

Малахова Екатерина Юрьевна

**ИССЛЕДОВАНИЕ И ИНТЕРПРЕТАЦИЯ
ПРИ ПОМОЩИ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ
МЕХАНИЗМОВ ОПИСАНИЯ ЗРИТЕЛЬНЫХ ОБРАЗОВ В
НИЖНЕВИСОЧНОЙ КОРЕ ГОЛОВНОГО МОЗГА ПРИМАТОВ**

1.5.5 – Физиология человека и животных

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата биологических наук

Санкт-Петербург — 2025

Работа выполнена в Лаборатории физиологии зрения Федерального государственного бюджетного учреждения науки Институт физиологии им. И. П. Павлова Российской академии наук

Научный руководитель:

Шелепин Юрий Евгеньевич, доктор медицинских наук,
профессор, заведующий лабораторией физиологии зрения
Института физиологии им. И. П. Павлова РАН

Оппоненты:

Кропотов Юрий Дмитриевич, доктор биологических наук,
профессор, заведующий лабораторией нейробиологии
программирования действий ФГБУН «Институт мозга человека
им. Н. П. Бехтерева РАН»

Бондарь Игорь Вячеславович, доктор биологических наук,
профессор, заведующий лабораторией физиологии сенсорных систем,
заместитель директора по научной работе
ФГБУН «Институт Высшей Нервной Деятельности и Нейрофизиологии РАН»

Ведущая организация: Федеральное государственное
бюджетное учреждение науки «Институт эволюционной
физиологии и биохимии им. И. М. Сеченова Российской
академии наук»

Защита состоится «___» _____ 2025 года в ____ часов на заседании
диссертационного совета _____ при ФГБУН Институт физиологии им. И.П.
Павлова РАН (199034, Санкт-Петербург, наб. Макарова, 6, тел. (812)328-07-01, электронная
почта pavlov.institute@infran.ru, сайт <http://infran.ru>).

С диссертацией можно ознакомиться в научной библиотеке ФГБУН Институт
физиологии им. И. П. Павлова РАН (Санкт-Петербург, наб. Макарова, д.6) и на сайте
Института <http://infran.ru>

Автореферат разослан «___» _____ 2025 года

Ученый секретарь Диссертационного совета,
доктор биологических наук

Дик Ольга Евгеньевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Зрительная система позволяет живым организмам получать информацию о состоянии окружающей среды, осуществлять целенаправленную деятельность и эффективно взаимодействовать с другими организмами и окружающей средой. Понимание принципов зрительного восприятия является одной из ключевых задач нейробиологии.

Распознавание образов напрямую связано с выделением статистических свойств воспринимаемой (наблюдаемой) зрительной сцены. Особенности окружающей среды также оказывают существенное влияние и на формирование самой зрительной системы (Глезер, Цуккерман, 1961; Field, 1989, 1999). Изучению свойств натуральных сцен и их взаимосвязи с характеристиками нейронов в зрительной коре посвящено большое количество научных работ в отечественной и зарубежной литературе.

Сравнительно новым подходом к изучению статистических свойств воспринимаемых изображений является использование искусственных нейронных сетей. Построенные на принципах, присущих зрительной системе приматов, сверточные нейронные сети (Fukushima, 1980; LeCun et al., 1989) вывели задачу распознавания образов на новый уровень, зарекомендовав себя не только как универсальный инструмент для работы с изображениями, но и как наиболее точная модель зрительной системы. Современные исследования (Cadieu et al., 2014; Khaligh-Razavi, Kriegeskorte, 2014; Yamins et al., 2014; Schrimpf et al., 2018) демонстрируют высокую степень функционального сходства между глубокими слоями сверточных нейронных сетей и нижневисочной корой приматов.

Получение новых сведений о принципах обработки информации в зрительной коре головного мозга с использованием нейросетевых моделей и их связи с реальными физиологическими механизмами восприятия представляет интерес как с научной точки зрения, так и для практического применения. Это может способствовать более глубокому пониманию процессов восприятия и принятия решений в биологических системах, а также улучшению стабильности и адаптивности компьютерных систем обработки изображений, что особенно важно в таких областях, как клиническая диагностика (Шелепин и др., 1985), транспортные средства с автоматическим управлением, системы обеспечения безопасности и т. д.

Приоритетным направлением в нейрофизиологии является и разработка методов повышения качества процесса исследования при помощи искусственного интеллекта. Помимо технологических инноваций, направленных на повышение точности работы оборудования и анализ больших массивов данных, можно выделить проекты, нацеленные на усовершенствование самого экспериментального поиска. Например, методы управления ходом эксперимента, такие как поиск оптимальных настроек экспериментального оборудования при помощи байесовской оптимизации (Lancaster et al., 2018) либо использование алгоритмов активного поиска для максимизации получаемой информации за счет адаптивного подбора стимулов (DiMattina, Zhang, 2013; Gardner et al., 2015). В диссертационном исследовании развивается данное направление и рассматривается возможность направленного создания стимульного материала во время проведения экспериментальной сессии при помощи генеративных методов искусственного интеллекта.

Актуальность данного исследования обусловлена значительным фундаментальным и практическим интересом к пониманию механизмов обработки информации в мозге человека и приматов, а также конвергенцией физиологии и вычислительного моделирования (Hassabis et al., 2017) в изучении этих процессов. Внимание к данной проблематике отражается в многочисленных международных конференциях и научных журналах, посвященных нейронаукам и междисциплинарным подходам к исследованию работы мозга. Развитие этого направления не только углубляет понимание о нейронных механизмах восприятия, но и способствует созданию более точных моделей для применения в искусственном интеллекте и биомедицине.

Степень разработанности темы

На сегодняшний день хорошо изучены функции нейронов первичной зрительной коры (поле 17 по Бродману BA17), известна организация их рецептивных полей (Hubel, Wiesel, 1962; Шелепин, 1981; Шевелев, 1984; Olshausen, Field, 1996), функциональная и структурная организация проводящих каналов зрительной системы, в том числе наличие специализированных путей распространения информации (Глезер, 1966; Кок, 1967; Mishkin, Ungerleider, 1982; Праздникова и др., 1985).

Значительный вклад внесен в понимание принципов пространственно-частотной фильтрации на начальных этапах обработки зрительного сигнала (Глезер, Цуккерман, 1961; Цуккерман, 1975; Шелепин, 1981) и последующей согласованной фильтрации на более высоких уровнях зрительной системы (Глезер, 1985). Разработаны модели согласованной фильтрации и распознавания объектов в условиях шума и вариаций входных паттернов (Красильников 1958, 1986; Grossberg, 1976).

Проводятся исследования процессов и на высших уровнях зрительной системы – заднетеменной и нижневисочной коры (Кок, 1967; Глезер и др., 1975; Яковлев, 1982; Праздникова и др., 1989). Внесен большой вклад в изучение механизмов инвариантности зрительного восприятия как основополагающего принципа распознавания объектов (Глезер, 1966; Шелепин, 1973; Яковлев, 1982; Tanaka, 1996).

Исследования высших отделов зрительной системы, таких как нижневисочная кора, позволили выявить сложную иерархическую организацию переработки зрительной информации (Felleman & Van Essen, 1991). Показано, что на более высоких уровнях коры визуальные представления объектов становятся все более абстрактными, что играет ключевую роль в механизмах инвариантности восприятия (Tanaka, 1996; Logothetis et al., 1995; DiCarlo et al., 2012). В частности, клетки нижневисочной коры избирательно реагируют на сложные стимулы, такие как формы, лица и объекты, обеспечивая устойчивое восприятие их идентичности независимо от изменений ракурса или масштаба (Gross et al., 1972; Logothetis et al., 1995). В ряде работ было продемонстрировано, что для распознавания объектов в коре формируются представления, чувствительные к углу обзора, но инвариантные к ряду других признаков (Logothetis, Sheinberg, 1996); инвариантность может достигаться за счет непрерывного кодирования признаков близлежащими нейрональными колонками (Tanaka, 1996); кодирование объектов может осуществляться при помощи различных комбинаций активных и неактивных колонок для отдельных признаков (Tsunoda et al., 2001). Показано, что нейроны в пределах одной функциональной зоны имеют сходную

избирательность, но различаются на более локальном уровне, формируя мозаичную структуру представления объектов (Sato, Uchida, Tanifuji, 2009; Sato et al., 2013).

Для анализа нейрональной активности разработаны различные методы, включая индексы чувствительности и селективности (Aparicio, Issa, DiCarlo, 2016; Bell et al., 2011), построение кривых настройки (Doya et al., 2011), однако эти методы имеют ограничения при изучении высокоуровневой нейронной активности. Так как нейроны высших областей реагируют на более сложные стимулы, что делает затруднительной их интерпретацию через простые функции, как в случае с ориентационной избирательностью нейронов первичной зрительной коры.

