

ПОИСК ОПТИМАЛЬНЫХ ЗНАЧЕНИЙ ПАРАМЕТРОВ ПАКЕТА АНАЛИЗА ДАННЫХ КАЛЬЦИЕВОЙ ВИЗУАЛИЗАЦИИ MINIAN С ИСПОЛЬЗОВАНИЕМ МОДЕЛЬНЫХ ДАННЫХ

© 2023 г. А. И. Ерофеев^{1, *, †}, М. В. Петрушан^{2, †}, Л. В. Лысенко²,
Е. К. Винокуров¹, О. Л. Власова¹, И. Б. Безпрозванный^{1, 2, 3, **}

¹Лаборатория молекулярной нейродегенерации Санкт-Петербургского политехнического университета Петра Великого, Санкт-Петербург, Россия

²Лаборатория синаптической биологии, Научно-исследовательский технологический центр нейротехнологий Южного федерального университета, Ростов-на-Дону, Россия

³Отделение физиологии юго-западного медицинского центра Техасского университета, Даллас, США

*e-mail: alexander.erofeev@gmail.com

**e-mail: ilya.bezprozvanny@utsouthwestern.edu

Поступила в редакцию 05.06.2023 г.

После доработки 25.06.2023 г.

Принята к публикации 03.07.2023 г.

Визуализация кальция *in vivo* широко применяется в нейробиологии для оценки активности нейронных ансамблей. Появление однофотонного миниатюрного флуоресцентного микроскопа (минископа) сделало возможным проведение прижизненной визуализации кальция у свободно передвигающихся животных. Для анализа данных, полученных с помощью минископа, были разработаны различные алгоритмы и пакеты анализа. В данной работе на примере модельных данных с разным уровнем шума исследуется связь между точностью обнаружения нейронов и значениями параметров “Minian” — пакета для анализа данных, полученных с помощью минископа. На основании полученных результатов даются рекомендации по изменению значений параметров “Minian” в зависимости от уровня шума в обрабатываемых данных. Результаты, полученные в данном исследовании, являются предварительным руководством по выбору соответствующих значений параметров “Minian” при обработке экспериментальных данных. Ожидается, что результаты данного исследования будут актуальны для нейробиологов, занимающихся прижизненной визуализацией кальция у свободно передвигающихся животных.

Ключевые слова: минископ, флуоресценция, кальций, Minian, моделирование, выбор параметров

DOI: 10.31857/S0044467723050039, **EDN:** OBLOPP

ВВЕДЕНИЕ

Одной из основных задач в нейробиологии является понимание того, как нейроны головного мозга кодируют, модифицируют, хранят и извлекают информацию. Для решения этой задачи требуются инструменты и методы, способные прижизненно регистрировать нейронную активность. Одним из таких методов регистрации клеточной активности является кальциевая визуализация (Ca^{2+}), которая представляет собой оптиче-

ское измерение концентрации кальция с помощью специальных флуоресцентных индикаторов (Russell, 2011; Grienberger, Konnerth, 2012). Динамические колебания уровня кальция коррелируют с такими важными клеточными событиями, как генерация потенциала действия, экзоцитоз нейротрансмиттеров, изменения синаптической пластичности и транскрипция генов (Resendez, Stuber, 2015). Стоит отметить, что в нейробиологии визуализация кальция в нейронах и нейронных ансамблях особенно важна, поскольку кальциевые сигналы выполняют свои высокоспецифичные

[†] Эти авторы внесли равный вклад в данную работу и разделяют первое авторство.

функции в четко определенных клеточных компартментах (Wu et al., 2021).

Существует несколько методов прижизненной визуализации кальция, такие как однофотонная микроскопия (Iwasaki, Ikegaya, 2018), двухфотонная микроскопия (Denk et al., 1990; ter Veer et al., 2017; Zátönyi et al., 2020) и волоконная микроскопия (Pochechuev et al., 2018). Технология визуализации с помощью двухфотонной микроскопии широко используется в нейровизуализации благодаря таким преимуществам, как высокая скорость сканирования, высокое разрешение и глубина проникновения (Denk et al., 1990; ter Veer et al., 2017). Несмотря на преимущества двухфотонной микроскопии, у этого метода было существенное ограничение — во время эксперимента животное должно быть зафиксировано. Это ограничение было преодолено после создания легких и миниатюрных микроскопов для изучения поведения свободно передвигающихся животных. Например, были разработаны миниатюрные двухфотонные микроскопы (Engelbrecht et al., 2008; Zong et al., 2021; Vogt, 2022; Zong et al., 2022), однофотонные флуоресцентные миниатюрные микроскопы (минископы) (Ghosh et al., 2011; Liberti et al., 2017; Wu et al., 2021; Barbera et al., 2022). На текущий момент минископы получили наиболее широкое распространение по сравнению с миниатюрными двухфотонными микроскопами из-за трудностей, связанных с изготовлением и стоимостью последних (Aharoni et al., 2019).

