

УДК 612.82, 004.93

Джоконда и неопределенность распознавания улыбки человеком и искусственной нейронной сетью

© 2019 г. **О. В. Жукова***, канд. психол. наук; **Е. Ю. Малахова***;
Ю. Е. Шелепин*, **, доктор мед. наук

*Институт физиологии им. И.П. Павлова Российской академии наук, Санкт-Петербург

** Санкт-Петербургский государственный университет, Санкт-Петербург

E-mail: volgazhukova@gmail.com

Поступила в редакцию 03.07.2019

DOI:10.17586/1023-5086-2019-86-11-40-50

Проведен сравнительный анализ возможностей распознавания улыбки человеком и искусственной нейронной сетью в условиях неопределенности. Методом функциональной магнитно-резонансной томографии исследованы основные паттерны мозговой активности. Существуют фундаментальные ограничения, свойственные естественным и искусственным нейронным сетям, поэтому получены общности результатов распознавания тестовых изображений в надпороговых и пороговых условиях. Так, вероятность распознавания улыбки в обычных условиях достаточно высока, но в условиях неопределенности (пороговые и зашумленные изображения) и у человека, и у искусственной нейронной сети она снижается. Например, распознавание улыбки в мимике Джоконды человеком и искусственной нейросетью происходит с вероятностью 0,69. Предполагаем, что важнейшим принципом работы в обеих сетях является механизм согласованной фильтрации как меры соответствия предъявленного изображения с выученным нейронной сетью шаблоном, в частности, улыбки.

Ключевые слова: улыбка, искусственная нейронная сеть, распознавание, крупномасштабная нейронная сеть, паттерн мозговой активности.

Коды OCIS: 100.4996, 170.6960, 330.5020

ВВЕДЕНИЕ

Исследование особенностей распознавания мимики зрительной системой человека и их моделирование в искусственных нейронных сетях остается важнейшей фундаментальной проблемой. Существующие алгоритмы достаточно успешно справляются с задачами в надпороговых условиях принятия решения. В условиях же неопределенности эффективность подобных решений падает.

Примером неопределенности восприятия мимики является «улыбка» Джоконды. Леонардо да Винчи довел изменение изгиба губ, имитируя улыбку, до пороговой величины. При пороговых значениях сигнала (улыбки) под влиянием внутренних шумов зрительной системы мимика вызывает у наблюдателя неопределенность (есть улыбка или ее нет). Ситуация усложняется еще и тем, что картина написана в художественной технике сфумато, при которой слои краски наносятся

таким образом, что не остается видимых мазков при переходах тонов или четких линий, разделяющих светлые и темные области. Таким образом, возникает живопись без линий и границ.

Наиболее удачным решением в пороговых условиях наблюдения являются сверточные нейронные сети глубокого обучения. В отличие от алгоритмов компьютерного зрения предыдущего поколения, нейронные сети эффективно работают с естественными изображениями, подобными тем, что встречаются человеку ежедневно и не требуют ручного программирования признаков: то, какие характеристики изображения выделять, и как они соотносятся с конечной задачей, определяется автоматически во время обучения модели. Нейронные сети также обладают возможностью работать с образами независимо от того, при каком освещении они были запечатлены, с какого ракурса и какую часть пространства на снимке занимают. Данные техни-

ческие решения прошли долгий путь развития, но одними из главных механизмов, реализованных в них, остаются механизм обучения перцептрона [1] и иерархическая обработка зрительного сигнала [2–3]. Как итог, внутренняя структура и качество конечного результата работы нейронных сетей позволяют рассматривать их как возможную модель зрительной коры головного мозга, изучать происходящие процессы и проводить двусторонние аналогии между указанными системами.

Таким образом, цель данного исследования — проведение сравнительного анализа возможностей биологических и искусственных нейронных сетей при распознавании мимики в неопределенных условиях принятия решения.

МЕТОДИКА

В нейрофизиологической серии исследования принимали участие 24 испытуемых (8 мужчин и 16 женщин), возрастной диапазон и мужчин, и женщин — от 19 до 33 лет (средний возраст мужчин $24,5 \pm 2,8$ года, средний возраст женщин $24,8 \pm 3,2$, средний возраст по группе $24,7 \pm 3$). Исключающим критерием являлись наличие патологических анатомических особенностей головного мозга, черепно-мозговые травмы и заболевания нервной системы. Все испытуемые дали согласие на участие в эксперименте.

Стимулы. В качестве стимулов использованы изображения лиц виртуальных людей. Все манипуляции с изображением выполнены на базе программного обеспечения — FaceGen (Singular Inversions, Canada). Валидность реализованных программных алгоритмов морфинга была неоднократно доказана в многочисленных психологических исследованиях. Все стимулы были разбиты на две группы. В первую группу включены изображения лиц с выраженной мимикой радости и грусти и поворотами головы направо-налево на шесть градусов по встроенной шкале FaceGen (надпороговые условия принятия решения), а во вторую группу — изображения тех же лиц, но с нейтральной мимической экспрессией, анфас (неопределенные условия принятия решения).

Оборудование и процедура. Для проведения комплексного исследования функциональных и когнитивных возможностей зрительной системы применялась функциональная магнитно-резонансная томография (фМРТ) — сканирование всего головного мозга на 1.5T MR-сканере *Symphony*, *Siemens*, с градиентами 40 мТл/м. Для выявления активированных участков использовался метод отображения уровня кислорода в крови (*Blood-oxygen-level-dependent imaging* — BOLD), базирующийся на различиях магнитных свойств оксигемоглобина и дезоксигемоглобина. Получаемые карты активации обрабатывались в программе

SPM8 — *Statistical parametric mapping* (Wellcome Trust Centre of Neuroimaging, London, UK), работающей в среде Matlab 8.0 (2012b). Данные, полученные со сканера в формате DICOM, конвертировались в формат NiFTI. Чтобы скорректировать движения головы между отдельными сканами на первой стадии обработки, все полученные изображения сводились к первому скану, после чего получали усредненное EPI-изображение — эхопланарное изображение (*echo planar imaging*), которое было нормировано к стандартному EPI-изображению, предоставленному Монреальским нейробиологическим институтом (MNI-пространство). Далее выполнялись сегментация и пространственное сглаживание, в процессе которого нормированные изображения с размером вокселя $2 \times 2 \times 2$ мм размывались гауссовой функцией диаметром 8 мм. Эта процедура применялась к данным каждого испытуемого.