Следует отметить, что хотя указанные основополагающие работы и значительно продвинули понимание роли и взаимодействия различных областей мозга в распознавании объектов, тем не менее, точные принципы и механизмы этого процесса до сих пор не до конца понятны и остаются предметом активных исследований. Для изучения работы зрительной системы применяются методы *in silico* моделирования и машинного обучения, позволяющие охватить различные аспекты сенсорного восприятия, исследовать механизмы представления информации и характеристики отдельных нейронов.

Обнаруженные сходства между работой сверточных нейронных сетей, оптимизированных для задач распознавания, и ответами нейронов зрительной коры (Yamins et al., 2014; Kriegeskorte, 2015) открыли целое направление исследований, посвященное изучению свойств моделей и их совершенствованию. Развитие компьютерных наук и применение новых моделей создало возможности для появления совершенно новых подходов к изучению механизмов кодирования информации в нейронных сетях (Ponce et al., 2019; Bashivan, Kar, DiCarlo, 2019), включая выявление избирательности отдельных нейронов к различным признакам зрительных стимулов.

Целью диссертационной работы является изучение и интерпретация механизмов описания зрительных образов нейронными сетями высших отделов зрительной системы приматов, а именно нижневисочной коры, обеспечивающих распознавание объектов и оценку их значения для наблюдателя, в модельных исследованиях с применением искусственных нейронных сетей, выполняющих аналогичные задачи распознавания.

Основные задачи диссертационной работы:

1. Планирование исследований, предварительная обработка и статистический анализ данных нейрональной активности нижневисочной коры для задач моделирования процесса распознавания зрительных образов.

2. Моделирование ответа нейронов нижневисочной коры на предъявляемые зрительные стимулы при помощи сверточных нейронных сетей; сопоставление модели с нейрофизиологическими данными.

3. Разработка и реализация методов и алгоритмов исследования кодирования зрительной информации и интерпретации работы моделей зрительной системы на различных этапах распознавания, применимых как для изучения свойств искусственных нейронных сетей, так и для обработки данных нейрофизиологических исследований.

4. Исследование взаимосвязи статистических свойств изображений и функций нейронов на различных уровнях обработки зрительного сигнала.

Научная новизна

В работе предложен новый подход к исследованию функций нейронов высших отделов зрительной системы посредством применения генеративных нейронных сетей, а также разработан программный комплекс, позволяющий создание стимулов непосредственно во время проведения нейрофизиологических экспериментов на основе регистрируемой нейрональной активности.

Впервые показана возможность контролируемой активации нейрональных колонок нижневисочной коры при помощи создания абстрактных экспериментальных стимулов во время проведения нейрофизиологического эксперимента. Проанализированы характеристики подхода и их влияние на нейрональный ответ.

Проведен многосторонний анализ применения искусственных нейронных сетей в модельных исследованиях высших отделов коры. Получены новые данные о пространстве описания информации как отдельными нейронами, так и популяциями, включая влияние характеристик сигнала и задачи на формирование этого пространства. Описан характер преобразования зрительной информации на различных этапах обработки сигнала.

Положения, выносимые на защиту

1. Разработанный метод адаптивной генерации стимулов, использующий генеративно-состязательные сети в сочетании с моделями нейронных ответов нижневисочной коры, позволяет создавать изображения, вызывающие целенаправленную активацию нейронных ансамблей, и открывает возможность исследования функциональной специализации нейронов без ограничения фиксированными наборами экспериментальных стимулов.
2. Показаны оппонентные отношения между пространством описания сигнала и пространством постановки задачи, определяющие структуру и динамику преобразования информации на разных этапах обработки зрительных стимулов в нейронных сетях. В частности, описание зрительной сцены на низкоуровневых этапах (начальные сверточные слои) отражает статистику входного сигнала, тогда как высокоуровневые (как сверточные, так и полносвязные слои) – адаптируются под цели и задачи деятельности наблюдателя.
3. Преобразование зрительной информации носит фазовый характер: описание зрительных образов, изначально подчиняющееся структуре входного сигнала, подвергается резкой перестройке и переходит в продолжительную стадию, сопровождающуюся трансформацией представлений, максимальной сложностью описания и распределенным кодированием, после чего происходит процесс интенсивной консолидации признаков и переструктурирование описания в соответствии со стоящей перед испытуемыми задачей.

Теоретическая и практическая значимость работы

Теоретическая значимость работы заключается в углублении понимания нейрофизиологических механизмов кодирования зрительной информации в высших отделах зрительной системы приматов. Методологически значимым является предложенный подход к изучению функций нейронов высших областей зрительной

системы, основанный на визуализации высокоуровневых репрезентаций с использованием генеративно-состязательных нейронных сетей.

Практическая значимость заключается в разработке и реализации предложенных методов в виде программных решений, применимых при проведении нейрофизиологических исследований зрительной системы. На основе данных микроэлектродной регистрации нейронов нижневисочной коры макак выполнено моделирование зрительного восприятия, включая распознавание образов и предсказание нейронных откликов с учетом семантики и контекста зрительных репрезентаций. Показана схожесть откликов искусственных нейронных сетей с характеристиками зрительного восприятия как на уровне принятия решений, так и на уровне активности кортикальных колонок нижневисочной коры. Рассмотрены особенности работы с искусственными нейронными сетями, как моделью зрительной системы приматов, выделены сходства, отличия и ограничения моделей.

Собран и систематизирован обширный материал, касающийся механизмов распознавания зрительных образов, что способствует дальнейшему изучению кодирования и интерпретации нейронной активности. Разработанные подходы позволяют углубить понимание взаимосвязи между сигналами, задачами и процессами формирования репрезентаций, а также найти новые пути для интеграции нейрофизиологических данных с методами искусственного интеллекта.

Методология и методы исследования

В методологию исследования включены методы экспериментальной работы – микроэлектродная запись нейрональной активности зрительной коры макак-резус; анализ данных параметрическими и непараметрическими статистическими методами, методы математического анализа и машинного обучения для исследования статистических закономерностей в данных; моделирование процессов распознавания изображений на основе искусственных нейросетей сверточной архитектуры, генеративно-состязательных и диффузионных сетей; изучение функциональных характеристик полученных моделей.

Степень достоверности и апробация результатов исследования

Достоверность научных положений и выводов, полученных в данной диссертационной работе, обеспечивается результатами экспериментальных исследований, успешным представлением основных положений в докладах на ведущих международных конференциях, согласованностью результатов диссертационной работы с результатами других авторов.

Основные результаты по теме диссертационной работы опубликованы в 16 печатных изданиях и доложены на 14 устных и 6 стендовых докладах следующих научно-практических конференций:

Neuroscience International Conference 2019, Chicago, USA; Vision Science Society International Conference 2019, St. Pete Beach, Florida, USA; Computational and Mathematical Models in Vision (MODVIS) Workshop at Vision Sciences Society International Conference 2018, St. Pete Beach, Florida, USA; Mutual Benefits of Cognitive and Computer Vision (MBCC) Workshop at The Computer Vision Foundation (CVPR) 2018, Salt Lake City, Utah, USA; The 5th IEEE International Conference on Video and Audio Signal Processing in the Context of

Neurotechnologies (SPCN) 2020, Taoyuan, Taiwan, 2017, 2018, 2020, 2021, St. Petersburg, Russia; WiML workshop at Neural Information Processing Systems (NeurIPS) 2018, Montreal, Canada; Всероссийская конференция с международным участием Интегративная Физиология 2019, Санкт-Петербург, Россия; XXIII Съезд Физиологического общества им. И. П. Павлова, 2018, Воронеж, Россия; Международная научная конференция Прикладная Оптика 2018, Санкт-Петербург, Россия; Современные аспекты интегративной физиологии 2018, Санкт-Петербург, Россия; Симпозиум, посвященный 100-летию Физиологического общества им. И.П. Павлова 2017, Санкт-Петербург, Россия; Технологическая перспектива в рамках Евразийского пространства 2017, Санкт-Петербург, Россия.

Публикации автора по теме диссертации

Основные результаты по теме диссертации изложены в 16 печатных изданиях. Из них 5 статей опубликованы в рекомендованных ВАК Минобрнауки РФ рецензируемых научных журналах; 2 – в научных изданиях, индексируемых в наукометрических базах Scopus, WoS; 7 — в тезисах докладов Всероссийских и международных конференций; 2 — в других печатных изданиях.

Личный вклад состоит в выполнении основного объема приведенных в диссертационной работе теоретических и экспериментальных исследований. Автор непосредственно участвовала в формулировании цели и задач исследования, разработке дизайна экспериментов и выборе методов исследования. Нейрофизиологические данные были предоставлены лабораторией М. Танифуджи, RIKEN Brain Science Institute (Япония) по итогам проведенных ранее экспериментов, а также по результатам эксперимента, разработанного и проведенного при участии автора работы. Автором лично были проведены обработка нейронных данных и статистический анализ, применены методы машинного обучения и проведено прототипирование архитектур нейронных сетей, предложены подходы к исследованию функциональных характеристик ответа нейронов, интерпретированы полученные данные и результаты экспериментов, создано и настроено необходимое программное обеспечение, изучена и проработана отечественная и зарубежная литература по теме диссертации, сформулированы результаты и выводы, подготовлены публикации по диссертационной теме. В методе интерпретации и аппроксимации функции нейронов через набор признаков в пространстве локальных ориентаций и цветов автором был проведен статистический анализ данных, а также проведено моделирование ответа искусственной нейронной сети.

