instantaneous values of the modulation function of contrast, orientation and spatial frequency in the image, may be sufficient to identify only some classes of scenes. In general, within the framework of the proposed neural network architectures, the use of modulation function values for solving the problem of object classification turned out to be ineffective. Thus, the hypothesis that second-order visual filters encode features that allow identifying an object was not confirmed. This result makes it necessary to test an alternative hypothesis that the role of second-order filters is limited to participation in the construction of saliency maps, and the filters themselves are windows through which information comes from the outputs of first-order filters. **Practical significance.** The possibility of using second-order models of visual mechanisms in computer vision systems was assessed.

Keywords: visual processing mechanisms, texture, convolutional neural network, classifier neural network, machine vision

Acknowledgment: the study was financially supported by the Russian Foundation for Basic Research, project № 18-29-22001 МК "An investigation of neurocognitive technologies of attentional control and formation of mental representations of visual web content".

For citation: Yavna D.V., Babenko V.V., Gorbenkova O.A., Plavelsky I.V., Voronaya V.D., Stoletniy A.S. Categorization of objects and scenes by a neural network whose input modules are pre-trained to decode spatial texture inhomogeneities [in Russian] // *Opticheskii Zhurnal*. 2023. V. 90. № 1. P. 37–48. <http://doi.org/10.17586/1023-5086-2023-90-01-37-48>

OCIS codes: 100.4996, 330.5370

ВВЕДЕНИЕ

Долгое время считалось, что преаттентивная зрительная обработка, осуществляемая параллельно по полю зрения, завершается выделением во входном изображении градиентов яркости разной локализации, ориентации и пространственной частоты. При этом пространственное объединение таких локальных признаков осуществляется последовательно с помощью механизмов высшего уровня [1]. Однако на рубеже нынешнего столетия появились убедительные свидетельства существования еще одного преаттентивного уровня, на котором происходит пространственное объединение локальной информации [2–4 и др.]. Этот этап обработки получил название «зрительные механизмы второго порядка» [5]. Далее будем придерживаться следующих устоявшихся определений: информация первого порядка — градиенты яркости, информация второго порядка — пространственные модуляции градиентов яркости [6]. Именно объединение выходов фильтров первого порядка (постстриарным нейронам) позволяет фильтрам второго порядка (постстриарных нейронов) обнаруживать пространственные изменения градиентов яркости (их контраста, ориентации или пространственной частоты). Схема, объясняющая принцип работы механизмов второго порядка (рис. 1), выглядит следующим образом: выпрямленные выходы простых нейронов проекционной коры формируют вход нейронов второго порядка, имеющих оппонентно организованные рецептивные поля значительно большего размера. Это обеспечивает их настройку на более низкие пространственные частоты огибающей; пиковые пространственно-частотные настройки фильтров второго порядка в 8–16 раз ниже [7] настроек фильтров первого порядка. Инвариантность зрительных механизмов второго порядка к масштабу и ориентации достигается благодаря пространственно-частотной и ориентационной многоканальности [8], а также связи частотных и ориентационных настроек механизмов первого и второго порядков.

Итак, сначала простые нейроны-фильтры проекционной коры пропускают сигналы градиента яркости на определенной пространственной частоте, а затем клетки второго порядка фильтруют получившийся выпрямленный результат на сниженной частоте. Эта схема фактически с минимальными отличиями воспроизводит принцип работы хорошо известного радиолюбителям приемника сигнала, модулированного по амплитуде (АМ). Он обнаруживает изменения огибающей посредством настроенного на частоту несущей полосового фильтра, операции выпрямления и фильтра низких частот. Основные отличия схемы зрительных механизмов второго порядка заключаются лишь в том, что нервные клетки второго порядка выполняют функцию именно полосовых фильтров, а не фильтров низких частот. Кроме того, АМ-детектор и зрительный механизм второго порядка имеют дело с сигналами, отличающимися пространственной размерностью: изменения яркости и контраста в изображении двумерны, а мгновенные значения радиосигнала одномерны (меняются во времени).

При обработке изображений схема, приведенная на рис. 1, особенно хорошо подходит для объяснения принципов детектирования модуляций контраста, с которых и начались исследования зрительных механизмов второго порядка. Их обнаружение наблюдателями не могло быть объяснено в терминах популярной тогда теории множественных детекторов [9], основанной на классических (*David Hubel*, *Torsten Wiesel*) исследованиях свойств клеток проекционной коры и хорошо подходящей для объяснения принципов обнаружения модуляций яркости.

Однако классическая схема «фильтрация — выпрямление — фильтрация» не объясняет механизмы восприятия пространственных неоднородностей ряда локальных характеристик изображения, отличных от контраста. Так, механизм, имеющий в своей основе именно эту схему, будет лучше всего обнаруживать

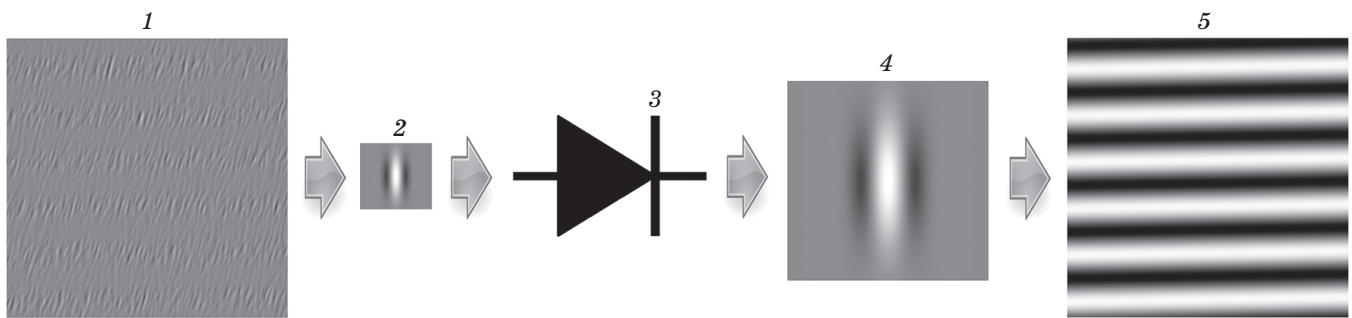


Рис. 1. Теоретическая схема модели зрительных механизмов второго порядка. 1 — исходное изображение, модулированное по контрасту, 2 — фильтр первого порядка, 3 — выпрямление, 4 — фильтр второго порядка, 5 — результат демодуляции