Процедура стимуляции включала этап обучения и этап проведения исследования. Во время обучения испытуемым разъяснялась инструкция и давался тренировочный набор тестовых изображений в надпороговых условиях предъявления. Во время исследования испытуемым предъявляли изображения четырех фаз стимуляции. В фазах 1 и 3 многократно демонстрировались изображения 36 разных людей с нейтральной мимической экспрессией, анфас. В фазах 2 и 4 — изображения этих же лиц с разной мимической экспрессией и поворотами головы. Межстимульный интервал составлял 1000 мс (рис. 1). Каждая фаза состояла из 10 сканирований по 3 с. Каждое исследование начиналось с предъявления слайда с точкой фиксации. Через 30 с происходила смена стимуляции — предъявление в случайном порядке изображений одной из четырех фаз. Первый слайд каждой фазы — инструкция. Далее — стимулы. Изображения каждой фазы предъявляли по 10 раз. Длительность всего сканирования составила 25 мин.

Инструкция. Испытуемые получали инструкцию определять мимику или поворот головы виртуального человека. Выбор регистрировался нажатием испытуемым левой или правой клавиши мыши. Испытуемые не были проинструктированы о наличии изображений с нейтральным выражением лица, анфас. Данный материал воспринимался как стимулы с «не заметной для глаза» минимальной степенью выраженности признака. Таким образом, в фазах 1 и 3 были смоделированы условия крайней степени неопределенности.

Статистическая обработка данных. В рамках статистического анализа фМРТ-данных использовалось несколько стандартных методик, которые являются общепринятыми в мире и наиболее распространенными, что дает возможность статистически корректно выявлять наборы областей мозга, вовлекаемых в обеспечение исследуемой деятельности. Использовался стандартный статистический

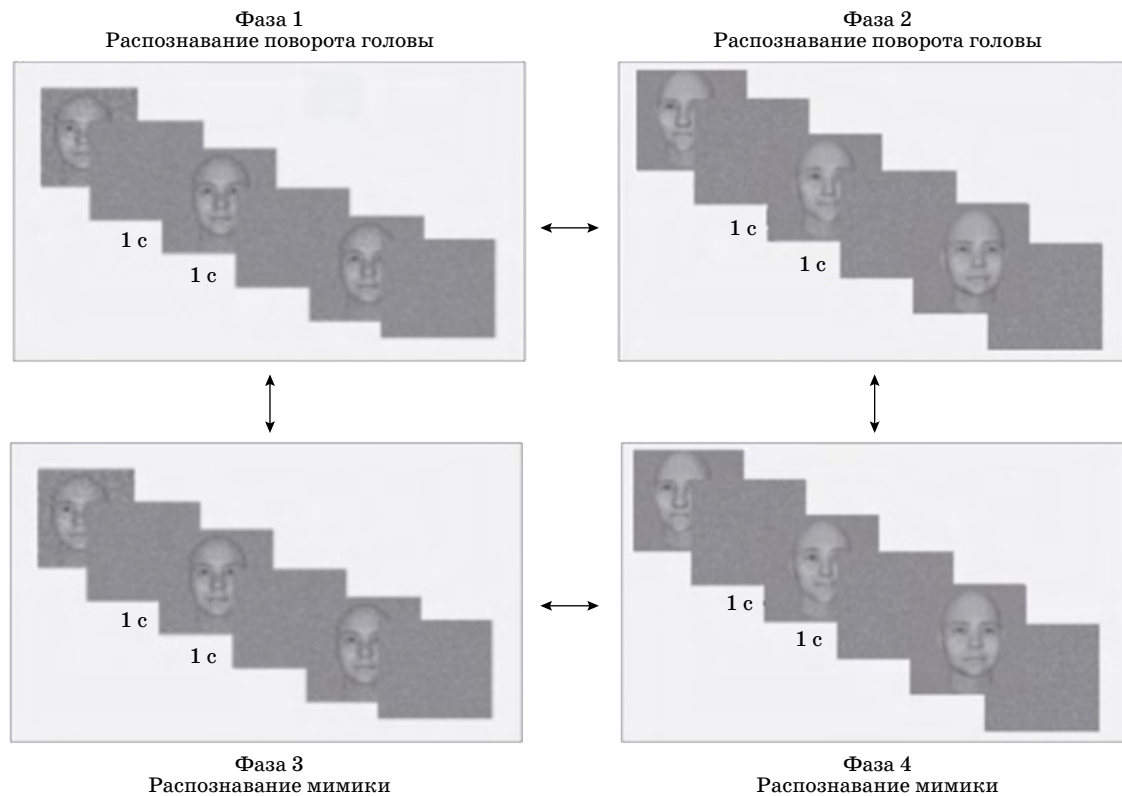


Рис. 1. Параметры стимуляции и возможные типы вычитания 4 фаз стимуляции при проведении фМРТ. Фаза 1 — предъявление изображений с нейтральным выражением лица (анфас) в задачах определения поворота головы, фаза 2 — предъявление изображений лиц, выражающих разные эмоции, с различными поворотами головы в задачах определения наличия поворота, фаза 3 — предъявление изображений с нейтральным выражением лица (анфас) в задачах определения эмоции, фаза 4 — предъявление изображений лиц, выражающих разные эмоции, с различными поворотами головы в задачах определения эмоции.

метод обработки фМРТ-данных — метод множественной регрессии в рамках обобщенной линейной модели (*generalized linear models*). Данный метод позволяет выявлять параметры регрессии каждого вокселя изображения, которые являются статистической мерой соответствия между появлением пробы и соответствующими изменениями зарегистрированного BOLD-сигнала. В качестве регрессоров выступали параметры регистрируемого сигнала (тип пробы и его длительность), а также параметры смещения положения головы и некорректные пробы. Далее осуществлялась коррекция на множественность сравнений методом выявления ожидаемой доли ложных отклонений (*false discovery rate* — FDR). Цель данного этапа — статистическое параметрическое картирование. На данном этапе проводилась проверка гипотезы о случайном распределении данных. Оценка локальной активности относительно всего мозга проводилась методом двухкомпонентного *t*-теста (*t*-критерий Стьюдента, уровень значимости отличий $p < 0,001$). По результатам анализа были определены области мозга с максимальным уровнем активации.

В нейротехнологической серии исследования для реализации процесса распознавания эмоций

с помощью искусственных нейронных сетей была выбрана сверточная архитектура, имеющая функциональные сходства с вентральным зрительным путем [4–5]. В качестве базовой модели использовалась глубокая нейронная сеть VGG Face [6], обученная задаче распознавания общеизвестных людей на датасете из 2,6 млн. изображений, собранных в интернете. Данная сеть была выбрана из-за предполагаемой релевантности признаков, выделяемых при идентификации лиц для задачи распознавания мимических экспрессий.