Структура и объем диссертации

Диссертация состоит из введения, четырех глав, содержащих описания материалов и методов исследования, полученных результатов, а также обсуждения результатов, выводов, заключения, списка сокращений и списка использованных литературных источников. Объем диссертации составляет 141 страницу печатного текста, включая 3 таблицы и 57 рисунков. Список цитируемой литературы содержит 171 источник.

Финансовая поддержка работы

Автор участвовала в качестве исполнителя работ по гранту РФФИ 14-15-00918, «Технологии оптимизации и восстановления когнитивных функций человека виртуальной средой», 2014–2016 гг., а также в Программе ПРАН, 2014–2017, проект «Сенсорно-моторные механизмы деятельности человека в реальном и виртуальном пространстве» под руководством Шелепина Ю.Е. Также автором было получено 2 гранта на участие в международных конференциях в США от организаторов конференций.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

1. Сбор экспериментальных данных и статистический анализ откликов нейронов нижневисочной коры

Автором проведен ряд модельных и экспериментальных исследований с целью изучения и интерпретации механизмов описания зрительных образов нейронными сетями высших отделов зрительной системы приматов (Рис. 1).

1.1. Регистрация нейрональной активности

Исследования функциональных характеристик нейронов высших отделов зрительной системы проводились на приматах (макаки), зрительная система которых во многом схожа со зрительной системой человека. Данные были получены методом внеклеточной микроэлектродной записи в серии экспериментов в лаборатории Танифуджи М. в RIKEN Brain Science Institute (Япония).

Регистрация нейрональной активности производилась в области нижневисочной коры у трех макаков в серии экспериментов (Sato, Uchida, Tanifuji, 2009; Sato et al., 2013). Экспериментальный протокол был одобрен комитетом по экспериментальным животным института RIKEN и соответствовал рекомендациям института RIKEN и Национальных институтов здоровья (NIH).

Электроды располагались в передней части нижневисочного отдела коры (Anterior Inferotemporal Cortex, область TEad), захватывающей заднюю стенку верхнетеменной борозды (Superior Temporal Sulcus), предполагая вовлеченность области в обработку сложных зрительных стимулов, включая распознавание лиц и тел.

Регистрация нейронального ответа производилась в 190 точках записи у трех макаков (у Н1 – 33 точки записи, Н2 – 133, у Н3 – 24 соответственно) (Sato, Uchida, Tanifuji, 2009; Sato et al., 2013). Регистрация выполнялась при помощи массивов из восьми микроэлектродов Utah (Blackrock Neurotech), внедренных в кору головного мозга перпендикулярно ее поверхности. Полученный сигнал усиливался и пропусклся через полосовой фильтр с частотой 500 Гц – 3 кГц, после чего оцифровывался с частотой 25 кГц. Наличие спайка фиксировалось, в случае если электрический сигнал превышал среднее значение более чем на 3,5 стандартных отклонения.

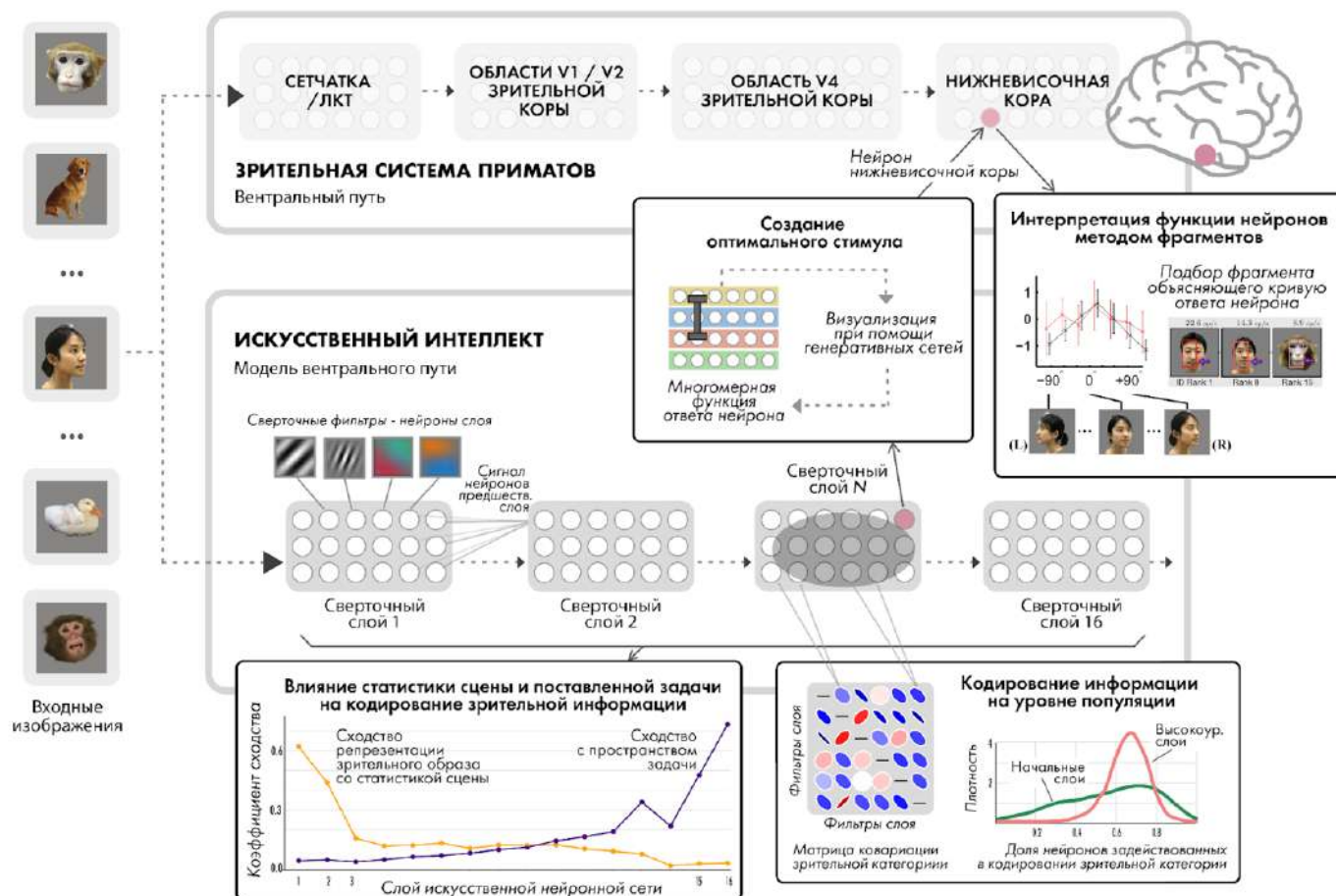


Рисунок 1. Исследование процесса распознавания образов в зрительной системе и искусственных нейронных сетях

1.2. Дизайн эксперимента и стимульный материал

Для нейронов нижневисочной области периодом фоновой активности рассматривался интервал от -50 до 80 мс после выведения стимула на экран (Nam et al., 2021), период от 100 до 220 мс рассматривался как период релевантного нейронального ответа. Активность учитывалась в целом по нейрональной колонке – данные, полученные с одного микроэлектрода усреднялись по 8-ми каналам. Для получения вызванного ответа из частоты пульсации в релевантный период вычиталась частота пульсации в предварительный период. Вызванный нейрональный ответ был усреднен по блокам с повторениями стимулов. Общая схема предъявления стимульного материала приведена на Рис. 2.

В экспериментах использовался набор изображений Takayuki-1550 (Sato, Uchida, Tanifuji, 2009), состоящий из 1550 стимулов, в числе которых следующие категории: лица людей (535 изображений), лица макак (285), тела макак (50), тела людей (50), другие животные (200), цветы и растения (120), пейзажи и природные объекты (120), искусственно созданные предметы и объекты (150), перевернутые лица обезьян и людей (40). Включение в набор данных широкого спектра категорий позволяет более точно определить избирательность нейронов к тем или иным признакам.