Fig. 1. Theoretical scheme of the second order visual mechanisms. (1) The original contrast modulated image, (2) first order filter, (3) rectification block, (4) second order filter, (5) the output

модуляции контраста. При этом модуляции ориентации и пространственной частоты также могут быть им обнаружены, но предполагаемая система фильтров не имеет средств для различения модуляций. В настоящее время имеется достаточное количество экспериментальных данных, позволяющих рассматривать механизмы обнаружения разных модуляций как относительно независимые [10–13 и др.], причем функционирующие автоматически на предвнимательном уровне [14]. Некоторыми авторами были предложены соответствующие корректировки классической модели [12, 15]. Они не затрагивают общий принцип ее устройства и касаются лишь особенностей организации рецептивных полей фильтров второго порядка, реализующих специфичность механизмов к пространственно модулируемой характеристике изображения. Специфичность механизмов второго порядка успешно моделируется также путем обучения сверточных нейронных сетей [16, 17].

Традиционно зрительные механизмы второго порядка рассматривались исследователями в первую очередь в контексте проблематики обнаружения и различения текстур. В настоящее время представления о роли этих механизмов существенно расширились: в ряде работ показано их участие в восстановлении трехмерной формы объектов, в фигурофоновой сегментации, а также в управлении зрительным вниманием [18, 19]. Особенно перспективен взгляд на механизмы второго порядка как на инструмент нахождения во

входном изображении пространственных неоднородностей в качестве «областей интереса». В этой связи, говоря о возможной роли механизмов второго порядка в зрительном восприятии, можно рассматривать две возможные гипотезы: 1) фильтры второго порядка, обнаруживая пространственные неоднородности, кодируют их характеристики и передают на лежащие выше уровни обработки эту информацию в виде вторичных зрительных признаков, 2) роль фильтров второго порядка ограничивается нахождением областей интереса, а сами фильтры играют роль окон внимания, через которые на последующие уровни обработки передается информация с выходов фильтров первого порядка.

Цель представленной работы — проверить первую гипотезу, ответив на вопрос, будет ли предварительное обучение входных модулей нейросетевого классификатора демодуляции пространственных модуляций градиентов яркости способствовать категоризации объектов и сцен. Положительный результат будет свидетельствовать о том, что информации о параметрах пространственной модуляции, кодируемой моделями механизмов второго порядка, может быть достаточно для классификации их содержимого.

Применялся метод компьютерного моделирования: нейронную сеть, вход которой образован предварительно созданными и обученными моделями зрительных механизмов второго порядка, обучали классификации объектов 20-ти классов. Авторы настоящей статьи отдают себе отчет в том, что такого рода

модели являются лишь грубым приближением в попытке понять принципы работы живого мозга. Однако результаты моделирования могут оказаться полезными в понимании информационных процессов, осуществляемых определенными мозговыми механизмами или структурами.

МОДЕЛИРОВАНИЕ ЗРИТЕЛЬНЫХ МЕХАНИЗМОВ ВТОРОГО ПОРЯДКА

Модели механизмов второго порядка представляли собой полностью сверточные автокодировщики, обученные демодуляции пространственных модуляций контраста, ориентации и пространственной частоты на искусственных текстурах. Кодировочная часть каждой модели помимо входного включала 3 сверточных слоя, содержащих по 32 «нейрона» (ядра свертки) размерами соответственно 3×3 , 3×3 и 5×5 . После каждого сверточного слоя располагался слой понижающей дискретизации по максимальному значению 2×2 . Выходной слой кодировщика содержал 8 ядер 5×5 . Конфигурация входного слоя декодировщика соответствовала выходному слою кодировщика. Декодировочная часть также содержала 3 сверточных слоя по 32 ядра размерами 5×5 , 5×5 и 3×3 , после каждого из которых выполнялась повышающая дискретизация 2×2 ; выходной слой был сверточным

с одним ядром 3×3 . Для всех слоев использовалась функция блока линейной ректификации (ReLU). Ширина и высота входного слоя кодировщика не были определены, что позволяет потенциально использовать автокодировщик для обработки изображений любого размера. Схема автокодировщика приведена на рис. 2.

Обучение, валидация и тестирование моделей зрительных механизмов второго порядка проводились на искусственных модулированных текстурах, созданных путем случайного наложения микропаттернов Габора. Алгоритм синтеза таких текстур подробно описан в публикации [21], он основывается на внесении изменений (огибающая) в параметры синтезируемых микропаттернов (несущая) в зависимости от их местоположения. В текстурах случайно варьировались модулируемая локальная характеристика (контраст, ориентация или пространственная частота), наклон осей несущей и огибающей, длины волн несущей и огибающей, фазовый сдвиг огибающей, средняя яркость и общий контраст. В качестве модуляторов использовались непрерывные тригонометрические функции. Пиксельные размеры текстур составляли 256×256 , пиксели представляли градации серого.