Стимулы. Было проведено переобучение нейронной сети на специализированном датасете меньшего размера, включающем в себя изображения виртуальных и реальных лиц с разной степенью выраженности мимики и разными поворотами головы. Обучающий датасет состоял из двух классов по 500 стимулов в каждом. В первую группу вошли изображения лиц с мимикой радости, во вторую группу — с мимикой грусти. В процессе обучения сети данные были разбиты на две группы: тренировочное множество, которое использовалось для переобучения сети, и тестовое — необходимое для контроля качества обучения. В качестве контрольной выборки методом экспертной

оценки были отобраны изображения лиц, представленных в портретной живописи, например, изображение Джоконды (Леонардо да Винчи).

Задача данной серии экспериментов заключалась в том, чтобы узнать, как поведет себя искусственная нейронная сеть, обученная распознавать улыбку и грусть в пороговых условиях (условиях неопределенности).

РЕЗУЛЬТАТЫ

Обработка изображений нейрофизиологической серии исследований

Для исследования особенностей распознавания стимулов в условиях неопределенности применялись четыре типа вычитания (рис. 1). Под «вычитанием» понимается сравнение BOLD-активности в ответ на предъявление изображений лиц с явно выраженной мимикой (фазы 2 и 4) с BOLD-активностью на предъявление лиц с неопределенной мимикой (фазы 1 и 3). Разница BOLD-отклика, полученная путем вычитания, представлена на статистических картах изображений головного мозга (рис. 2, 3).

Первый тип вычитания — сравнение BOLD-активности в двух фазах стимуляции, во время которых инструкция оставалась одной и той же, но менялась степень выраженности признаков (наличие очевидной или неопределенной мимики и поворотов головы) (рис. 2).

В задаче распознавания поворота головы (из фазы 1 вычитали фазу 2) получены два паттерна активации, обозначенные на рисунке разными цветами. Красным цветом отмечена активация в условии вычитания «фаза 1 минус фаза 2», а голубым — обратное вычитание: «фаза 2 минус фаза 1». В первом случае видна активация в нижневисочной (фузиформная извилина, BA37), а также во фронтальной (BA10, BA9, BA8, BA32), теменной (BA40, BA7) и премоторной (BA6) коре головного мозга человека. Во втором случае наблюдается отличный паттерн крупномасштабной нейронной сети, который включает медиальную префронтальную кору (BA9, BA10, BA11, BA24), височно-теменную стык (BA39, BA40, BA22), предклинье и заднюю поясную извилину (BA23, BA29, BA30, BA31).

В задаче распознавания мимики (из фазы 3 вычитали фазу 4) получена активация в теменной коре (BA7, BA40, BA31). При обратном вычитании активации не видно, что, вероятно, связано с минимальными различиями в тестовых изображениях в фазах 3 и 4.

Второй тип вычитания — это сравнение BOLD-активности в двух фазах стимуляции, во время которых степень выраженности признаков была одной и той же, но менялась инструкция (опре-

деление либо поворота головы, либо мимики). Полученные в процессе вычитания карты активации представлены на рис. 3.

В условиях предъявления изображений лиц с явной мимической экспрессией и поворотами головы (фаза 2 минус фаза 4) получены два паттерна активации, обозначенные на рис. 3 разными цветами. Желтым цветом обозначена активация при выполнении задачи определения поворота, а голубым — эмоции. Напомним, что в стимулах, предъявляемых в обеих фазах стимуляции, присутствовали оба явно выраженных признака — и поворот, и мимическая экспрессия. Можно предположить, что при наличии обоих признаков происходит автоматическое их сравнение и ранжирование, например, по степени релевантности, значимости, сложности. Видимо, с этим связана активация паттерна «сети базового режима» в задаче определения поворота головы. В задаче же определения эмоции наблюдается активация в областях мозга, отвечающих за эмоциональное восприятие стимулов (например, в островковой доле конечного мозга — инсуле).

В условиях предъявления изображений лиц с нейтральной мимикой, анфас, (фаза 1 минус фаза 3) оба изучаемых признака отсутствовали (рис. 3). В итоге получены два паттерна крупномасштабной нейронной сети. Оранжевым цветом отмечена активация в задаче распознавания поворота, что ожидаемо вызвало вовлечение в работу дорсальные зрительные пути, ответственные за описание пространственных отношений. А голубым цветом отмечена активация в задаче распознавания мимической экспрессии, что повлекло вовлечение областей, ответственных за эмоциональное восприятие стимулов, характерных для наших условий стимуляций (например, островковой доли конечного мозга).

Анализ изображений в психофизических исследованиях

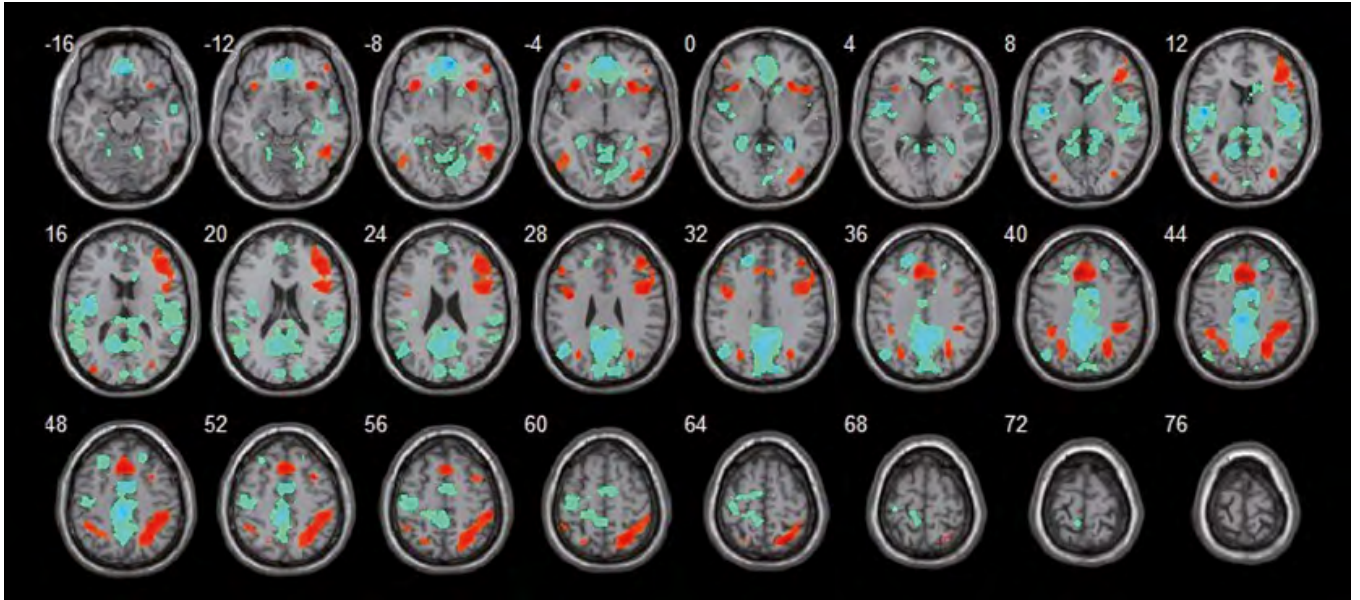
Данные по ответам и времени реакции испытуемых представлены на рис. 4. Как видно из представленных данных, распознавание мимических экспрессий и поворотов головы в фазах 1 и 3 (в условиях неопределенности) представляло значительную сложность для испытуемых, что привело к более низкой скорости реакции. Вероятность распознавания мимики снизилась до уровня случайного гадания — 50%.