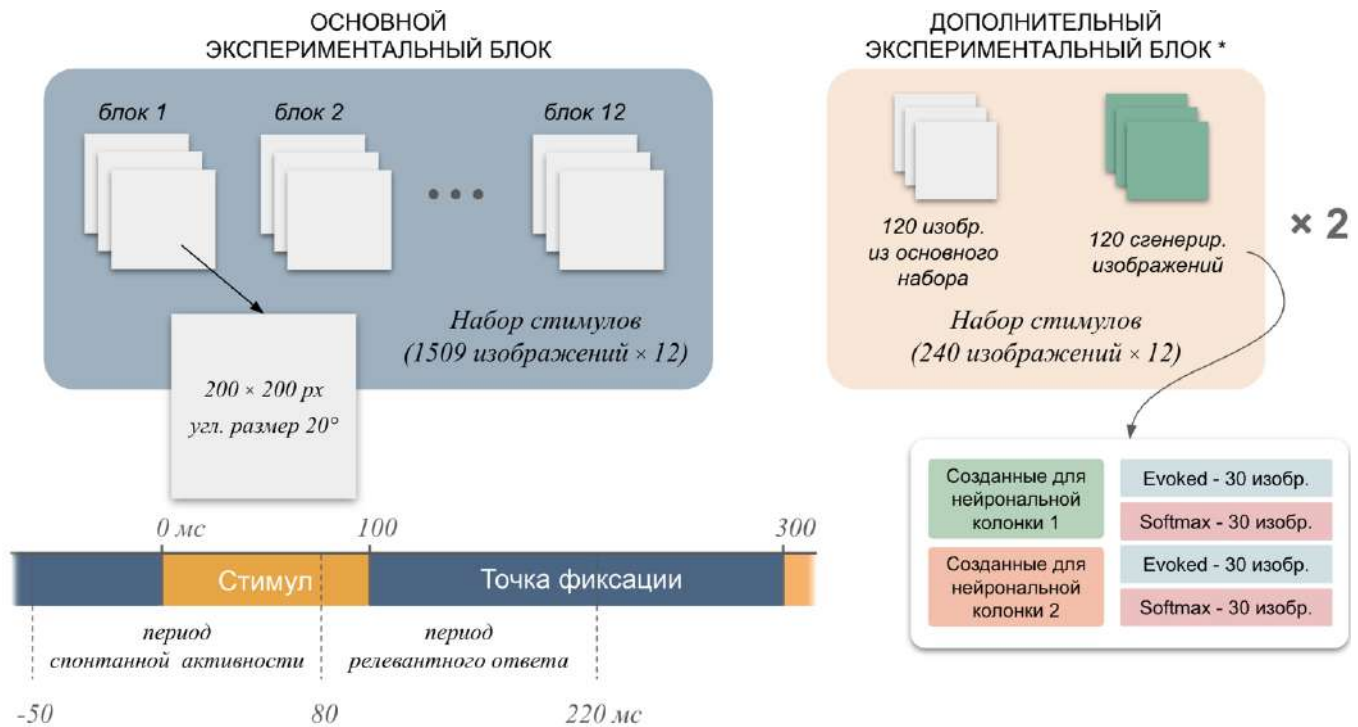


Рисунок 2. Дизайн экспериментального исследования нейрональной активности в нижневисочной коре. *Дополнительные экспериментальные блоки были включены только в часть исследования, посвященную оценке эффективности сверточных и генеративно-состязательных нейронных сетей

1.3. Характеристики импульсной активности нейронов

Регистрация вызванного нейронального ответа производилась в 190, однако из них в анализе рассматривались только 143 точках записи, остальные были исключены по техническим причинам.

Стабильность ответа оценивали с помощью t-критерия Стьюдента, сравнивая отклики на идентичные стимулы в четных и нечетных блоках. В 78 точках зафиксирована стабильная реакция ($p < 0,05$).

Анализ пространственной организации активности выявил закономерную связь: чем ближе расположены электроды, тем выше корреляция их сигналов (коэф. Спирмена = 0,41, $p \leq 0,001$). Аппроксимация данных показала, что наиболее адекватно взаимосвязь описывается степенной ($R^2 = 0,248$) и логарифмической ($R^2 = 0,239$) зависимостями, что согласуется с локальной организацией кортикальных ансамблей.

Исследование вариативности ответов установило зависимость между средней вызванной активностью и стандартным отклонением: нейроны с более сильным откликом демонстрировали большую изменчивость (коэф. Спирмена = -0,58, $p \leq 0,001$). Обнаружена область средней активности (5–15 спайков/с), в которой наблюдается широкий разброс стандартного отклонения, что может отражать сложную динамику работы нейронов.

1.4. Избирательность к категориям

Сравнение ответов на изображения лиц и других объектов показало, что 130 из 143 нейронов демонстрируют усиленную активность на лица (разница 8,04 спайков/с, SD = 6,88, $p < 0,001$), а 118 – увеличение вариативности (разница 3,51, SD = 3,70, $p < 0,001$). Это подтверждает специализацию данной области в обработке информации о лицах.

Расчет индекса избирательности (K) выявил, что 100 точек имели $K \geq 0,1$, а 59 – $K \geq 0,2$. Максимальная избирательность отмечена для лиц обезьян (71 точка) и лиц людей (39 точек), тогда как только 2 точки проявили избирательность ко всем трем лицевым категориям (лица людей, лиц обезьян, перевернутые лица), что свидетельствует о специализированных популяциях нейронов. Выраженная избирательность $K \geq 0,2$ для лиц обезьян (7 точек) подтверждает существование специфических нейронных ансамблей, ориентированных на биологически значимые стимулы. Напротив, стимулы искусственных объектов и ландшафтов вызвали отрицательные значения индекса ($-0,1$ и $-0,14$) по отношению к другим категориям, указывая на снижение активности при их предъявлении.

2. Применение генеративно-состязательных сетей в целях создания оптимального стимула для модели кортикальной колонки

Исследуются возможности применения сверточных и генеративно-состязательных нейронных сетей для изучения функций нейронов нижневисочной коры приматов. В основе подхода лежит моделирование нейрональных ответов с использованием глубоких сверточных сетей, анализируемых методами интерпретации искусственного интеллекта. Генеративно-состязательные сети применяются для поиска стимулов, вызывающих специфические реакции в моделируемых нейронах, что позволяет изучать их избирательность.

Подход предполагает создание изображений, направленных на максимизацию активации целевых нейронов (модельных или биологических). Нахождение такого изображения, в свою очередь, помогло бы понять функциональное назначение нейрона. Создание стимула начинается с генерации случайного вектора в пространстве признаков. Вектор затем переводится в изображение и предъявляется наблюдателю, регистрируется ответ. Затем осуществляется пошаговое изменение вектора и, соответственно, стимула таким образом, чтобы усилить признаки, наиболее значимые для целевого нейрона, что приводит к постепенному формированию характерного паттерна. Этот процесс позволяет выявить особенности избирательности нейрона, показывая, какие элементы структуры или текстуры оказывают наибольшее влияние на его активацию. Разработанная методология была протестирована в двух вариантах:

- ☐ офлайн-метод – построение модели и анализ ответов нейронов на основе ранее собранных данных,
- ☐ онлайн-метод – динамическое коррекция стимуляции во время эксперимента.

2.1. Моделирование ответа нейрональных колонок

Реализация метода визуализации при помощи генеративно-состязательных сетей предполагает оптимизацию стимула на основе ответа наблюдателя. В технических науках этот процесс осуществляется через сверточные нейронные сети или другие модели, непосредственно оптимизирующие подаваемый входной сигнал. Такая оптимизация может выполняться быстро, проходя большое количество итераций. Однако в реальных экспериментах количество предъявляемых наблюдателю стимулов существенно ограничено. Поэтому целесообразным является введение прокси-модели, обученной на ответах нижневисочной коры и представляющей собой аппроксимацию нейрональных колонок этой области мозга. Создание стимулов осуществляется посредством

взаимодействия генеративных моделей и моделей наблюдателя, после чего созданный материал предъявляется непосредственно биологическому наблюдателю. Такой подход позволяет значительно ускорить процесс экспериментального поиска и сократить количество необходимых предъявлений живому субъекту.

Для офлайн моделирования использовались данные об активности 143 кортикальных колонок нижневисочной коры макака. Усредненный по 12 блокам отклик использовался как общий нейрональный ответ. Данные были разделены на тренировочное (1240) и тестовое (310) множества, на котором оценивалась точность предсказаний.

Модель включала сверточный блок, использующий предобученные слои VGG16 для извлечения признаков, и полносвязную надстройку из 143 отдельных подсетей, что снижало корреляцию предсказаний и лучше соответствовало нейробиологическим данным. Средний коэффициент корреляции модели с реальными нейронными откликами составил 0,68 ($p < 0,01$), при этом точность предсказаний зависела от стабильности ответа нейронов ($r = 0,7$, $p < 0,01$).

Качество предсказаний варьировалось между колонками (среднее $r = 0,45$), при этом в некоторых колонках (E1, E2, H8, M5, E7) достигало $> 0,8$, что подтверждает применимость модели для определенных нейронных популяций. Далее модель использовалась для поиска оптимальных стимулов с помощью генеративно-состязательных сетей, ограничивающих пространство поиска для создания естественно выглядящих изображений.

2.2. Визуализация функциональных характеристик модели

После обучения модели было проведено исследование ее характеристик. Приведем пример оригинальных экспериментальных стимулов и сгенерированных изображений, ассоциирующихся с активацией кортикальных колонок № 25 и № 27 (Рис. 3). Заметно, что сгенерированные стимулы фокусируются преимущественно на глазах, а также на абстрактных формах, которые, вероятно, отражают другие признаки, выделяемые рассматриваемым нейрональным участком. Например, во всех стимулах, созданных для колонки № 25 (Рис. 3 а, б) прослеживается наличие диагональной линии, направленной с левого верхнего угла в правый нижний.