Минископ представляет собой портативную систему визуализации, состоящую из камеры, оптического тракта и съемной градиентной (GRIN) линзы. Данные, полученные с помощью минископа, представляют собой видеофайл, содержащий покадровую запись изменений флуоресценции индикатора, чувствительного к кальцию (Ca^{2+}), такого как генетически кодируемый индикатор кальция GCaMP. Поскольку в данных однофотонной визуализации преобладает шумный, неравномерный и флуктуирующий фон, их обработка является достаточно сложной задачей (Lu et al., 2018b). Процесс обработки данных минископа включает в себя шумоподавление, коррекцию сдвигов изображения, идентификацию клеток и количественную оценку сигналов кальция (Robbins et al., 2021). Для реализации этих этапов обработки используются различные алгоритмы и пакеты анализа. Среди алгоритмов (подходов) можно выделить: метод главных компонент или

метод независимых компонент (PCA-ICA) (Mukamel et al., 2009), подход на основе выбора области интереса (ROI), ограниченную неотрицательную матричную факторизацию (CNMF) (Pnevmatikakis et al., 2016), расширенную ограниченную неотрицательную матричную факторизацию (CNMF-E) (Zhou et al., 2018) и другие (Pachitariu et al., 2017; Pnevmatikakis, Giovannucci, 2017; Bao et al., 2021; Li et al., 2022; Sità et al., 2022). Среди популярных пакетов анализа можно выделить: CalmAn — пакет анализа, написанный на языке Python, с открытым исходным кодом для обработки данных кальциевой визуализации, который включает алгоритмы CNMF и CNMF-E (Giovannucci et al., 2019); MIN1PIPE — пакет анализа, реализованный в Matlab, с открытым исходным кодом для обработки данных кальциевой визуализации, включающий алгоритмы CNMF и специальный алгоритм удаления фоновой флуоресценции во время предварительной обработки данных (Lu et al., 2018b); Minian — пакет анализа, написанный на языке Python, с открытым исходным кодом, интерактивным интерфейсом выбора параметров и проверки результатов обработки данных минископа (Dong et al., 2022), и другие (Radstake et al., 2019; Cantu et al., 2020; Erofeev et al., 2021; Friedrich et al., 2021; de Kraker et al., 2022). С полным списком инструментов анализа данных кальциевой визуализации можно ознакомиться по ссылке https://github.com/bahanonu/imaging_tools#imaging-tools.

При использовании подходов, основанных на PCA/ICA и CNMF-E, отмечаются проблемы с ложноотрицательными и ложноположительными результатами при идентификации области интереса (нейронов) (Lu et al., 2018b). Для различных подходов общей проблемой является сложность интерпретации извлеченного сигнала как артефакта или как биологически значимого сигнала. Кроме того, часть из упомянутых выше алгоритмов подразумевает сложную настройку различных параметров обработки исходных данных.

Настоящая работа посвящена проблеме выбора значений параметров в пакете анализа данных миниатюрной флуоресцентной микроскопии “Minian” (далее Minian). Данная проблема не была всесторонне изучена в предыдущих исследованиях (Robbins et al., 2021; Dong et al., 2022). Мы рассматриваем влияние параметров и их значений на точность обнаружения нейронов с помощью F -меры. F -мера