Всего было сгенерировано 15000 изображений — по 5000 на каждую модулируемую характеристику. Из этих 5000 для собственно обучения брались 3500, остальные же

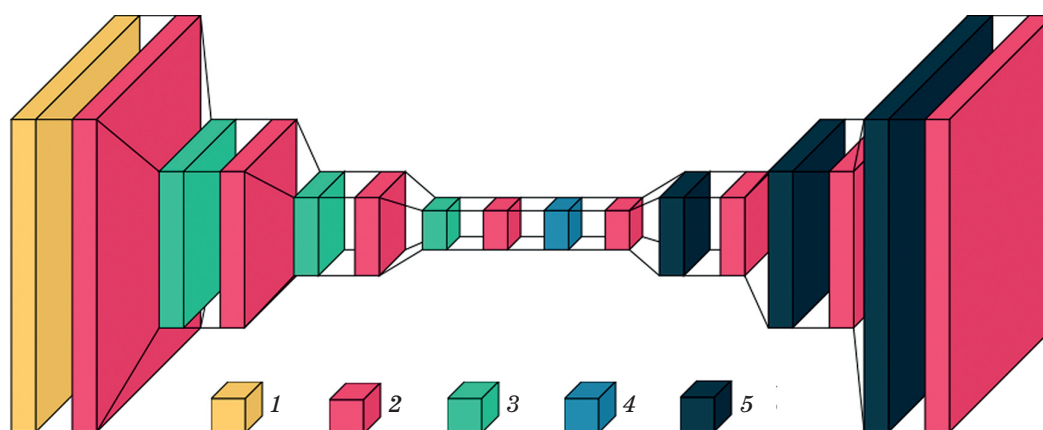


Рис. 2. Схема полностью сверточного автокодировщика, реализующего модель зрительного механизма второго порядка, специфичного к модулируемой характеристике изображения. 1 — входной слой, 2 — сверточный слой, 3 — слой понижающей дискретизации (пулинг), 4 — входной слой декодировщика, 5 — слой повышающей дискретизации. Построено с использованием библиотеки visualkeras [20]

Fig. 2. Scheme of a fully convolutional autoencoder that implements a second-order visual mechanism model specific to the modulated image feature. (1) Input layer, (2) convolutional layer, (3) pooling layer, (4) decoder input layer, (5) upsampling layer. Built using the visualkeras library [20]

использовались для валидации и тестирования. Таким образом, количество изображений в обучающей выборке составило 10500. При обучении на вход подавались текстуры, на выход — карты значений функции модуляции в каждой точке текстуры, сохраняемые при их синтезе и кодируемые таким образом, что минимуму соответствовал 0, максимуму — 255. Процесс обучения был направлен на подбор параметров модели, обеспечивающих наибольшее сходство между ожидаемыми значениями функций модуляции и результатами

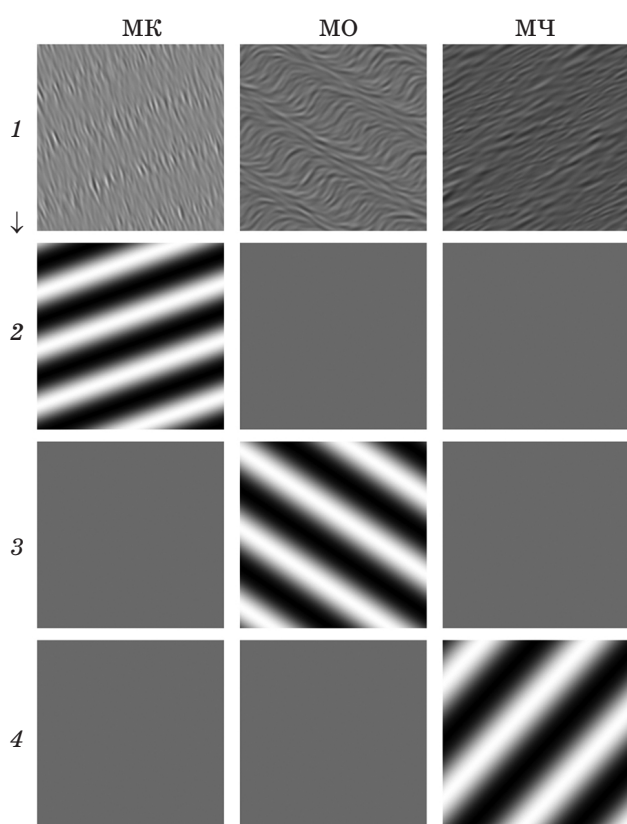


Рис. 3. Примеры обучающих текстур и ожидаемых результатов обучения. 1 — модулированные текстуры: МК — по контрасту, МО — по ориентации, МЧ — по пространственной частоте; ожидаемые результаты для модели, специфичной к модуляциям контраста (2), для модели, специфичной к модуляциям ориентации (3), для модели, специфичной к модуляциям пространственной частоты (4)

Fig. 3. Examples of training textures and expected results. (1) Modulated textures: (МК) contrast-modulated, (МО) orientation-modulated, (МЧ) spatial frequency-modulated; expected results: (2) for the model specific to modulations of contrast, (3) for the model specific to modulations of orientation, (4) for the model specific to modulations of spatial frequency

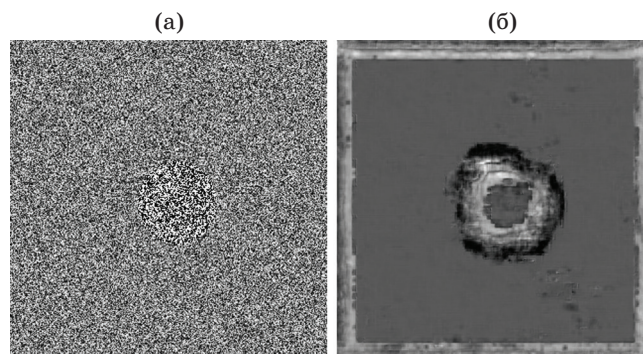


Рис. 4. Реакция модели (б), специфичной к модуляциям контраста, на однородную текстуру с высококонтрастной областью (а)

Fig. 4. The response of (б) the contrast modulation specific model to (а) a uniform texture with a highcontrast region

операций кодирования-декодирования. Примеры обучающих текстур и результатов, ожидаемых от каждой специфичной к модулируемой характеристике модели, приведены на рис. 3.

Несмотря на то, что модели обучались на текстурах, модулированных гармоническими функциями, в результате обучения вырабатывался достаточно универсальный набор фильтров, обеспечивающий реакцию моделей на шаблоны, расположенные в произвольных участках изображения и имеющие разные ориентацию, размер и др. На рис. 4 демонстрируется реакция модели, обученной демодуляции пространственных модуляций контраста, на высококонтрастное шумовое пятно на низкоконтрастном фоне.