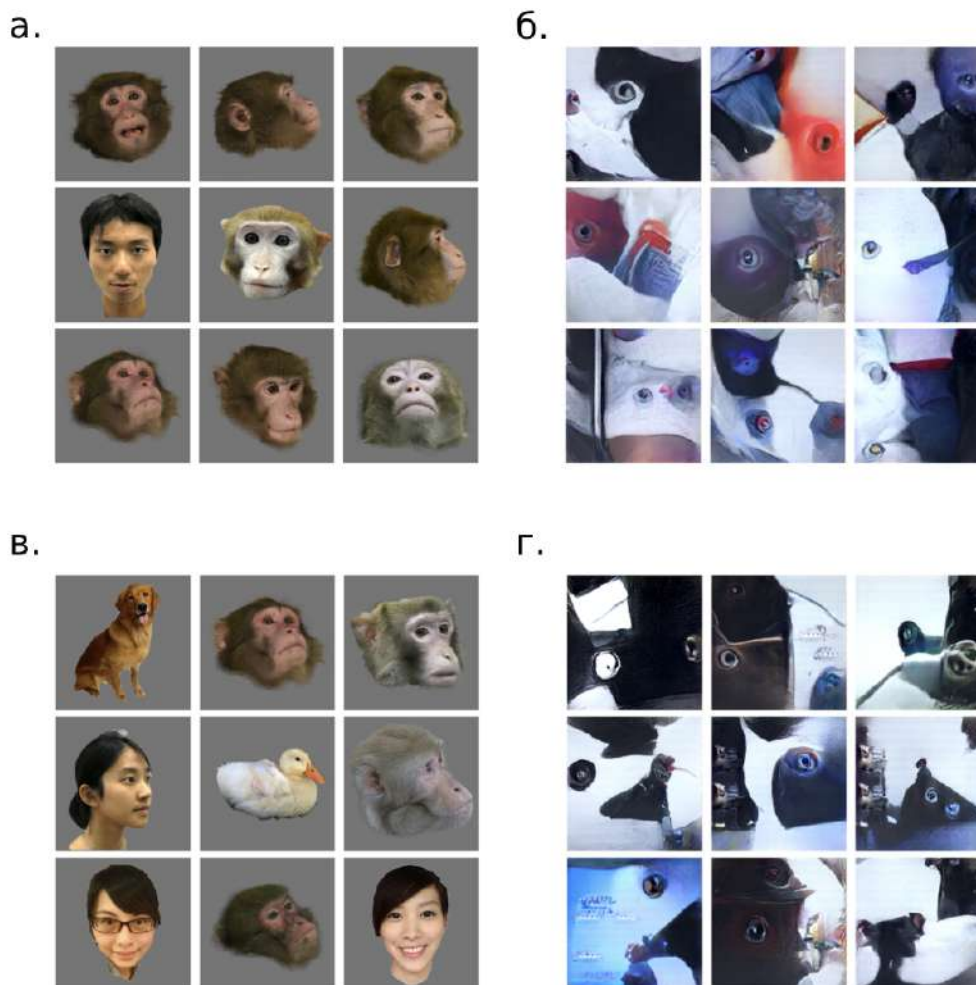


Рисунок 3. Сопоставление стимулов, вызывающих высокую активацию нейрональной колонки и изображений, сгенерированных при помощи модели для интерпретации функций нейрона: а) топ-9 экспериментальных стимулов для кортикальной колонки № 25; б) топ-9 сгенерированных изображений, вызывающих реакцию в № 25; в) топ-9 экспериментальных стимулов кортикальной колонки № 27; г) топ-9 сгенерированных изображений, вызывающих реакцию в № 27

2.3. Применение подхода во время экспериментальной сессии

Для валидации этого подхода был проведен эксперимент (см. Рис. 2, дополнительные блоки), в котором модель обучалась на данных основного блока с естественными изображениями, а затем использовалась для генерации стимульного материала в дополнительных блоках экспериментальной сессии. Предполагалось, что если модель эффективно предсказывает нейронные отклики, а метод визуализации с использованием генеративно-состязательных сетей применим, то сгенерированные стимулы должны вызывать отклик в целевых нейронах.

Эксперимент проводился с регистрацией нейрональной активности в нижневисочной коре макаки при помощи массива Utah из 8 электродов. Он состоял из трех блоков: в первом предъявлялись естественные изображения из набора Такауки-1550, а в последующих – как естественные, так и сгенерированные изображения, созданные на основе обучения модели по данным предыдущих блоков.

Для генерации стимулов было разработано программное решение, включающее модули сбора данных, оптимизации и генерации изображений. Стимулы создавались с использованием двух функций потерь: Evoked (максимизация активации в конкретной точке

регистрации) и Softmax (активация в целевой точке при минимизации ответа в других). Результаты показали, что обученная модель обладала предсказательной способностью (коэффициент корреляции 0,49 между предсказанными и реальными ответами), хотя эффективность была ниже, чем на тестовом подмножестве (0,73).

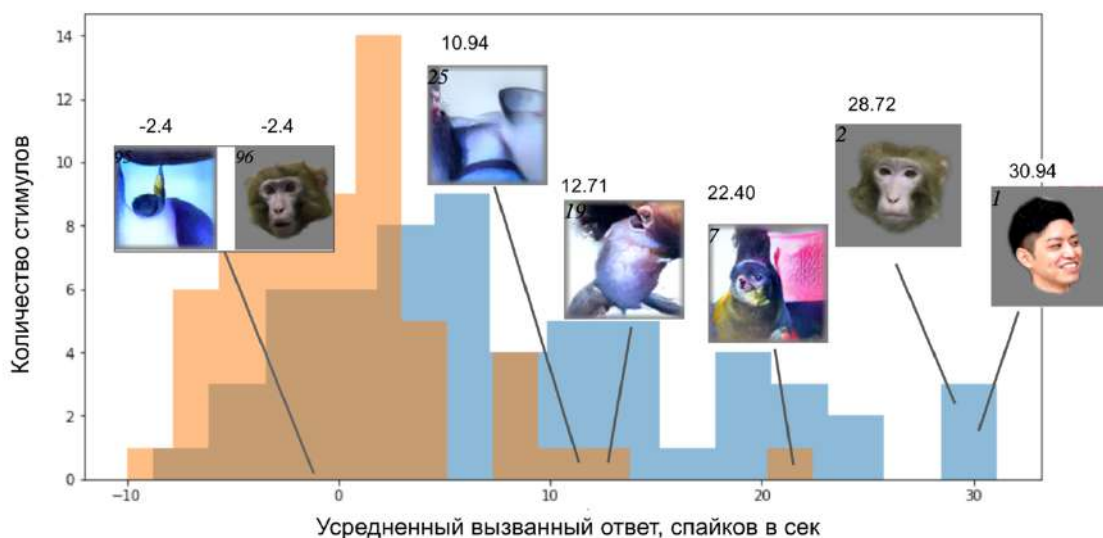


Рисунок 4. Распределение вызванного нейронального ответа для естественных и сгенерированных стимулов в точке регистрации Site 1, блок 1. Голубым цветом показано распределение ответов на естественные стимулы, оранжевым – на сгенерированные изображения

Анализ распределений нейрональных ответов подтвердил, что некоторые сгенерированные стимулы вызывали активацию, сопоставимую с естественными изображениями. Особый интерес представляло наблюдение, что абстрактные сгенерированные изображения без явных черт лиц могли вызывать нейрональные ответы, сравнимые с фотографиями реальных лиц, что позволяет предположить реакцию нейронов на определенные геометрические паттерны, а не на целостные образы. Разработанный метод демонстрирует принципиальную эффективность для создания целевых стимулов в нейрофизиологических исследованиях и открывает новые возможности для понимания механизмов кодирования информации в высших отделах зрительной коры.

3. Модельные исследования работы нейронных сетей в процессе распознавания образов объектов

Для исследования кодирования информации в процессе зрительного восприятия был разработан метод, основанный на оценке схожести пространств описания. Данный подход позволяет проследить изменения в представлении информации на разных уровнях нейронной сети путем сопоставления сходства объектов, спроецированных в пространство признаков. Предложенный метод включает расчет матриц сходства для изображений на различных этапах обработки и их сравнение с референтными матрицами, отражающими статистику сигнала и функциональное пространство задачи.

3.1. Исследование пространства представления информации посредством оценки схожести пространств описания

В модельных исследованиях с использованием сверточной нейронной сети VGG16 установлены оппонентные отношения между пространством описания сигнала и пространством постановки задачи (Рис. 5). Начальные слои сети демонстрируют высокое сходство с пространством сигнала (коэф. корр. = 0,62), которое значительно снижается после первых двух уровней нелинейных преобразований (коэф. корр. = 0,16). Одновременно с этим, сходство с пространством задачи постепенно возрастает, достигая значения 0,34 в последнем сверточном слое, что свидетельствует о значительной адаптации представлений под решаемую задачу.