является средним гармоническим между точностью и полнотой детектирования нейронов. Она может быть определена через количество корректных и ложных детектирований нейронов, а также пропусков нейронов. F -мера часто используется как интегральная оценка эффективности детектирования нейронов (Lu et al., 2018). Подробно вычисление F -меры и способ определения корректных и ложных детектирований и пропусков нейронов описаны в разделе “Материалы и методы”, подразделе “определение оптимальных значений параметров Minian”, пункте 4. Мы ставим задачу определить такие значения параметров, при которых F -мера будет максимальной (оптимальные значения параметров). Кроме того, мы исследуем, как оптимальные значения параметров изменяются при различных уровнях шума в исходных данных. Под ними мы подразумеваем данные миниатюрной флуоресцентной микроскопии, которые могут быть как экспериментальными, так и модельными. В нашем исследовании в качестве исходных мы используем набор модельных данных. Модельные данные (табл. 1) часто используются в качестве эталонных для проверки алгоритмов и пакетов анализа (Zhou et al., 2018; Dong et al., 2022). В нашей работе мы демонстрируем целесообразность изменения значений параметров в пакете анализа Minian на модельных данных с разным уровнем шума.

МЕТОДИКА

Одной из основных целей данной работы было определение оптимальных значений параметров пакета анализа данных миниатюрной флуоресцентной микроскопии “Minian” (далее Minian). Набор параметров для оптимизации и диапазон их значений представлены в разделе 2.3 (табл. 2). Исследование проводилось на модельном наборе данных. Паттерны активности были синтезированы с использованием модели, описанной в разделе 2.1. Параметры модели генератора выбирались таким образом, чтобы пространственно-временные характеристики смоделированных паттернов активности соответствовали характеристикам, наблюдаемым в реальных данных. С этой целью мы использовали экспериментальный набор данных из репозитория Minian (https://github.com/denisecailab/minian/tree/master/demo_movies). Оптимальные параметры Minian определялись с помощью проверки соответствия меж-

ду пространственными характеристиками активности, полученными с помощью Minian, и известными параметрами генерации смоделированных паттернов.

Следующей целью данной работы было исследование взаимосвязи между оптимальными значениями параметров и уровнем шума исходных данных. Для этого в каждом кадре модельных данных мы добавляли случайный шум с равномерным распределением в заданном диапазоне. Оптимальность значения параметра определяли на основе F -меры. В качестве оптимального значения параметра выбирали такое, при котором значение F -меры было максимальным.

Создание модельных данных

Динамику кальциевой активности в момент времени t в пространственных координатах $\bar{x}$ на сериях изображений, моделирующих данные минискапа, определяли следующим уравнением:

$$f(t, \bar{x}) = \text{Relu}[F + B + UN(\text{ampl})], \quad (1)$$

где F — модель активности нейронов, представленная в виде суммы N_g пространственно-временных паттернов активности (компонентов):

$$F = \sum_{g=1}^{N_g} H(t - t_0) A_g T_g(t) SP_g(\bar{x}), \quad (2)$$

где N_g — количество компонентов активности (количество активных нейронов); g — индекс нейрона; t — время; $\bar{x} = \{x, y\}$ — вектор пространственных координат; A_g — амплитуда активности g -го нейрона; t_0 — время начала активности g -го нейрона (индекс g при t_0 в уравнении опущен); $T_g(t)$ — временная динамика активности g -го нейрона; H — функция Хевисайда, которая устанавливает активность g -го нейрона равной нулю до момента начала активности t_0 (индекс g при t_0 в уравнении опущен).

В нашем исследовании мы не моделировали временную динамику, а напрямую использовали динамику кальциевой активности, полученную при обработке экспериментальных данных в пакете Minian со значениями параметров по умолчанию. Пространственное распределение активности g -го нейрона $SP_g(\bar{x})$ моделировали следующим образом:

Таблица 1. Список наборов данных кальциевой визуализации для миниатюрной флуоресцентной однофотонной микроскопии
Table 1. List of datasets for single-photon calcium imaging