ОБУЧЕНИЕ НЕЙРОСЕТЕЙ-КЛАССИФИКАТОРОВ

В качестве классификатора использовалась сеть MobileNetV2 [22]. Входной слой принимал изображения, масштабированные до пиксельного размера 224×224 , что связано с наличием в сети полносвязной части. Модели зрительных механизмов второго порядка располагались параллельно после входа сети, причем их веса при обучении замораживались. После слоя объединения следовали слои MobileNetV2, инициализированные весами ImageNet, 86 начальных слоев были заморожены, коэффициент «ширины» сети alpha,

задающий количество фильтров на слое, составлял 0,75. На выходе располагался полносвязный слой из 20 «нейронов» с логистической активационной функцией для многомерного случая, кодирующих принадлежность к классу (рис. 5а). Слой прореживания, расположенный перед выходным, в итоговой версии модели имел очень малый уровень и фактически не влиял на обучение. В контрольной серии исследования проводилось обучение классификатора без моделей зрительных механизмов второго порядка на входе (рис. 5б).

Для обучения сети-классификатора использовалась коллекция изображений LSUN [23]. Обучение проводилось на одном ускорителе Nvidia Tesla M40 12 Gb, что делало практически невозможным использование всего объема коллекции, содержащей 51181482 изображений. Кроме того, LSUN содержит значительное количество неверно классифицированных объектов: по оценкам авторов коллекции оно составляет около 10% [23]. Была сделана слу-

чайная выборка, в которой каждая из 20-ти категорий коллекции представляла 11000 обучающих, 2200 валидационных и 2200 тестовых изображений. Обучение классификатора стало возможным только после ручной очистки выборки от ошибочно категоризованных объектов; в итоговую выборку вошли 181717 изображений для обучения, 34601 — для валидации и 36360 — для тестирования (82% от исходной случайной выборки).

Обучение проводилось в 200 эпох, причем производительность модели оценивалась на валидирующем наборе после каждой эпохи по показателю точности классификации. Если производительность превышала наилучшую из ранее достигнутых, веса модели сохранялись.

Сеть, имеющая модели механизмов второго порядка на входе, не выработала универсальных репрезентаций признаков класса и продемонстрировала сильное переобучение: несмотря на близкую к 100% точность на обучающей выборке, она классифицировала тестовые объекты на случайном уровне 5% (рис. 6). При этом в контрольной серии был достигнут вполне ожидаемый от MobileNetV2 уровень

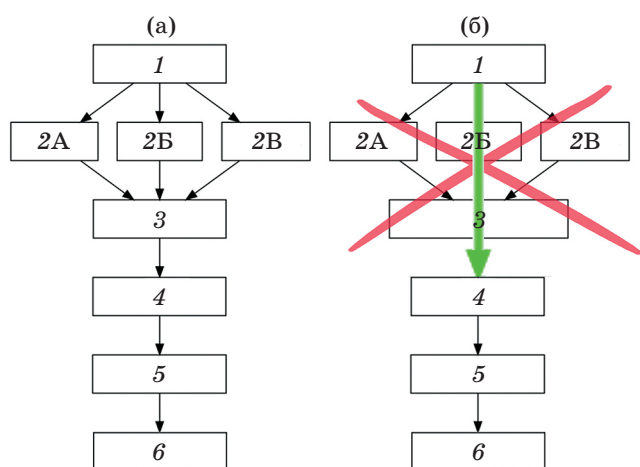


Рис. 5. Блок-схемы классификаторов на основе MobileNetV2 с моделями зрительных механизмов второго порядка на входе (а) и без таковых (б). 1 — входной слой, 2А, 2Б, 2В — модели, специфичные соответственно к модуляциям контраста, ориентации и пространственной частоты, 3 — слой конкатенации, 4 — модель MobileNetV2, 5 — слой прореживания, 6 — выходной слой

Fig. 5. Flowcharts of classifiers based on MobileNetV2 with the models of second order visual mechanisms at (a) the input and (b) without them. (1) Input layer, (2A), (2B), (2B) models specific to contrast, orientation and spatial frequency modulations respectively, (3) concatenation layer, (4) MobileNetV2 model, (5) dropout layer, (6) output layer

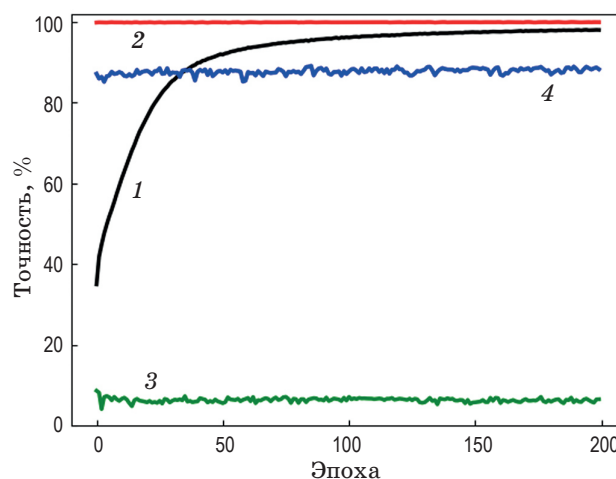


Рис. 6. Динамика точности классификаторов в основной и контрольной сериях обучения. 1 — обучение в основной серии, 2 — обучение в контрольной серии, 3 — валидация в основной серии, 4 — валидация в контрольной серии

Fig. 6. Classifier accuracy graphs for the main and control series of training. (1) Training in the main series, (2) training in the control series, (3) validation in the main series, (4) validation in the control series

обучения: точность составляла 100% на обучающей и 84% на тестовой выборках. Полученные результаты свидетельствуют не в пользу гипотезы о достаточности признаков второго порядка для классификации объектов.