Анализ показал, что в процессе обработки повышается уровень сходства между репрезентациями экземпляров одного класса (от $K=0.05\pm0.27$ в пространстве сигнала до $K=0.18\pm0.11$ в последнем сверточном слое). При этом увеличивается количество ортогональных репрезентаций, что говорит о декорреляции представлений объектов разных классов. В полносвязных слоях происходит дальнейшая консолидация признаков, и корреляция с пространством задачи возрастает до 0,74 в слое предсказания.

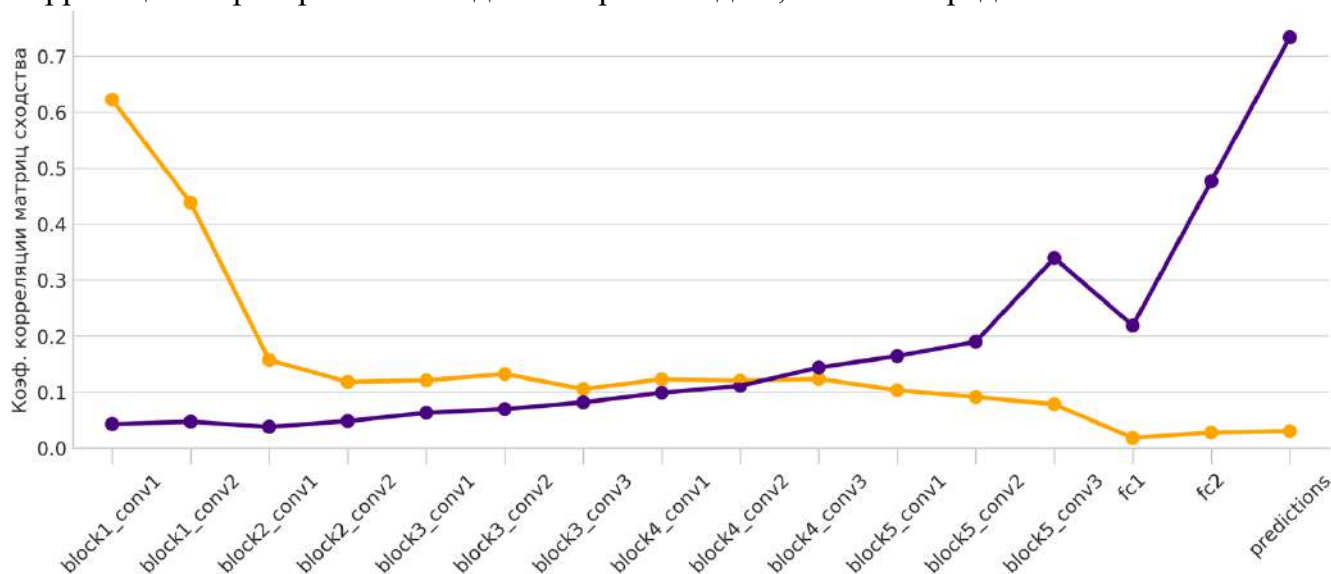


Рисунок 5. Влияние оппонентных факторов (сигнала и задачи) на формирование представлений информации в слоях нейронной сети. По оси абсцисс – слой нейронной сети VGG16, по оси ординат – корреляция между матрицей сходства изображений внутри слоя и референтной матрицей. Фиолетовым цветом обозначена корреляция с матрицей сигнала, оранжевым – с матрицей задания

Обнаружен фазовый характер обработки зрительной информации: сначала происходит резкий разрыв связей с входными данными, затем следует этап трансформации представлений и распределенного кодирования, а на заключительном этапе наблюдается консолидация признаков под задачи распознавания. Интересно, что для некоторых семантически близких категорий (например, два различных класса птиц) уровень сходства репрезентаций в высших слоях возрастает, несмотря на задачу их различения. Это указывает на формирование инвариантных признаков, характерных для родительского класса объектов, что можно рассматривать как создание «образа категории», где выделяются ключевые характеристики объекта при игнорировании несущественных различий.

Таким образом, представление информации в нейронной сети претерпевает значительные изменения от начальных слоев к конечным, постепенно трансформируясь от пространства, близкого к статистике входного сигнала, к пространству, организованному в соответствии с решаемой задачей. Промежуточные слои, где происходит наиболее сложная трансформация представлений, представляют особый интерес для понимания механизмов обработки зрительной информации в биологических системах.

3.2. Сложность представления зрительной категории

Для исследования представления зрительных категорий на уровне популяции нейронов был разработан метод, основанный на анализе ковариационных матриц активаций фильтров. Данный подход учитывает не только согласованность активаций, но и сложность распределенного паттерна их взаимодействия, формируя прототипы категорий путем усреднения ковариационных матриц группы изображений.

Сравнение данного метода с традиционным подходом, основанным на средних активациях фильтров, показало его преимущество в задаче классификации. Особенно значительная разница в эффективности (15–20%) была выявлена на промежуточных этапах обработки сигнала, что указывает на важность учета взаимосогласованной детекторов в этом периоде обработки данных. На поздних этапах различия между методами нивелируются, что может свидетельствовать о концентрации информации в отдельных детекторах.

Дополнительный анализ с использованием метода главных компонент позволил оценить сложность представления категорий в различных слоях сети. На первом слое репрезентации категорий сжаты в небольшое количество главных компонент (10 направлений объясняют 95% дисперсии), что указывает на их низкую сложность. На средних слоях (5–12) требуется гораздо больше главных компонент для описания представления категории (среднее 68%, максимум 77% в 10-м слое), что подтверждает гипотезу о высокой степени распределенности информации. На поздних слоях количество необходимых компонент вновь снижается, достигая 62% в последнем слое, что говорит о происходящем сжатии информации до наиболее важных для классификации осей.

Также был проведен анализ значимых направлений, который показал, что количество значимых направлений, представляющих категорию внутри слоя, постепенно увеличивается при продвижении вглубь сети. На тренировочных данных до 93% фильтров демонстрируют отличную от базовой линии реакцию активности, тогда как для валидационных данных это количество снижается до 57%–70%, что указывает на тонкую подстройку к обучающим изображениям и оптимизацию вычислительных ресурсов сети.

Полученные результаты свидетельствуют об ограниченной применимости методов, рассматривающих один нейрон в качестве детектора категории, и подчеркивают важность анализа коллективной активности для понимания принципов кодирования информации в нейронных сетях. Предложенный подход открывает новые возможности для сопоставления механизмов обработки информации в искусственных и биологических системах.

3.3. Интерпретация функции нейронов посредством фрагментов натуральных изображений

Исследование популяционного кодирования выявило сложные распределенные представления, где категории описываются согласованной активностью множества

нейронов. Для понимания работы такой системы необходимо установить, какие конкретные зрительные признаки формируют эти представления. Предложенный метод, разработанный в лаборатории Танифуджи М. в RIKEN Brain Science Institute, позволяет аппроксимировать функции нейронов через набор признаков в пространстве локальных ориентаций и цветов, извлеченных из фрагментов натуральных изображений.

Методология включала создание словаря из 560 000 фрагментов различного размера из базы данных VOC 2010, преобразованных в семь каналов (четыре ориентационных и три цветовых). Стимульные изображения также преобразовывались в аналогичное представление. Для каждой пары "фрагмент-стимул" вычислялись расстояния с учетом инвариантности к положению и масштабу, на основе чего формировался предсказанный нейрональный ответ. Оптимальные признаки отбирались по наивысшим коэффициентам корреляции между предсказанными и реальными нейрональными ответами.

Результаты показали, что признаки, кодируемые нейронами IT-коры, могут быть успешно аппроксимированы комбинациями локальных ориентаций и цветов. Модель объяснила до 81% вариабельности нейрональных ответов (показатель объясненной дисперсии 65,7%) и точно предсказала избирательность нейронов к определенным ракурсам лиц, причем 84,6% участков показали значимую корреляцию между нейронными и предсказанными кривыми настройки на поворот лица. Выделенные признаки относились к двум категориям: состоящие из множества локальных компонентов и локальные с небольшим числом компонентов.