Название статьи	Год	Набор данных		Примечание
		модельные	экспериментальные	
Semantic Segmentation of Spatiotemporal Neural Signals Extraction of Miniscope Imaging Data https://arxiv.org/abs/1704.00793	2017	Набор данных с разным уровнем шума	Набор данных коры головного мозга мышей	Смоделировано 100 нейронов в виде сферической двухмерной Гауссовой функции со средним диаметром нейрона, сравнимым с реальными данными. Фон был извлечен из реальных видеозаписей с микроскопом для однофотонной идентификации. Отношение амплитуды между сигналом и фоном (уровень сигнала) варьировалось от 0.05 до 0.8 с шагом 0.05
MINIPIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline	2018	Один демонстрационный файл модельных данных в формате <i>.mp4</i> с разным уровнем сигнала https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6084484/bin/NIHMS978786-supplement-2.mp4 16 видеороликов, имитирующих динамику кальция в нейронах в формате <i>.mp4</i>	Демонстрационный файл (<i>demo_minipipe.m</i>) с данными, полученными с помощью микроскопа UCLA в формате <i>.m</i> https://github.com/JinghaoLu/MINIPIPE#dataset Набор данных коры головного мозга мышей и зоны Х зебрафиш	16 видеоизображений были синтезированы в виде серии 5-минутных видеороликов с частотой кадров в секунду, равной 10, содержащих 100 нейронов, имеющих характеристики, сопоставимые с экспериментальными данными. Нейроны были смоделированы с помощью сферической двухмерной Гауссовой функции. Фон извлекался из реальных видеозаписей однофотонной миниатюрной флуоресцентной микроскопии. Отношение амплитуды между сигналами и фоном (уровень сигнала) варьировалось от 0.05 до 0.8 с шагом 0.05
Efficient and accurate extraction of <i>in vivo</i> calcium signals from microendoscopic video data	2018	Набор данных в формате <i>.mat</i> https://caiman.flatironinstitute.org/~neuro/caiman_downloads/sim.mat	Набор данных стриатума, префронтальной коры, гиппокампа, ядра ложа терминальной полоски (BNST) в формате <i>.mat</i> https://datadryad.ori.rrd.org/dataset/doi:10.5061/dryad.kr17k	Нейроны моделировались с помощью сферической двухмерной Гауссовой функции со средним диаметром, сопоставимым с реальными данными. Данные моделировались с разным коэффициентом соотношения сигнал-шум (SNR) и количеством клеток. Используемые значения SNR: 1, 2, 3, 4, 5, 6. Количество клеток: 2, 50, 200 и количество кадров 1000 и 2000

Таблица 1. Окончание

Название статьи	Год	Набор данных		Примечание
		модельные	экспериментальные	
CaImAn an open-source tool for scalable calcium imaging data analysis	2019	Набор данных в формате <i>.mat</i> <i>test_sim.mat</i> https://zenodo.org/record/1659149 <i>sim.mat</i> https://caiman.flatironinstitute.org/~neuro/caiman_downloadables/sim.mat	Аннотированные вручную наборы данных из различных областей мозга (гиппокамп, зрительная кора, теменная кора) с использованием нескольких вариантов GCaMP https://github.com/CodeNeuro/neurofinder Демо-данные в формате <i>.npz</i> https://github.com/flatironinstitute/Calman/blob/master/testdata/groundtruth.npz	Набор данных состоит из 9 экспериментальных наборов двухфотонных данных мыши, полученных <i>in vivo</i> . Каждый набор данных проверен вручную 3–4 валидаторами. На последующем этапе результаты валидации были объединены для создания “консенсусного” набора данных. Консенсусный набор данных принимался за “истину”. https://zenodo.org/record/1659149#XC_Wcs9Ki9s https://users.flatironinstitute.org/~neuro/caiman_paper/
Minian, an open-source miniscope analysis pipeline	2022	Набор данных в формате <i>.nc</i> с разным соотношением сигнал-шум. https://cdn.elifesciences.org/articles/70661/elifesciences-fig16-data2-v2.zip	Набор данных гиппокампа CA1 мышей в формате <i>.avi</i> https://archive.softwareheritage.org/browse/directory/e11e29a1ec322ff836c3a0ca0abc3c2b0303fb81/?origin_url=https://github.com/-denisecailab/min-ian&path=demo_mov-ies&revision=a6d3339932df3c63aa46811fc70fc00aca09218d&snapshot=695d1c8410ede7ac197b94946b8f59eb122cbbb8	Нейроны моделировались с помощью сферической двумерной Гауссовой функции со средним диаметром, сопоставимым с реальными данными. Данные смоделированы с разным соотношением сигнал-шум (SNR) и количеством клеток. Используемые SNR: 0.2, 0.4, 0.6, 1.0, 1.4 и 1.8. Количество клеток: 100, 300 и 500, количество кадров: от 4000 до 28 000 с шагом 8000

.nc — независимый двоичный формат с открытым исходным кодом, используемый инструментами NetCDF под управлением Unidata.

.mat — формат контейнера двоичных данных MATLAB.

.npz — архивный формат для сохранения и загрузки массивов данных в библиотеке NumPy в языке программирования Python.