Однако можно ли полностью исключить участие зрительных механизмов второго порядка в процессах классификации вообще? В ходе работы было проведено похожее исследование с изображениями сцен. Сеть с моделями зрительных механизмов на входе идентифицировала ряд классов с высокой точностью. Изображения были отобраны из коллекции SUN [24], содержащей всего 397 классов сцен. Отбор не был полностью случайным — каждый класс сцен должен был содержать также объекты одной базовой категории: автостроды, беседки, телефонные будки и др. К сожа-

лению, из-за специфики коллекции (большое число классов и малое число образцов) обучающая выборка содержала только 7432, валидационная — 1100 картинок. Всего было отобрано 20 классов.

Модели-классификаторы состояли из сверточной и полносвязной частей (рис. 7). Входной слой принимал изображения, масштабированные до пиксельного размера 256×256 . На их входе располагались три параллельные ветви, каждая из которых начиналась с предварительно обученной модели механизмов второго порядка, специфичной к модуляциям контраста, ориентации или пространственной частоты. Веса моделей замораживались. Каждая из ветвей включала один сверточный слой в 32 ядра размером 7×7 и слой пулинга 2×2 . Отдельные ветви соединялись в общую, со-

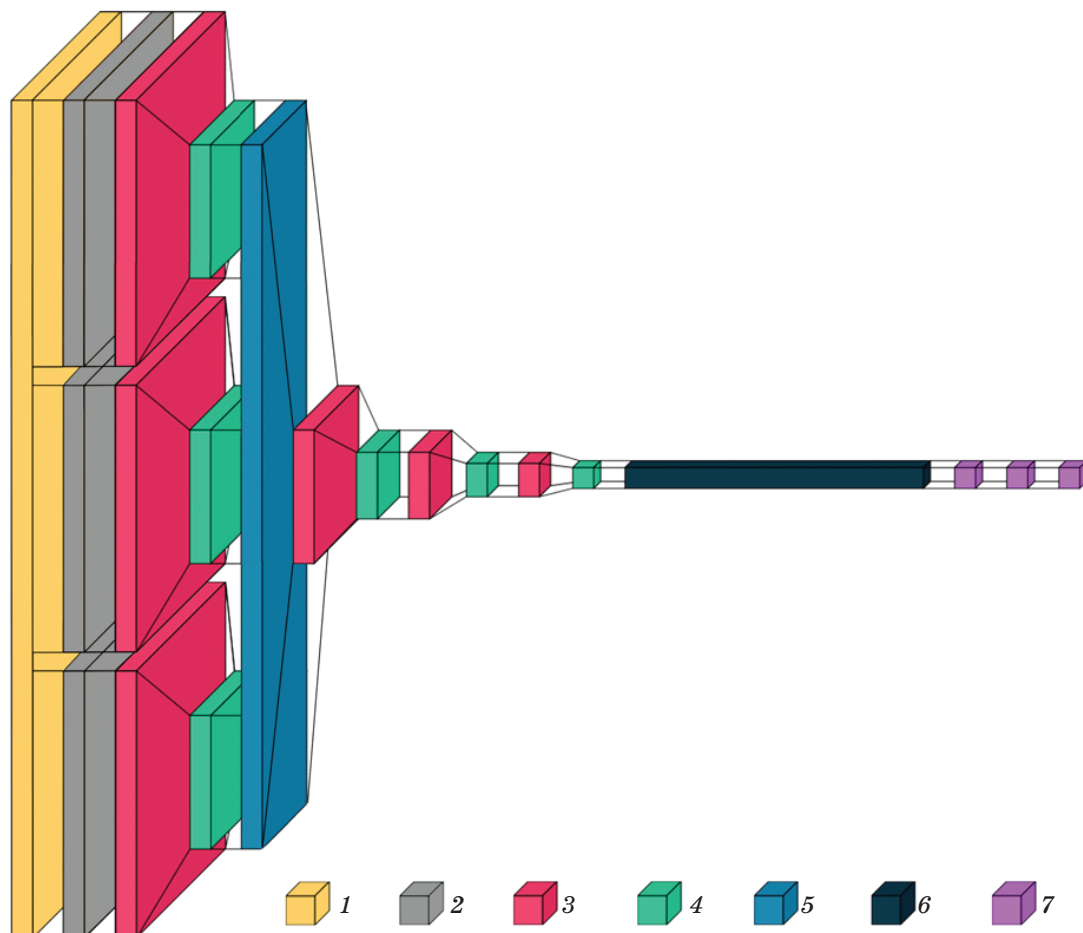


Рис. 7. Схема сети, использованной для классификации сцен. 1 — входной слой, 2 — слой механизмов второго порядка, 3 — сверточный слой, 4 — слой понижающей дискретизации, 5 — конкатенация, 6 — слой выпрямления, 7 — полносвязный слой

Fig. 7. Scheme of the network used to classify scenes. (1) Input layer, (2) layer of the second order mechanisms, (3) convolutional layer, (4) pooling layer, (5) concatenation layer, (6) flatten layer, (7) fully connected layer

державшую последовательно сверточные слои в 128, 64 и 16 ядер 7×7 , после каждого из которых следовал пулинг 2×2 . Сверточная часть соединялась с тремя полносвязными слоями в 80, 40 и 20 «нейронов». Принадлежность изображения к классу кодировалась активацией одного из 20 «нейронов» выходного слоя. В моделях механизмов второго порядка, как и во всех внутренних слоях классификаторов, применялась функция блока линейной ректификации (ReLU); в выходном слое использовалась логистическая активационная функция для многомерного случая. В сети, обучаемой в целях контроля, предобученных моделей на входе не было (рис. 7).

Обучение моделей проводилось в 200 эпох, наилучшие достигнутые результаты сохранялись после каждой эпохи. Оказалось, что уже примерно к 20-й эпохе модели начинают распознавать несколько классов валидирующих изображений, достигая на обучающем наборе 100% точности, что связано с эффектом переобучения. После обучения были определены следующие классы сцен с объектами, идентифицируемые сетями на тестовой выборке на уровне 20% и более (при уровне случайного угадывания, равном 5%):

– с механизмами второго порядка на входе — автострады (56%), башни (56%), лестницы (28%) ветряные мельницы (28%) и айсберги (20%). При этом 5 классов не идентифицировались вообще;

– без механизмов второго порядка на входе — склады бутылок (88%), виноградники (84%), коралловые рифы (80%), беседки (76%), автострады (48%), башни (44%), заправочные станции (32%), телефонные будки (32%), ветряные мельницы (28%), лестницы (24%). Наихудший показатель идентификации составлял 8%.