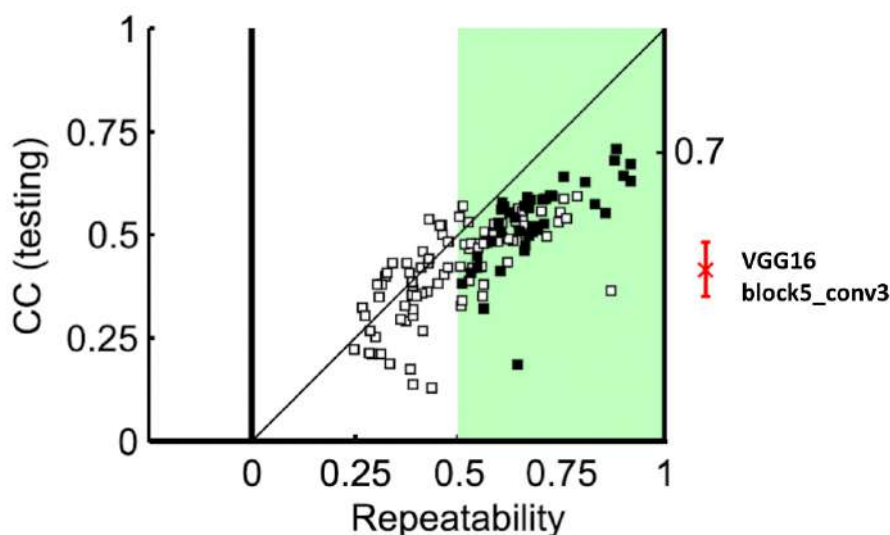


Рисунок 6. Эффективность предложенного метода фрагментов в предсказании нейронального ответа. Горизонтальная ось – повторяемость каждой колонки лиц, оцениваемая по корреляции между четными и нечетными усредненными ответами на изображения. Среди 152 колонок с избирательностью к лицам > 0.3 (черные прямоугольники), были отобраны 88 колонок с повторяемостью $> 0,5$ для проведения теста перекрестной валидации. Красным обозначена эффективность применения метода для интерпретации активаций ИНС. Источник: (Nam et al., 2021)

При применении подхода к сверточным нейронным сетям (VGG16 и VGG-Face) была показана возможность описания признаков, кодируемых глубокими слоями, при помощи фрагментов. Корреляция между предсказанными и фактическими векторами откликов

составила $0,42 \pm 0,13$ ($p < 0.05$ в 510 фильтрах из 512), что подтверждает эффективность метода для анализа искусственных нейронных сетей.

Сравнение распределения фильтров, избирательных к изображениям лиц, показало, что в VGG16 наибольшее количество таких нейронов сосредоточено в средних слоях, достигая ≈ 80 в слоях block4_conv3 и block5_conv2. Интересно, что в VGG-Face, несмотря на обучение исключительно на изображениях лиц, количество избирательных нейронов оказалось меньше. Это свидетельствует о том, что в VGG16 такие нейроны появляются как универсальный механизм обработки естественных изображений, тогда как в VGG-Face они выполняют функцию различения лиц, а не их детекции.

Данный подход продемонстрировал, что комбинирование визуальных признаков, полученных при помощи фрагментов, позволяет объяснить инвариантное к ракурсу распознавание лиц как в нижневисочной коре, так и в искусственных нейронных сетях. Результаты подтверждают, что высшие уровни обработки в вентральном зрительном пути могут быть аппроксимированы вычислениями, аналогичным работе неглубокой искусственной нейронной сети.

3.4. Распознавание образов в условиях неопределенности

Исследования зрительного восприятия в условиях неопределенности показывают, что зрительная система человека обладает механизмами адаптивной фильтрации, позволяющими минимизировать влияние внутреннего шума и фоновых помех. При бинокулярной интеграции объединение информации от обоих глаз снижает уровень внутреннего шума, улучшая детекцию зрительных стимулов. Зрительная система компенсирует шум за счет динамической регуляции чувствительности к различным пространственным частотам и подавления низкоуровневого шума на первичной стадии восприятия.

Было проведено исследование процессов распознавания образов в условиях неопределенности с использованием комбинации сверточных и диффузионных генеративных моделей и сопоставление полученных результатов с данными психофизических экспериментов. Диффузионная нейронная сеть применялась для создания стимульного материала с различной степенью зашумленности, имитирующего визуальную неопределенность. Сверточная нейронная сеть выполняла функцию наблюдателя, решая задачу классификации объектов на зашумленных изображениях в различных условиях восприятия. Такой подход позволил количественно оценить закономерности распознавания визуальных паттернов при дефиците сенсорной информации.

Для исследования использовалась диффузионная модель Stable Diffusion, которая позволяет пошагово восстанавливать изображение из шума. На начальных этапах генерации модель преимущественно восстанавливала низкочастотные компоненты изображения, формируя общую структуру объектов в сцене. Более сложные детали и высокочастотные элементы появлялись на поздних стадиях обработки.

Анализ точности распознавания объектов моделью наблюдателя выявил скачкообразный переход: на первых этапах генерации изображения классифицировались случайным образом, требуется более 60% итераций, чтобы изображения начали корректно классифицироваться. По достижению 80% происходит качественный скачок и изображения практически всегда (> 0.8) классифицируются корректно. Этот эффект напоминает

механизм порогового восприятия у человека, как например, в так называемом феномене «инсайта» в восприятии, когда разрозненные фрагменты информации неожиданно складываются в осмысленный образ. В начальных 40% итераций создания изображения диффузионными сетями структурная информация, измеряемая посредством SSIM, практически отсутствует. Во второй половине процесса происходит наиболее интенсивное формирование структурных элементов и других характеристик зрительной сцены.

Скачкообразный переход в точности распознавания может быть обусловлен тем, что сверточные нейронные сети в первую очередь опираются на локальные признаки изображения, а не на его глобальную структуру. В отличие от этого, зрительная система человека интегрирует информацию на нескольких уровнях, что позволяет гибко переключаться между локальными деталями и общей структурой объекта. Человеческая зрительная система с рождения функционирует в условиях постоянного зашумления сенсорного сигнала, где шум не является исключением, а представляет собой неотъемлемую часть процесса обработки информации. Мозг использует вероятностные стратегии, адаптивные механизмы подавления помех и интеграцию контекста, что позволяет стабильно распознавать объекты даже при значительных искажениях.

Исследование распознавания мимики в условиях неопределенности показало, что в надпороговых условиях точность распознавания эмоций человеком составляла 83%, а при предъявлении пороговых изображений вероятность распознавания мимики снижалась до случайного уровня (50%). ФМРТ-данные показали, что при восприятии выраженной мимики активируются фронтальные и теменные области, однако в условиях неопределенности возрастает активность нижневисочной коры и зон, связанных с избирательным вниманием, что свидетельствует о попытках компенсировать недостаток информации за счет дополнительного анализа визуальных признаков. Моделирование при помощи сверточной нейронной сети показало сопоставимые результаты, в условиях неопределенности сеть демонстрировала нестабильные активации, что указывает на недостаток семантической интерпретации, присущей биологической зрительной системе.

ЗАКЛЮЧЕНИЕ

В диссертационной работе исследуются механизмы кодирования и распознавания зрительных образов в нижневисочной коре зрительной системы приматов. Предложены новые методологические подходы для изучения функциональных характеристик нейронов и нейронных сетей, существенно расширяющие возможности экспериментального исследования. Работа объединяет классические нейрофизиологические методы и современные вычислительные подходы, преодолевая разрыв между анализом ответа отдельных нейронов и рассмотрением процесса зрительного распознавания в целом.

Разработан адаптивный метод генерации стимулов с использованием генеративно-состязательных сетей в сочетании с моделями нейронных ответов. Данный подход позволяет создавать изображения, вызывающие целенаправленную активацию нейронных ансамблей, и открывает новые возможности для исследования функциональной специализации нейронов без ограничения фиксированными наборами стимулов.

В ходе исследования выявлены оппонентные отношения между пространством описания сигнала и пространством постановки задачи, определяющие структуру и

динамику преобразования информации в нейронных сетях. Количественно оценена мера и степень влияния статистики входного сигнала на низкоуровневые этапы обработки, а также адаптации высокоуровневых этапов под цели наблюдателя. Установлен фазовый характер преобразования информации: от структуры, подчиняющейся статистике сигнала, через стадию трансформации с максимальной сложностью описания, до консолидации признаков в соответствии с задачей.

Исследование кодирования категорий показало, что вопреки ожиданиям о возрастающей компактности описания по мере продвижения по иерархической системе, структура представлений на промежуточных и высоких уровнях остается многомерной, что свидетельствует о сложности распределенного кодирования информации.

Применение метода предсказания нейронных ответов посредством фрагментов естественных изображений подтвердило, что функции нейронов нижневисочной коры могут быть успешно аппроксимированы комбинациями локальных ориентаций и цветов. Аналогичные закономерности обнаружены при применении этого подхода к искусственным нейронным сетям, что указывает на сходство в механизмах обработки зрительной информации.

Проведено исследование процесса распознавания образов в условиях неопределенности с применением современных диффузионных моделей. Установлено, что точность распознавания зашумленных изображений моделью зависит от меры накопления структурной информации, и подчиняется характерному скачкообразному переходу, наблюдаемому при пороговом восприятии в сенсорных системах, когда при преодолении определенного информационного уровня классификация трансформируется из случайной в практически однозначную.

Сопоставление результатов, полученных на ответах нейронов зрительной коры и искусственных нейронных сетях, раскрывает как фундаментальные сходства, так и существенные различия в принципах организации этих систем. Хотя сверточные нейронные сети демонстрируют высокую эффективность в моделировании зрительного восприятия, они имеют ряд функциональных ограничений. Тем не менее, подобные модели следует рассматривать как неотъемлемый и ценный инструмент нейрофизиологических исследований, существенно расширяющий возможности для моделирования, анализа и совершенствования экспериментальных парадигм.