Таблица 2. Перечень использованных параметров Minian
Table 2. List of Minian analysis pipeline parameters

Название параметра	Значение по умолчанию	Диапазон значений	Краткое описание
'ksize'	7	3, 5, 9, 11	Параметр, управляющий медианной фильтрацией
'noise_freq'	0.06	0.005, 0.01, 0.02, 0.1, 0.2, 0.3, 0.45, 0.6, 0.8	Параметр, управляющий частотой среза
spatial 'sparse_penal'	0.01	0.0025, 0.0050, 0.0075, 0.05, 0.075, 0.1	Параметр, управляющий общей разреженностью пространственных следов (отпечатков)
temporal 'sparse_penal'	1	0.25, 0.50, 0.75, 1.25, 1.5, 1.75	Параметр, управляющий общей разреженностью кальциевой динамики и деконволюцией пиков
'wnd'	15	5, 7, 9, 11, 13, 17, 19	Параметр, управляющий морфологической операцией “tophat” для оценки фона 'wnd' — радиус элемента диска

Примечание:

1. Медианная фильтрация — это метод обработки сигналов или изображений, который использует медиану значений пикселей вокруг определенного пикселя для устранения выбросов или шума. Вместо вычисления среднего значения пикселей в окне фильтра медианная фильтрация заменяет значение пикселя медианой из соседних пикселей. Это помогает сгладить изображение и устранить аномалии, сохраняя при этом резкие границы и детали.

2. Частота среза (или граничная частота) — это частота, при которой амплитуда или энергия сигнала, проходящего через фильтр или систему, снижается до определенного уровня. В контексте сигнальной обработки частота среза определяет границу между пропуском сигнала и его подавлением фильтром. Например, в случае фильтра низких частот, частота среза указывает на частоту, ниже которой сигнал проходит через фильтр с минимальными изменениями, а выше которой сигнал подавляется или ослабляется.

3. Пространственный след (или пространственный отпечаток) относится к представлению пространственной информации в сигнале. Общая разреженность пространственных следов (отпечатков) относится к степени разреженности или распределению следов (отпечатков) в пространстве. Она характеризует, насколько равномерно или разреженно распределены следы или отпечатки на поверхности или в пространстве, а также сколько информации содержится в каждом следе или отпечатке. Высокая общая разреженность означает, что следы или отпечатки распределены с большим пространственным интервалом или имеют меньшую плотность, в то время как низкая общая разреженность указывает на более плотное распределение следов или отпечатков.

4. Разреженность в данном контексте относится к свойству результирующего кальциевого сигнала или динамики, где большинство значений или пиков имеют нулевые или очень низкие значения, в то время как некоторые из них имеют более высокие значения. Такая разреженность может быть результатом физиологических особенностей клеток или биологических процессов. В анализе кальциевых сигналов разреженность может быть использована для выделения и анализа релевантных событий или активаций.

5. Морфологическая операция “tophat” (преобразование в виде цилиндра) — это операция, используемая в математической морфологии и обработке цифровых изображений, для извлечения мелких элементов и деталей из заданных изображений.

1. Median filtering — a signal or image processing technique employed to eliminate outliers or noise, utilizing the median of pixel values surrounding a specific pixel. Rather than computing the average pixel value within the filter window, median filtering substitutes the pixel value with the median of adjacent pixels. This method aids in the enhancement of image smoothness and the elimination of irregularities, all while preserving distinct boundaries and intricate details.

2. The cutoff frequency represents the frequency at which the amplitude or energy of a signal, as it traverses a filter or system, is attenuated to a specific level. Within the realm of signal processing, the cutoff frequency establishes the demarcation between permitting the signal to pass through or impeding its progression by the filter. In the instance of a low-pass filter, for instance, the cutoff frequency signifies the frequency below which the signal undergoes minimal alteration while traversing the filter, above which the signal is suppressed or attenuated.

3. Spatial trace (or spatial footprint) denotes the depiction of spatial information within a signal. The collective spatial sparsity of spatial traces (or footprints) pertains to the extent of sparsity or distribution observed among the traces (or footprints) in space. It characterizes the uniformity or sparsity of the trace or print distribution across a surface or within a spatial context, as well as the amount of information encapsulated within each individual trace or print. A heightened overall sparsity indicates a wider spatial interval or reduced density in the distribution of traces or prints, whereas a diminished overall sparsity implies a denser allocation of traces or prints.