Таким образом, в нейрофизиологической серии исследования получены различные паттерны крупномасштабной нейронной сети головного мозга человека в зависимости от экспериментальной парадигмы (чередование очевидных и неопределенных для принятия решения условий). При наличии в тестовых изображениях и поворотов головы, и мимической экспрессии, вероятно, происхо-

Фаза 1 – Фаза 2



Инструкция «поворот»



Фаза 3 – Фаза 4



Инструкция «эмоция»

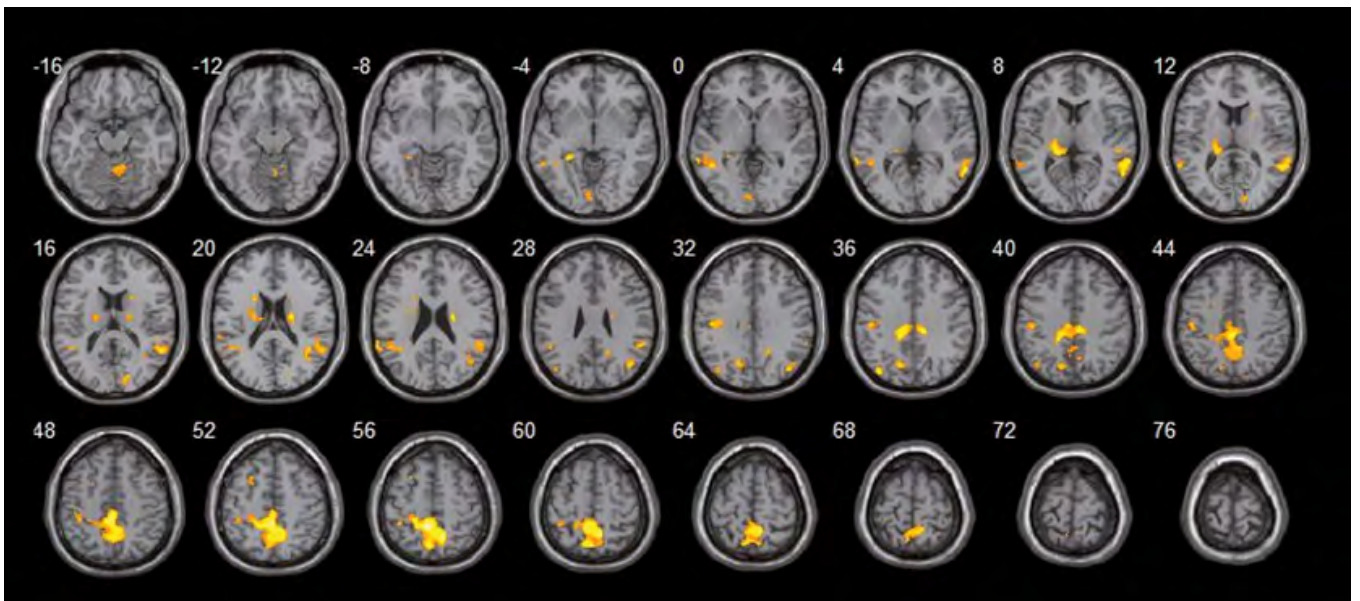


Рис. 2. Карты активации, полученные при вычитании фаз стимуляции в задаче распознавания поворота головы (фаза 1 минус фаза 2) и мимической экспрессии лица (фаза 3 минус фаза 4). Усредненные данные группы испытуемых из 24 человек (без FDR-corr., $p < 0,001$).