Общая производительность сети находится на уровне 13,9% при производительности контрольного классификатора, равной 30,4%.

Таким образом, производительность контрольной модели оказалась большей. Однако

важно отметить, что информации второго порядка оказывается достаточно для уверенной идентификации ряда классов сцен, содержащих объекты. Дополнительно считаем важным указать, что сеть, основанная на классификаторе MobileNetV2 и имеющая модели зрительных механизмов на входе, также оказалась способна классифицировать сцены, но с очень низкой производительностью (19,6% на тестовой выборке при производительности контрольного классификатора без предобученных моделей на входе, составившей 86%).

ЗАКЛЮЧЕНИЕ

Предварительное обучение сети обнаружению пространственных модуляций контраста, ориентации и пространственной частоты градиентов яркости не способствует решению задачи категоризации объектов, хотя и делает возможным категоризацию некоторых классов сцен. Вероятно, это связано с наличием в изображениях сцен характерных признаков, по сути являющихся текстурными. Таким образом, информации о значениях амплитуды модуляции градиентов яркости, вычисляемой моделями механизмов второго порядка в каждой точке изображения, для уверенной классификации оказывается в общем случае недостаточно. Полученный результат не подтверждает первую из сформулированных выше гипотез.

Следующим этапом исследования будет проверка второй гипотезы, согласно которой зрительные механизмы второго порядка не кодируют параметры пространственных неоднородностей в качестве признаков второго порядка, а играют роль «ворот внимания», благодаря которым зрительная система направляется и считывает необходимый объем информации из областей наибольшей пространственной неоднородности [25–27]. В планируемом исследовании зрительные механизмы второго порядка будут использоваться не непосредственно как входы классификатора, а для поиска таких областей.

СПИСОК ИСТОЧНИКОВ

1. Treisman A.M., Gelade G. A feature-integration theory of attention // *Cognitive Psychology*. 1980. V. 12. № 1. P. 97–136.
2. Sutter A., Beck J., Graham N.V. Contrast and spatial variables in texture segregation: Testing

REFERENCES

1. Treisman A.M., Gelade G. A feature-integration theory of attention // *Cognitive Psychology*. 1980. V. 12. № 1. P. 97–136.
2. Sutter A., Beck J., Graham N.V. Contrast and spatial variables in texture segregation: Testing

- a simple spatial-frequency channels model // *Percept. Psychophys.* 1989. V. 46. № 4. P. 312–332.
3. Mareschal I., Baker C.L. Temporal and spatial response to second-order stimuli in cat area 18 // *J. Neurophysiol.* 1998. V. 80. № 6. P. 2811–2823. <https://doi.org/10.1152/jn.1998.80.6.2811>
 4. Landy M.S., Oruç I. Properties of second-order spatial frequency channels // *Vision Res.* 2002. V. 42. № 19. P. 2311–2329. [https://doi.org/10.1016/s0042-6989\(02\)00193-1](https://doi.org/10.1016/s0042-6989(02)00193-1)
 5. Derrington A. Second-order visual processing // *Optics & Photonics News.* 2001. V. 12. № 1. P. 18. <https://doi.org/10.1364/OPN.12.1.000018>
 6. Huang P.-C., Chen C.-C. A comparison of pedestal effects in first- and second-order patterns // *J. Vision.* 2014. V. 14. № 1. P. 9–9. <https://doi.org/10.1167/14.1.9>
 7. Sutter A., Sperling G., Chubb C. Measuring the spatial frequency selectivity of second-order texture mechanisms // *Vision Res.* 1995. V. 35. № 7. P. 915–924. [https://doi.org/10.1016/0042-6989\(94\)00196-s](https://doi.org/10.1016/0042-6989(94)00196-s)
 8. Шелепин Ю.Е., Чихман В.Н., Вахрамеева О.А., Пронин С.В., Фореман Н., Пэсмор П. Инвариантность зрительного восприятия // *Экспериментальная психология.* 2008. Т. 1. № 1. С. 7–33. <https://elibrary.ru/item.asp?id=3019577>
 9. Graham N.V. Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years // *Vision Res.* 2011. V. 51. № 13. P. 1397–1430. <https://doi.org/10.1016/j.visres.2011.02.007>
 10. Babenko V.V., Ermakov P.N. Specificity of brain reactions to second-order visual stimuli // *Vis Neurosci.* 2015. V. 32. P. E011. <https://doi.org/10.1017/S0952523815000085>
 11. Ellemberg D., Allen H.A., Hess R.F. Second-order spatial frequency and orientation channels in human vision // *Vision Res.* 2006. V. 46. № 17. P. 2798–2803. <https://doi.org/10.1016/j.visres.2006.01.028>
 12. Kingdom F.A.A., Prins N., Hayes A. Mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism // *Vis Neurosci.* 2003. V. 20. № 1. P. 65–76. <https://doi.org/10.1017/s0952523803201073>
 13. Schofield A., Cruickshank A. Transfer of tilt after-effects between second-order cues // *Spatial Vis.* 2005. V. 18. № 4. P. 379–397. <https://doi.org/10.1163/1568568054389624>
 14. Wolfe J.M. Visual search. Attention. Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis, 1998. P. 13–73.
 15. Бабенко В.В., Явна Д.В. Конкуренция за внимание пространственных модуляций градиентов яркости // *Российский психологический журнал.* 2018. Т. 15. № 3. С. 160–189. <https://doi.org/10.21702/rpj.2018.3.8>
 - a simple spatial-frequency channels model // *Percept. Psychophys.* 1989. V. 46. № 4. P. 312–332.
 3. Mareschal I., Baker C.L. Temporal and spatial response to second-order stimuli in cat area 18 // *J. Neurophysiol.* 1998. V. 80. № 6. P. 2811–2823. <https://doi.org/10.1152/jn.1998.80.6.2811>
 4. Landy M.S., Oruç I. Properties of second-order spatial frequency channels // *Vision Res.* 2002. V. 42. № 19. P. 2311–2329. [https://doi.org/10.1016/s0042-6989\(02\)00193-1](https://doi.org/10.1016/s0042-6989(02)00193-1)
 5. Derrington A. Second-order visual processing // *Optics & Photonics News.* 2001. V. 12. № 1. P. 18. <https://doi.org/10.1364/OPN.12.1.000018>
 6. Huang P.-C., Chen C.-C. A comparison of pedestal effects in first- and second-order patterns // *J. Vision.* 2014. V. 14. № 1. P. 9–9. <https://doi.org/10.1167/14.1.9>
 7. Sutter A., Sperling G., Chubb C. Measuring the spatial frequency selectivity of second-order texture mechanisms // *Vision Res.* 1995. V. 35. № 7. P. 915–924. [https://doi.org/10.1016/0042-6989\(94\)00196-s](https://doi.org/10.1016/0042-6989(94)00196-s)
 8. Shelepin Yu.E., Chikhman V.N., Vakhrameeva O.A., Pronin S.V., Foreman N., Pasmore P. Invariance of visual perception [in Russian] // *Experimental Psychology (Russia).* 2008. V. 1. № 1. P. 7–33. <https://elibrary.ru/item.asp?id=13019577>
 9. Graham N.V. Beyond multiple pattern analyzers modeled as linear filters (as classical V1 simple cells): Useful additions of the last 25 years // *Vision Res.* 2011. V. 51. № 13. P. 1397–1430. <https://doi.org/10.1016/j.visres.2011.02.007>
 10. Babenko V.V., Ermakov P.N. Specificity of brain reactions to second-order visual stimuli // *Vis Neurosci.* 2015. V. 32. P. E011. <https://doi.org/10.1017/S0952523815000085>
 11. Ellemberg D., Allen H.A., Hess R.F. Second-order spatial frequency and orientation channels in human vision // *Vision Res.* 2006. V. 46. № 17. P. 2798–2803. <https://doi.org/10.1016/j.visres.2006.01.028>
 12. Kingdom F.A.A., Prins N., Hayes A. Mechanism independence for texture-modulation detection is consistent with a filter-rectify-filter mechanism // *Vis Neurosci.* 2003. V. 20. № 1. P. 65–76. <https://doi.org/10.1017/s0952523803201073>
 13. Schofield A., Cruickshank A. Transfer of tilt after-effects between second-order cues // *Spatial Vis.* 2005. V. 18. № 4. P. 379–397. <https://doi.org/10.1163/1568568054389624>
 14. Wolfe J.M. Visual search. Attention. Hove, England: Psychology Press/Erlbaum (UK) Taylor & Francis, 1998. P. 13–73.
 15. Babenko V.V., Yavna D.V. Competition for attention among spatial modulations of brightness gradients [in Russian] // *Russian Psychological J.* 2018. V. 15. № 3. P. 160–189. <https://doi.org/10.21702/rpj.2018.3.8>