Результаты работы имеют теоретическое значение для понимания механизмов кодирования зрительной информации и практическую ценность для повышения эффективности нейрофизиологических исследований.

Перспективным направлением дальнейших исследований представляется изучение динамических аспектов распознавания образов. Понимание того, как меняется представление объекта в зрительной системе по мере поступления новой информации, смене контекста или выполняемой задачи, позволит получить более полную картину механизмов распознавания объектов в динамических условиях реального мира.

ВЫВОДЫ

1. Активность нейронов нижневисочной коры головного мозга приматов (область TEad), усиливается при предъявлении изображений лица по сравнению с

изображениями других объектов. Средний вызванный ответ на лица был выше в 130 из 143 точек регистрации активности нейрональных колонок. В 118 точках регистрации возрастало стандартное при предъявлении тестов в виде изображений лиц (коэф. корр. = 0,58, $p \leq 0,001$).

2. Индекс избирательности к категории лиц был высок и составил $K = 0,06$ для изображений лиц людей и $K = 0,10$ для изображений лиц обезьян. Показано, что искусственные объекты и ландшафты демонстрируют низкий уровень избирательности по сравнению с другими категориями, что проявляется в средних значениях индекса $K = -0,10$ и $-0,14$ соответственно.
3. Анализ пространственной организации показал, что нейрональная активность в нижневисочной коре имеет высокую корреляция ответов, характерную для локально организованных ансамблей (коэф. корр. = 0,41, $p \leq 0,001$), которая снижается с увеличением расстояния между точками регистрации и имеет нелинейный характер и может быть описана как степенной ($R^2=0.25$) так логарифмической зависимостью ($R^2=0.24$).
4. Моделирование нейрональных ответов с помощью сверточных нейронных сетей продемонстрировало высокую точность предсказаний реакций нейрональных колонок нижневисочной коры (коэф. корр. = 0,68, $p < 0,001$). Также выявлено соответствие между глубокими слоями сети (block5_conv2) и ответами нейрональных колонок. Полученные результаты указывают на сходство в механизмах обработки информации биологическими и искусственными нейронными сетями, выполняющими задачу зрительного распознавания. При этом точность модели зависела от стабильности ответа нейрональных колонок (коэф. корр. = 0,7, $p < 0,01$), рассчитанного на основе реакции в четных и нечетных блоках стимуляции.
5. Предложен и разработан адаптивный метод генерации стимулов на основе анализа ответов нейронов нижневисочной коры, позволяющий создавать оптимизированные изображения, вызывающие целенаправленную активацию нейронных ансамблей. В отличие от традиционных подходов, подобный метод «генерации стимулов на основе обратной связи» дает возможность тестирования функциональной специализации нейронов без ограничения фиксированными наборами стимулов. Анализ двух подходов к созданию стимулов, показал различия в эффективности их применения: функция потерь Evoked (коэф. корр. = 0,42) обеспечивала максимальную активацию нейронов, в то время как Softmax (коэф. корр. = 0,31) оказалась в среднем менее эффективной, но продемонстрировала большую устойчивость к «утомлению», а точнее снижению общего уровня нейронального ответа нижневисочной коры во втором блоке эксперимента (коэф. корр. = 0,35 против -0,09 для Evoked).
6. Применение разработанного метода предсказания ответов нейронов посредством фрагментов естественных изображений показало, что ответ нейронов нижневисочной коры можно предсказать на основе набора признаков V1/V2, включающих локальные ориентации и цвета (коэф. корр. = 0,51 для всех стимулов; 0,37 для лиц). Применение этого подхода к искусственным нейросетям позволило выявить аналогичные закономерности в обработке зрительной информации, демонстрируя возможность

предсказания ответа глубоких слоев (VGG-16, слой 13) через комбинации локальных ориентаций и цветов (коэф. корр. = 0,44).

7. В модельных исследованиях установлены оппонентные отношения между пространством описания сигнала и пространством постановки задачи, а также их влияние на представление информации на разных этапах зрительной обработки. Начальные слои сохраняют структуру, близкую к статистике входного сигнала (коэф. корр. = 0,62), тогда как последние сверточные слои адаптируются под задачу (коэф. корр. = 0,34), хотя традиционно рассматриваются как этап выделения зрительных характеристик. В завершающих полносвязных слоях происходит консолидация представлений (коэф. корр. = 0,74), подчиняясь структуре задачи. Обнаружен фазовый характер обработки: сначала резкий разрыв связей с входными данными, затем этап трансформации представлений и распределенного кодирования, а на заключительном этапе – консолидация признаков под задачи распознавания.

8. Исследование устойчивости к шуму в задаче распознавания образа показало наличие скачкообразного перехода у искусственных нейронных сетей, аналогично достижению порогового уровня восприятия, например феномену «инсайта» в экспериментах с Голлин-тестом. После 60% итераций изображения начинают распознаваться с высокой вероятностью, а после 80% шагов классификация становится практически однозначной. Однако причины подобного перехода у искусственных нейронных сетей могут объясняться различными механизмами кодирования признаков о категории объекта.

СПИСОК ОСНОВНЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

1. Вахрамеева О. А., Хараузов А. К., Пронин С. В., Малахова Е. Ю., Шелепин Ю. Е. Зрительный прайминг при распознавании мелких изображений в сцене, содержащей объекты разного размера // Физиология человека. 2016. Т. 42, № 5. С. 39–48.
2. Малахова Е. Ю. Визуализация информации, кодируемой нейронами высших областей зрительной системы // Оптический журнал. 2018. № 8. С. 61–66.
3. Жукова О. В., Малахова Е. Ю., Шелепин Ю. Е. Джоконда и неопределенность распознавания улыбки человеком и искусственной нейронной сетью // Оптический журнал. 2019. № 11.
4. Малахова Е. Ю. Пространство описания зрительной сцены в искусственных и биологических нейронных сетях // Оптический журнал. 2020. № 10.
5. Малахова Е. Ю. Представление категорий посредством прототипов согласованной активности нейронов в сверточных нейронных сетях // Оптический журнал. 2021. № 12.
6. Nam Y., Sato T., Uchida G., Malakhova E., Ullman Sh., Tanifuji M. View-tuned and view-invariant face encoding in IT cortex is explained by selected natural image fragments // Scientific Reports (Nature Portfolio). 2021. № 11. С. 7827.
7. Малахова К. Ю., Шелепин К. Ю., Шелепин Ю. Е. Обнаружение и распознавание изображений в условиях помехи // Оптический журнал. 2024. Т. 91, № 8. С. 60–74.

Тезисы докладов:

8. Малахова Е. Ю., Жукова О. В., Шелепин Ю. Е. Джоконда – неопределенность восприятия улыбки человеком и искусственной нейронной сетью // Тез. докл.

- Международной научной конференции «Технологическая перспектива в рамках Евразийского пространства: новые рынки и точки экономического роста». Санкт-Петербург, 26–28 октября 2017. – С. 309–312.
9. Малахова Е. Ю. Методы визуализации репрезентации информации в искусственных и биологических нейронных сетях // Тез. докл. Международной научной конференции «Технологическая перспектива в рамках Евразийского пространства: новые рынки и точки экономического роста». Санкт-Петербург, 26–28 октября 2017. – С. 321–324.
 10. Малахова Е. Ю. Представление информации в нейронных сетях при распознавании семантики изображений // Тез. докл. XXIII съезда Физиологического общества им. И. П. Павлова. Воронеж, 18–22 сентября 2017. – С. 1642–1644.
 11. Жукова О. В., Малахова Е. Ю. Разработка искусственной нейронной сети глубокого обучения при распознавании лиц в условиях неопределенности // Тез. симпозиума, посвященного 100-летию Физиологического общества им. И. П. Павлова. Санкт-Петербург, 17–19 апреля 2017. – С. 37–38.
 12. Малахова Е. Ю. Моделирование и анализ сверточных нейронных сетей для задачи детекции текста на естественных изображениях // Тез. симпозиума, посвященного 100-летию Физиологического общества им. И. П. Павлова. Санкт-Петербург, 17–19 апреля 2017. – С. 63–64.
 13. Малахова Е. Ю. Формирование зрительных категорий в нейронных сетях // Материалы Всероссийской молодежной конференции с международным участием «Современные аспекты интегративной физиологии». Санкт-Петербург, 9–11 октября 2018. – С. 124.
 14. Малахова К. Representation of categories in filters of deep neural networks // In Proceedings: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. 2018. – С. 1973–1975.

Другие печатные издания:

15. Малахова Е. Ю. Обработка зрительной информации в искусственных и биологических нейронных сетях / Нейротехнологии // ред. Шелепин Ю. Е., Чихман В. Н., СПб.: Изд-во ВВМ, 2018. С. 336–346.
16. Zhukova O. V., Malakhova K. Yu., Shelepin Yu. E. Gioconda's Smile — from biological to artificial neural networks / Neural Networks and Neurotechnologies // ред. Шелепин Ю. Е., Огородникова Е. А., Соловьев Н. А., Якимов Е. Г., СПб.: Изд-во ВВМ, 2019. С. 204–210.