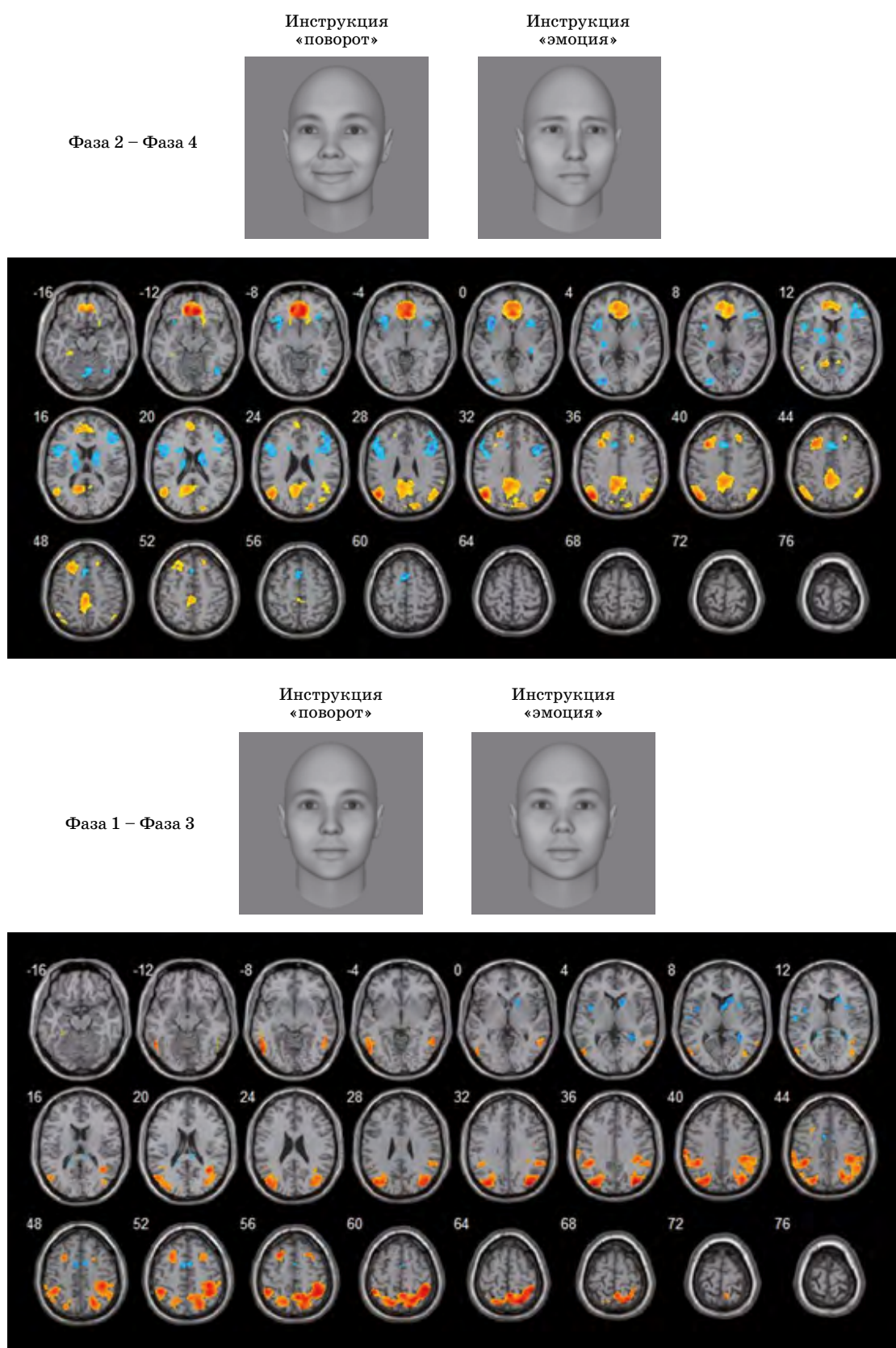


Рис. 3. Карты активации, полученные при вычитании фаз стимуляции. В обеих фазах предъявлялись изображения лиц с мимической экспрессией и с поворотами головы в разные стороны (фаза 2 минус фаза 4), в обеих фазах предъявлялись изображения лиц с нейтральным выражением, анфас (фаза 1 минус фаза 3). Менялась только инструкция. Усредненные данные группы испытуемых из 24 человек (без FDR-corr., $p < 0,001$).

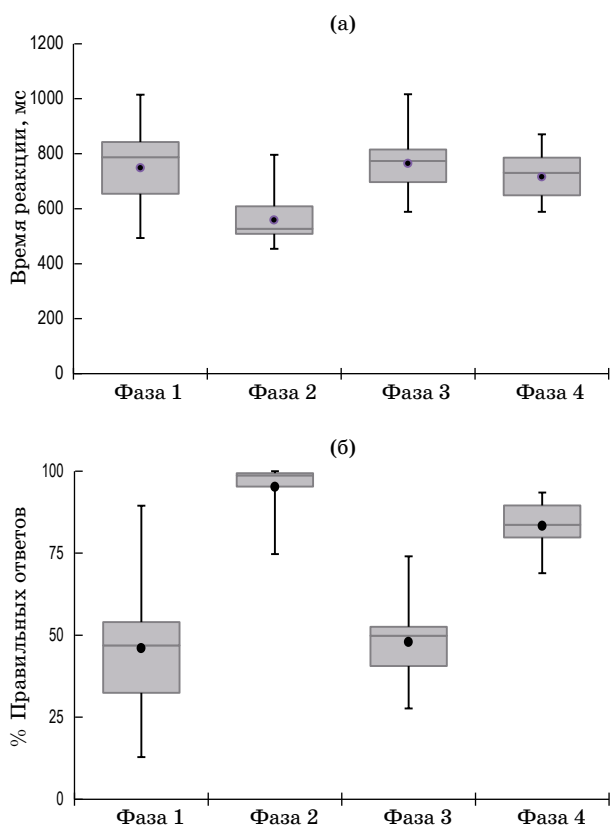


Рис. 4. Диаграммы размаха по времени реакции (а) и ответам 22 испытуемых при распознавании поворота головы (направо) и мимики (радость) на изображениях 4 фаз стимуляции (б).

дит автоматическое сравнение признаков и оценка их физических и семантических характеристик. Причем данное сравнение определяется не только инструкцией, но и другими параметрами. Так, в задаче определения мимики видна активация также дорсальных зрительных путей, ответственных за описание пространственных отношений, а в задаче распознавания поворота — активация в медиальной префронтальной коре, ответственной в том числе за осознанное и неосознанное сопереживание текущему эмоциональному состоянию другого человека.

Анализ изображений в модельных исследованиях на сверточной нейронной сети глубокого обучения

Архитектура обучаемой нейронной сети состояла из 13 сверточных и 3 полносвязных слоев. Сверточные слои обработали изображение фильтрами небольшого размера [7]. Каждый фильтр кодировал какой-либо графический признак в виде матрицы, например, если говорить о первом слое, это могло быть наличие наклонной линии под

определенным углом, цветового градиента и др. Суть операции свертки заключалась в умножении изображения на фильтр (ядро) свертки. Так как размер фильтра существенно меньше размера изображения, он применялся последовательно с небольшим шагом ко всему изображению, а результат суммировался и записывался в аналогичную позицию выходного изображения. Таким образом, была сформирована карта активации, которая отражает соответствие данного изображения фильтру. Фильтры первого сверточного слоя выделили простые признаки вроде цвета и наклона линии — они универсальны практически для всех сетей, работающих с изображениями, вне зависимости от выполняемой задачи (рис. 5а). На рис. 5б приведены примеры сопоставления фильтра с входным изображением — карты активации, где светлые участки означают максимальную схожесть между фильтром и фрагментом картины.

Фильтры второго слоя не являлись настолько же легко интерпретируемыми, так как размер рецептивного поля увеличился и изображения, подобранные для визуализации максимальной активации, содержали менее тривиальные формы. Нейроны последующих сверточных слоев охватывали все большую часть изображения и реагировали на все более сложные формы объектов, вплоть до выделения различных категорий объектов. Тем не менее, в указанных слоях была сохранена зависимость реакции нейрона от пространственного расположения признака. И только в полносвязных слоях сети было достигнуто инвариантное — независимое от положения и размера — распознавание образов. Каждый нейрон полносвязного слоя получил информацию от всех нейронов предыдущего слоя, что позволило учитывать все возможные пространственные локации объектов. Некоторые нейроны в обучаемой выборке могли определить признаки/объекты на изображениях, некоторые — нет. Последний из полносвязных слоев содержал количество нейронов, равное количеству классов в исходной задаче классификации, т.е. в случае задачи распознавания мимики лица на изображении этот слой состоял из двух нейронов: один, представляющий категорию «улыбка», другой — представляющий категорию «отсутствие улыбки».

Тестирование распознавания мимики с помощью данной нейронной сети показало результат категоризации изображений лиц с выраженной улыбкой, равный 97%. В условиях неопределенности, скрытой или едва намечаемой улыбки вероятность распознавания снизилась до уровня случайного гадания — 50% (есть улыбка или ее нет). Например, при распознавании неопределенной мимики Джоконды показано, что ее лицо с вероятностью 0,69 выражает улыбку.