16. Yavna D.V., Babenko V.V., Ikonopistseva K.A. Neural network models of second order visual filters//NeuralNetworksandNeurotechnologies. St. Petersburg, Russia: BBM, 2019. P. 198–203.
17. Явна Д.В., Бабенко В.В., Столетний А.С., Щетинина Д.П., Алексеева Д.С. Различение и декодирование пространственных модуляций текстур многослойными сверточными нейросетями // Вестник РФФИ. 2019. № 4(104). С. 94–104. <https://doi.org/10.22204/2410-4639-2019-104-04-94-104>
18. Frey H.-P., König P., Einhäuser W. The role of first- and second-order stimulus features for human overt attention // Percept Psychophys. 2007. V. 69. № 2. P. 153–161. <https://doi.org/10.3758/bf03193738>
19. Johnson A., Zarei A. Second-order saliency predicts observer eye movements when viewing natural images // J. Vision. 2010. V. 10. № 7. P. 526–526. <https://doi.org/10.1167/10.7.526>
20. Gavrikov P. Программное обеспечение Visualkeras [режим доступа]: <https://github.com/paulgavrikov/visualkeras>
21. Prins N., Kingdom F.A.A. Detection and discrimination of texture modulations defined by orientation, spatial frequency, and contrast // JOSA. A. 2003. V. 20. № 3. P. 401. <https://doi.org/10.1364/JOSAA.20.000401>
22. Sandler M., Howard A., Zhu M., et al. MobileNetV2: Inverted residuals and linear bottlenecks // arXiv:1801.04381 [cs]. 2019. Retrieved from <http://arxiv.org/abs/1801.04381>
23. Yu F., Seff A., Zhang Y., et al. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop // arXiv:1506.03365 [cs]. 2016. Retrieved from <http://arxiv.org/abs/1506.03365>
24. Xiao J., Hays J., Ehinger K.A., et al. SUN database: Large-scale scene recognition from abbey to zoo // 2010 IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition. 2010. P. 3485–3492. <https://doi.org/10.1109/CVPR.2010.5539970>
25. Victor J.D., Conte M.M., Chubb C.F. Textures as probes of visual processing // Annu. Rev. Vis. Sci. 2017. V. 3. № 1. P. 275–296. <https://doi.org/10.1146/annurev-vision-102016-061316>
26. Uejima T., Niebur E., Etienne-Cummings R. Proto-object based saliency model with second-order texture feature // 2018 IEEE Biomedical Circuits and Systems Conf. (BioCAS). Cleveland, OH: IEEE, 2018. P. 1–4. <https://doi.org/10.1109/BIOCAS.2018.8584749>
27. Williams C.C., Castelhana M.S. The changing landscape: High-level influences on eye movement guidance in scenes // J. Vision. 2019. V. 3. № 3. P. 33. <https://doi.org/10.3390/vision3030033>
16. Yavna D.V., Babenko V.V., Ikonopistseva K.A. Neural network models of second order visual filters//NeuralNetworksandNeurotechnologies. St. Petersburg, Russia: BBM, 2019. P. 198–203.
17. Yavna D.V., Babenko V.V., Stoletniy A.S., Shchetinina D.P., Alekseeva D.S. Differentiation and decoding of the spatial modulations of textures by the multilayer convolutional neural networks [in Russian] // Russian Foundation for Basic Research J. 2019. V. 4(104). P. 94–104. <https://doi.org/10.22204/2410-4639-2019-104-04-94-104>
18. Frey H.-P., König P., Einhäuser W. The role of first- and second-order stimulus features for human overt attention // Percept Psychophys. 2007. V. 69. № 2. P. 153–161. <https://doi.org/10.3758/bf03193738>
19. Johnson A., Zarei A. Second-order saliency predicts observer eye movements when viewing natural images // J. Vision. 2010. V. 10. № 7. P. 526–526. <https://doi.org/10.1167/10.7.526>
20. Gavrikov P. Visualkeras software [access mode]: <https://github.com/paulgavrikov/visualkeras>
21. Prins N., Kingdom F.A.A. Detection and discrimination of texture modulations defined by orientation, spatial frequency, and contrast // JOSA. A. 2003. V. 20. № 3. P. 401. <https://doi.org/10.1364/JOSAA.20.000401>
22. Sandler M., Howard A., Zhu M., et al. MobileNetV2: Inverted residuals and linear bottlenecks // arXiv:1801.04381 [cs]. 2019. Retrieved from <http://arxiv.org/abs/1801.04381>
23. Yu F., Seff A., Zhang Y., et al. LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop // arXiv:1506.03365 [cs]. 2016. Retrieved from <http://arxiv.org/abs/1506.03365>
24. Xiao J., Hays J., Ehinger K.A., et al. SUN database: Large-scale scene recognition from abbey to zoo // 2010 IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition. 2010. P. 3485–3492. <https://doi.org/10.1109/CVPR.2010.5539970>
25. Victor J.D., Conte M.M., Chubb C.F. Textures as probes of visual processing // Annu. Rev. Vis. Sci. 2017. V. 3. № 1. P. 275–296. <https://doi.org/10.1146/annurev-vision-102016-061316>
26. Uejima T., Niebur E., Etienne-Cummings R. Proto-object based saliency model with second-order texture feature // 2018 IEEE Biomedical Circuits and Systems Conf. (BioCAS). Cleveland, OH: IEEE, 2018. P. 1–4. <https://doi.org/10.1109/BIOCAS.2018.8584749>
27. Williams C.C., Castelhana M.S. The changing landscape: High-level influences on eye movement guidance in scenes // J. Vision. 2019. V. 3. № 3. P. 33. <https://doi.org/10.3390/vision3030033>