4. Sparsity — the resultant calcium signal or dynamics, whereby a majority of values or peaks possess zero or significantly low magnitudes, while a subset demonstrates elevated values. This sparsity phenomenon can arise from intrinsic cellular properties or biological processes. Within the domain of calcium signal analysis, leveraging sparsity can facilitate the isolation and examination of pertinent events or activations.

5. The morphological operation known as “tophat” (also referred to as cylinder transformation) is a technique employed within the domains of mathematical morphology and digital image processing. Its purpose is to extract small elements and intricate details from given images.

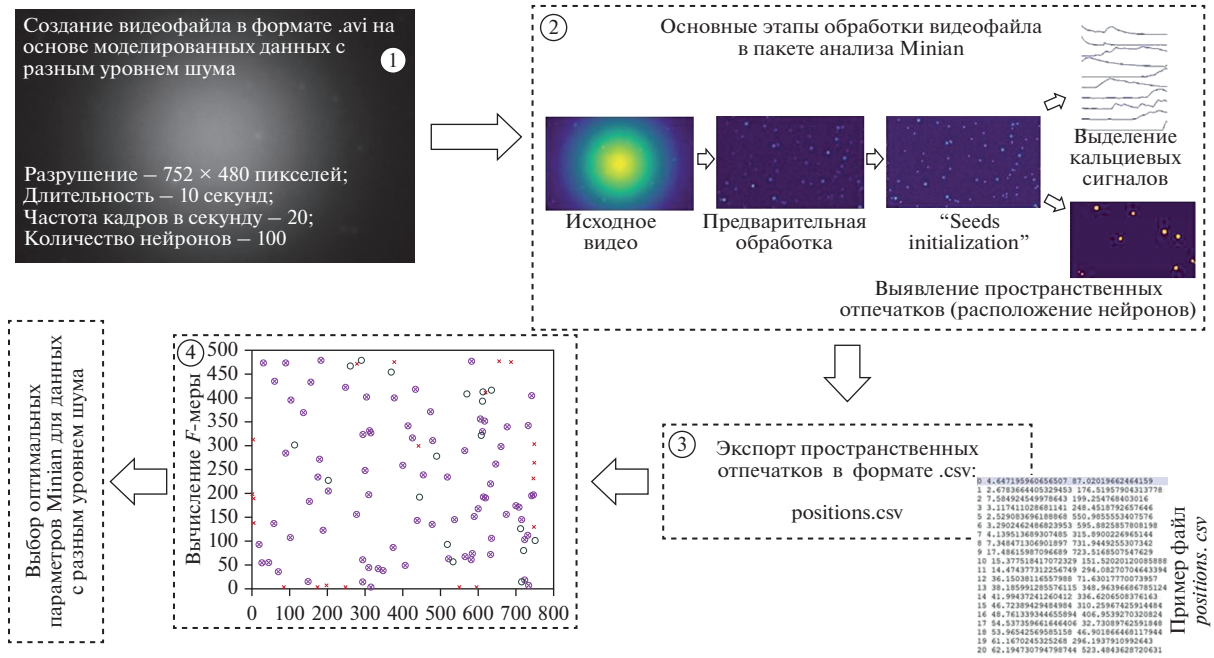


Рис. 1. Этапы, необходимые для поиска оптимальных параметров Minian: (1) создание видеофайла в формате AVI с модельными данными, (2) обработка видеофайла в пакете анализа Minian с разными значениями параметров, (3) экспорт данных из Minian в формате CSV и (4) вычисление F -меры.

Fig. 1. Pipeline for optimizing Minian parameters involves four steps: (1) generating a simulated video file in AVI format, (2) processing it in the Minian analysis pipeline with different parameters, (3) exporting the spatial footprints and calcium traces data from Minian in CSV format, (4) calculating the F1 score.

$$SP_g(\bar{x}) = e^{\frac{-(\bar{x} - \bar{m}_g)^T S_g^{-1} (\bar{x} - \bar{m}_g)}{2}}, \quad (3)$$

где $\bar{m}_g$ — координаты центра активности g -го нейрона; S_g^{-1} — обратная ковариационная матрица пространственного распределения активности g -го нейрона; B — модель фоновой активности, представленная в виде суммы N_b пространственно-временных фоновых паттернов (компонентов):

$$B = \sum_{b=1}^{N_b} K_b \sin(w_b t + \varphi_b) e^{\frac{-(\bar{x} - \bar{m}_b)^T S_b^{-1} (\bar{x} - \bar{m}_b)}{2}}, \quad (4)$$