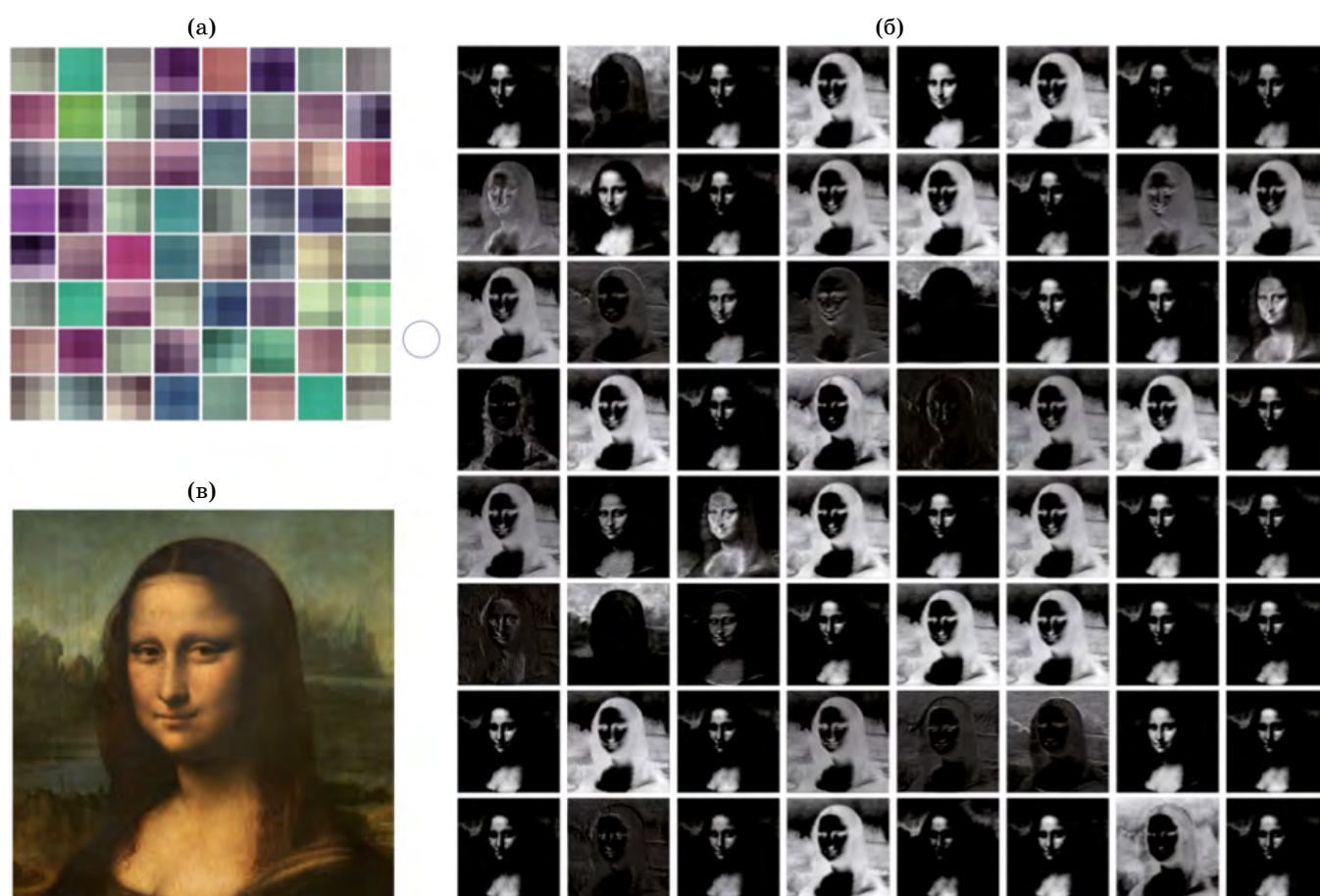


Рис. 5. Визуализация работы первого сверточного слоя нейронной сети. Визуализация фильтров первого сверточного слоя, размер фильтров 3×3 пиксела, 3 цветовых канала (а). Карты активации первого сверточного слоя, соответствующие приведенным фильтрам, где светлые участки означают максимальную схожесть между фильтром и фрагментом картины, темные — отсутствие активации (б). Оригинальное изображение (в).

ОБСУЖДЕНИЕ

Целью данного исследования было проведение сравнительного анализа возможностей биологических и искусственных нейронных сетей при распознавании мимики в условиях неопределенности. В соответствии с поставленной целью было проведено обучение искусственной нейросети по распознаванию мимических экспрессий и поворотов головы на датасете тестовых изображений. Аналогичный этап обучения проходили и испытуемые. После соответствующего обучения искусственной нейросети и испытуемым были предъявлены тестовые изображения лиц с очевидной и неопределенной мимикой и поворотами головы. В очевидных условиях принятия решения точность классификации искусственной нейронной сетью составила 97%, а испытуемыми — на уровне 95% при распознавании поворота головы и 83% — при распознавании мимики. В условиях неопределенной мимики эффективность распознавания и в случае искусственной нейросети, и в случае испытуемых опустилась до уровня случайного гадания. Можно предположить, что при дальнейшем обуче-

нии нейросетей восприятию мимики в пороговых условиях наблюдения точность распознавания повысилась бы значительно. Но в условиях предъявления данных, которые изначально не входили в тренировочный набор, были получены схожие результаты при работе как искусственных, так и биологических нейронных сетей.

Вероятно, объяснить подобный результат можно с позиции разрабатываемых функциональных моделей зрительной системы, основанных на предположении о существовании в зрительной системе механизма согласованной фильтрации как меры соответствия предъявляемого тестового изображения и эталона, хранящегося в памяти. Эта модель была предложена и успешно развита в СССР Н.Н. Красильниковым, а в США — Таннером и Светсом еще в 50-е гг. XX в. [8]. В соответствии с моделью сначала происходит первичная фильтрация в многочисленных параллельных фильтрах, затем идет согласованная фильтрация и, наконец, на высоких уровнях зрительной системы происходит принятие решения. Необходимо отметить, что помимо указанных звеньев модели

на основе согласованной фильтрации учитывают влияние внешнего и внутреннего шумов зрительной системы. В процессе принятия решения из-за влияния шума возможны ошибки, но они, как правило, имеют экологически обусловленный смысл [9]. Согласованный фильтр лица — «штрихкод лица», вероятно, задан частично генетически, так как новорожденные предпочитают смотреть на изображение лица в сравнении с другими типами изображений [9]. Необходимо отметить, что согласованные фильтры обучаемы. Операцию сравнения наблюдаемого и хранящегося в памяти лица выполняют обширные нейронные структуры. Выделяются инвариантное лицо, лицо вообще и лицо конкретного человека в определенные периоды его жизни. При этом лучше запоминаются фотографии лица, так как мимика в них фиксирована. Наличие согласованного фильтра, выделяющего лицо, позволяет в хаотических случайных структурах, например, в белом шуме, видеть лицо. В этом плане представляет интерес исследование, в рамках которого искусственной нейронной сети VGG Face [6] были предъявлены изображения шумов, отфильтрованных в областях высоких и низких пространственных частот. Результаты показали, что искусственная нейросеть распознает в шумах наличие улыбки, причем в отношении шума с фильтрацией в области низких частот уверенности в этом больше, чем при фильтрации в области высоких частот [9–10].