АВТОРЫ

Денис Викторович Явна — кандидат психологических наук, доцент, Южный федеральный университет, Ростов-на-Дону, 344006, Россия; Scopus ID 56034231500; <https://orcid.org/0000-0003-2895-5119>; yavna@fortran.su

Виталий Вадимович Бабенко — доктор биологических наук, профессор, Южный федеральный университет, Ростов-на-Дону, 344006, Россия; Scopus ID 7101752217; <https://orcid.org/0000-0002-3750-1277>; babenko@sfedu.ru

Ольга Алексеевна Горбенкова — студент, Южный федеральный университет, Ростов-на-Дону, 344006, Россия; <https://orcid.org/0000-0002-3302-563X>; gorbenkova@sfedu.ru

Игорь Владимирович Плавельский — магистрант, Южный федеральный университет, Ростов-на-Дону, 344006, Россия; <https://orcid.org/0000-0003-1594-1842>; nilpga@yandex.ru

Виктория Дмитриевна Вороняя — студент, Южный федеральный университет, Ростов-на-Дону, 344006, Россия; <https://orcid.org/0000-0002-9984-2245>; vikt.voronaya@yandex.ru

Александр Сергеевич Столетний — кандидат психологических наук, доцент, Южный федеральный университет, Ростов-на-Дону, 344006, Россия; Scopus ID 56801819200; <https://orcid.org/0000-0001-7955-8696>; stoletniynk@inbox.ru

Статья поступила в редакцию 30.03.2022

Одобрена после рецензирования 10.08.2022

Принята к печати 29.11.2022

AUTHORS

Denis V. Yavna — Candidat nauk (PhD in Psychology), Associate Professor, Southern Federal University, Rostov-on-Don, 344006, Russia; Scopus ID 56034231500; <https://orcid.org/0000-0003-2895-5119>; yavna@fortran.su

Vitaly V. Babenko — Doctor nauk (Doctor of Sciences in Biological Sciences), Professor, Southern Federal University, Rostov-on-Don, 344006, Russia; Scopus ID 7101752217; <https://orcid.org/0000-0002-3750-1277>; babenko@sfedu.ru

Olga A. Gorbenkova — Student, Southern Federal University, Rostov-on-Don, 344006, Russia; <https://orcid.org/0000-0002-3302-563X>; gorbenkova@sfedu.ru

Igor V. Plavelsky — Master Student, Southern Federal University, Rostov-on-Don, 344006, Russia; <https://orcid.org/0000-0003-1594-1842>; nilpga@yandex.ru

Victoria D. Voronaya — Student, Southern Federal University, Rostov-on-Don, 344006, Russia; <https://orcid.org/0000-0002-9984-2245>; vikt.voronaya@yandex.ru

Alexander S. Stoletniy — Candidat nauk (PhD in Psychology), Associate Professor, Southern Federal University, Rostov-on-Don, 344006, Russia; Scopus ID 56801819200; <https://orcid.org/0000-0001-7955-8696>; stoletniynk@inbox.ru

The article was submitted to the editorial office 30.03.2022

Approved after review 10.08.2022

Accepted for publication 29.11.2022