Представляет интерес локализация звеньев модели согласованной фильтрации, т.е. в привязке их к конкретным нейронным структурам с последующим уточнением алгоритма их работы.

При исследовании биологических нейронных сетей в условиях восприятия лиц ожидалось увидеть характерный паттерн активации, свойственный современным представлениям об иерархическом строении зрительной системы с прямым распространением сигнала от сетчатки к дорсальному ядру наружного коленчатого тела, далее к областям зрительной коры V1, V2, V4 и далее к нижневисочной коре (НК) головного мозга. Это достаточно упрощенное представление, так как существуют важные связи сетчатки еще и с другими подкорковыми ядрами и отделами коры. Есть и множественные связи сетчатки, минуя наружное коленчатое тело, в частности, идущие через подушку зрительного бугра (через пульвинар) и поступающие непосредственно в высшие зрительные центры, минуя первичную зрительную кору [11]. Нейроны на нижних уровнях иерархии имеют маленькие рецептивные поля (обрабатывают сигнал с малой части изображения), в то время как нейроны в более высоких областях видят все большую часть зрительного поля, аккумулируя сигналы с нижестоящих уровней. Таким образом, нейроны первичной зрительной коры выделяют

простые признаки в локальных областях (такие как ориентация линий, частота, цвет), а НК отвечает за комплексные образы (такие как человек, животные и др.). Однако результаты исследования показали отсутствие активации в первичной зрительной коре, а также в других областях мозга. Например, в области дна верхней височной борозды, где активность нейронов определяется не только наличием в поле зрения лиц, но весьма чувствительна к мимике. А также в области нижней затылочной извилины, в которой найдены статистически значимые отличия при восприятии изображений лиц в сравнении с другими объектами. Согласно методике проведения фМРТ карты активации рассчитываются на основе вычитания одной фазы стимуляции из другой. Соответственно на картах активации области, активируемые одинаково в обоих режимах, не видны и поэтому не окрашены. В обеих фазах предъявлялись изображения одних и тех же лиц, отличия в изображениях состояли только в степени поворота головы и изгиба уголков губ и бровей. Вероятно, именно поэтому нет активации в указанных выше областях мозга. Это является фундаментальным ограничением метода фМРТ.

В различных условиях стимуляции видна активация различных областей мозга, которые демонстрируют синхронное изменение кровотока. На этом основании получило развитие представление о крупномасштабных нейронных сетях головного мозга человека. Необходимо отметить, что данные сети можно представить и на микроуровне как совокупность нейронов, которые функционально объединены в нервной системе, и на макроуровне как совокупность функционально взаимосвязанных отдельных зон, областей и даже полушарий мозга. Так, результирующие карты активации любого фМРТ-исследования являются примером крупномасштабной нейронной сети мозга человека, состоящей из областей мозга, активность которых превысила заданное пороговое значение. В то же время важно помнить, что зоны мозга, не достигшие порога активации, также являются частью нейронной сети мозга и вступают в сложное взаимодействие с другими областями.

В условиях неопределенности (при предъявлении изображений лиц с нейтральной мимикой, анфас) наблюдается активация в НК, в области фузиформной извилины, а также во фронтальной, теменной и премоторной коре головного мозга человека. Активация в области фузиформной извилины была ожидаема. Достаточно давно в области НК были зарегистрированы специфические реакции на изображения лиц, отличавшиеся по своей топографии от реакций на изображения домов, мебели или букв [12]. На основании полученных данных была построена схема строения НК, состоящей из сверхколонок нейронов, которые реагируют

на разные вариации в изображении одного объекта. Роль фронтальной коры в задачах принятия решения также известна [13]. Активация же теменной коры, видимо, обусловлена спецификой экспериментальных условий. Напомним, что стимулы не были связаны с испытуемыми какими-либо личными отношениями. Соответственно можно предположить, что испытуемые выполняли поставленную задачу как пространственную через определение относительного положения уголка губ — вверх или вниз. Решение подобной задачи невозможно без участия механизмов, обеспечивающих избирательное внимание, поэтому на полученных картах фМРТ видна активация различных участков теменной коры.

При принятии решения в очевидных условиях (предъявление изображений лиц с мимической экспрессией и поворотами головы) наблюдается активация характерного паттерна, который в современных исследованиях получил название «сеть по умолчанию», или «сеть базового режима». Данный паттерн включает медиальную префронтальную кору, височно-теменной стык, предклинье и заднюю поясную извилину. Основная особенность данной сети заключается в том, что она наиболее активна при отсутствии внешних раздражителей в состоянии относительного покоя. В данных экспериментальных условиях покоя не было, но стимулы, предъявляемые в фазах 2 и 4 (лица с выраженной мимической экспрессией и поворотом головы), вероятно, казались испытуемым более очевидной задачей в сравнении с фазами 1 и 3, когда предъявлялись изображения лиц с нейтральной мимикой, анфас. Напомним, что испытуемые не были уведомлены об отсутствии изучаемых признаков (поворота головы и мимической экспрессии) в изображениях, что делало задачу крайне неопределенной в решении. Таким образом, можно предположить, что крупномасштабная нейронная сеть «базового режима мозга» активируется не только при отсутствии внешних раздражителей, но и в условиях монотонной простой работы. Вероятно, это своеобразный «автопилот», цель поведения которого — экономия времени и сил. Полученные данные позволили высказать предположение о справедливости применения в психофизике фундаментального физического принципа — «принципа наименьшего действия».

В условиях принятия решения при смене инструкции (предъявление в двух фазах стимуляции изображений нейтральных лиц, анфас) предполагалось, что при одном и том же наборе стимулов, но в условиях разных инструкций, будут выявлены отличающиеся карты активации нейронной сети. Действительно, результаты показали, что в задаче распознавания поворота головы активируются дорсальные зрительные пути, а в задаче распознавания мимики — области мозга, ответственные

за эмоциональное восприятие стимулов, характерных для данных условий стимуляций (активация островковой доли конечного мозга).

Таким образом, в зрительной системе человека зависимость между комбинацией и уровнем возбуждения на сетчатке и принятием решения не так однозначна. Перестройка нейронной сети определяется самой задачей, стимулами-изображениями, их физическими, геометрическими и семантическими характеристиками, структурами глаза и «зрительного мозга», обеспечивающими восприятие, узнавание, понимание и ответную реакцию.

При исследовании сверточных нейронных сетей глубокого обучения было показано, что они наиболее приближены к биологическим нейронным сетям по сравнению с другими парадигмами классификации изображений (например, линейные модели, деревья решений и др.). Этот новый класс искусственных нейронных сетей [14] является продолжением работ середины XX в. [1, 3, 15]. В отличие от других решений архитектура данной сети имеет несколько «скрытых» слоев (в связи с чем получила название — сеть глубокого обучения) и обучается методом обратного распространения ошибки. Со времен работ Фукусимы [16] в основу нейронных сетей глубокого обучения были положены особенности, которые заимствованы в живых зрительных системах. Во-первых, это наличие специфических детекторов деталей изображения, реализованных на основе специальной формы рецептивных полей нейронов. Во-вторых, это иерархическая организация процесса классификации изображения, в которой, также как и в зрительной системе человека, на высоких уровнях иерархии осуществляется обобщение информации, полученной с более низких уровней. В-третьих, адаптация детекторов к различным свойствам натуральных сцен. Они формируются автоматически в процессе обучения сети. Полученные фильтры можно условно подразделить на две категории: те, что являются универсальными и активируются практически во всех случаях, и более специфичные — выделяющие определенный класс объектов. Начальные этапы обработки содержат универсальные фильтры, отражающие статистику сигнала, специфичные же формируются в глубоких слоях из определенных комбинаций универсальных. Таким образом, фильтры слоев формируются в процессе обучения и не меняются после. Карты активации же зависят от изображения, поданного в сеть на обработку, отражая присутствие тех или иных признаков и их соответствие ранее виденным паттернам. В итоге на выходе получается сеть, которая способна классифицировать данные, изначально не входившие в тренировочный набор. В нашем случае в качестве подобного рода изображений было выбрано изображение Джоконды Леонардо да Винчи. Несмотря

на то, что обученная нейронная сеть повторила результат, полученный людьми, к сожалению, невозможно утверждать однозначного сходства в процессах обработки сигнала моделью и зрительной системой. Сложность архитектуры и нелинейность операций в современных искусственных нейронных сетях приводит к невозможности простой трактовки их работы и прямого соотнесения информации, поданной на вход, и полученного на выходе сети результата. Однако дальнейшее изучение возможностей работы искусственных нейронных сетей, возможно, прольет свет на механизмы работы биологических зрительных систем.

ЗАКЛЮЧЕНИЕ

Таким образом, методы визуализации и анализа карт активации крупномасштабных нейронных сетей позволили изучать процессы, происходящие внутри искусственных нейронных сетей и сравнивать их с нейрофизиологическими данными. Перестройка нейронной сети определяется самой задачей, стимулами-изображениями, их физическими, геометрическими и семантическими характеристиками, структурами глаза и «зрительного мозга», обеспечивающими восприятие, узна-

вание, понимание и ответную реакцию. Благодаря тому, что в основу искусственных нейросетей заложены некоторые особенности, заимствованные в живых зрительных системах, в некоторых задачах были получены схожие результаты по точности распознавания тестовых изображений. Показано, что важнейшим принципом работы в обеих сетях является механизм согласованной фильтрации как меры соответствия предъявленного изображения эталону. Восприятие таинственной «улыбки» Джоконды как для человека, так и для искусственных нейронных сетей является работой в условиях неопределенности на пороге распознавания.

Нейрофизиологическая часть исследования выполнена при финансовой поддержке научно-исследовательского проекта «Психофизиологические и нейролингвистические аспекты процесса распознавания вербальных и невербальных паттернов», проект Российского научного фонда №14-18-02135.

Моделирование на сверточной нейронной сети глубокого обучения выполнено при финансовой поддержке Программы фундаментальных научных исследований государственных академий на 2013–2020 гг. (ГП-14, раздел 63), Институт физиологии им И.П. Павлова.

ЛИТЕРАТУРА

1. Розенблатт Ф. Принципы нейродинамики (перцептрон и теория механизмов мозга). М.: Мир, 1965. 480 с.
2. Hubel D.H., Wiesel T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex // J. Physiology. 1962. V. 160. P. 106–154.
3. Fukushima K., Miyake S. Neocognitron. A self-organizing neural network model for a mechanism of visual pattern recognition / in Competition and cooperation in neural nets. Berlin: Springer Heidelberg, 1982. P. 267–285.
4. Guclu U., van Gerven M.A. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream // J. Neuroscience: The Official Journal of the Society for Neuroscience. 2015. V. 35. P. 10005–10014.
5. Kriegeskorte N. Deep neural networks: A new framework for modeling biological vision and brain information processing // Annual Review of Vision Sci. 2015. V. 1. P. 417–446.
6. Parkhi O.M., Vedaldi A., Zisserman A. Deep face recognition // In BMVC. 2015. V. 1. № 3. P. 6.
7. Малахова Е.Г. Обработка зрительной информации в искусственных и биологических нейронных сетях // Нейротехнологии / Под ред. Шелепина Ю.Е., Чихмана В.Н. СПб.: изд. ВВМ, 2018. С. 338–349.
8. Красильников Н.Н., Шелепин Ю.Е. Маскирование как результат согласованной фильтрации // Физиология человека. 1996. Т. 22. № 5. С. 99–103.
9. Шелепин Ю.Е., Борачук (Жукова) О.В., Пронин С.В. и др. Лицо и невербальные средства коммуникации // Петербургский психологический журнал. 2014. № 9. С. 1–43.
10. Жукова О.В. Закономерности перестройки крупномасштабной нейронной сети головного мозга человека при распознавании лиц в условиях неопределенности // Автореф. канд. дис. СПб.: СПбГУ, 2017. 25 с.
11. Подвигин Н.Ф., Макаров Ф.Н., Шелепин Ю.Е. Элементы структурно-функциональной организации зрительно-глазодвигательной системы. Л.: Наука, 1986. 252 с.
12. Tanaka K., Saito H., Fukada Y., et al. Coding visual images of objects in the inferntemporal cortex of the macaque monkey // J. Neurophysid. 1991. V. 66. № 1. P. 170–189.
13. Шелепин Ю.Е., Фокин В.А., Хараузов А.К. и др. Локализация методами нейроиконики механизмов принятия решений об упорядоченности текстур // Оптический журнал. 2014. Т. 81. № 1. С. 57–69.
14. Cadieu C.F., Hong H., Yamins D.L., et al. Deep neural networks rival the representation of primate IT cortex for core visual object recognition // PLoS Computational Biology 10:e1003963. 2014.
15. Kohonen T. Analysis of simple self-organizing process // Biological Cybernetics. 1982. V. 44. P. 135–140.
16. Fukushima K. Neural network model for selective attention in visual pattern recognition and associative recall // Appl. Opt. 1987. V. 26. № 23. P. 4985–4992.