где b — индекс фоновой составляющей; N_b — количество компонентов фоновой активности; K_b — амплитуда b -й составляющей фона; w_b — частота b -го компонента фона; φ_b — фаза b -го фонового компонента; $\bar{m}_b$ — пространственный центр b -го компонента фона; S_b^{-1} — обратная ковариационная матрица пространственного распределения b -го компонента фона; $UN(ampl)$ представляет собой аддитивный равномерный шум, выбранный

из диапазона $[-ampl, +ampl]$, функция $Relu$ определена как $Relu(x) = \max(0, x)$.

Определение оптимальных значений параметров Minian

Оптимальные параметры пакета анализа данных минискапа Minian были определены нами следующим образом (рис. 1):

1. Выбрали набор данных, полученных с помощью минискапа, для которых требовалось подобрать оптимальные значения параметров Minian.

2. Настроили параметры модели генератора, как описано в разделе 2.1.

Динамику кальциевой активности, полученную при обработке в Minian со значениями параметров по умолчанию, мы использовали как временную составляющую модели (2).

Пространственные компоненты включали центр области активности и ковариационную матрицу, характеризующую пространственное распределение амплитуды активности. Диагональные элементы ковариационной матрицы определяли размер моделируемой области активности эллиптической формы

вдоль горизонтальной и вертикальной осей. Недиagonальные элементы определяли наклон эллипса области активности.

В данном исследовании мы не ставили целью обеспечить реалистичное моделирование фона, так как это требует отдельного исследования, затрагивающего такие факторы, как яркостная неоднородность, вызванная оптическими свойствами минискапа, а также кальциевой динамикой вне фокальной плоскости. Мы использовали один низкочастотный фоновый компонент, чтобы увеличить визуальное сходство между модельными и экспериментальными данными.

Мы моделировали шум как случайный аддитивный компонент, применяемый независимо к каждому пикселю в каждом изображении. Значение шума выбирали из равномерного распределения в пределах ограниченного диапазона, который определялся относительно уровня сигнала. Уровень сигнала рассчитывали путем усреднения пиковой амплитуды сигнала всех нейронов за весь моделируемый период времени. Диапазон шума далее мы будем называть уровнем шума.

3. Синтезировали серию модельных данных с разным уровнем шума.

Например, диапазон шума может быть выбран равным 0.5 уровня сигнала (низкий уровень шума), равным уровню сигнала (средний уровень шума) или 1.5 уровня сигнала (высокий уровень шума).

4. Оценивали точность определения координат нейронов с помощью Minian, сравнив их с известными координатами модельного распределения.

Для оценки точности обнаружения активных областей использовали F -меру:

$$F = \frac{2tp}{2tp + fp + fn}, \quad (5)$$

где tp — количество истинно положительных, fp — количество ложноположительных, а fn — количество ложноотрицательных результатов идентификации активных областей (нейронов).

Выявление активной области (нейрона) с помощью Minian определяли как истинно положительное, если центр обнаруженной активной области находился в радиусе R от истинного положения моделируемого нейрона. Значение R определяли как максимум диагональных элементов ковариационной

матрицы пространственного распределения моделируемой активности. Количество ложноотрицательных обнаружений определяли на основе количества модельных нейронов в окрестности радиуса R , которые не были обнаружены с помощью Minian. Нейроны идентифицировали как ложноположительные, если в окрестности радиуса R не было соответствующего модельного нейрона. F -мера отображала степень соответствия истинного (модельного) пространственного распределения нейронов и распределения нейронов, обнаруженных с помощью Minian.

5. Выполняли шаг “4” на модельных данных, с разным уровнем шума и значениями параметров Minian.

Этот шаг позволял оценить зависимость точности обнаружения нейронов от уровня шума, а также возможность настройки параметров Minian для различных уровней шума.

6. Оценивали уровень шума на экспериментальных данных и использовали параметры Minian, определенные как оптимальные на модельных данных (шаг “5”) с ближайшим значением относительного уровня шума.

$$NoiseStrength = Average_{x,y,t} (abs(I_{x,y,t+1} - I_{x,y,t})), \quad (6)$$

$$\{x,y\} \in BMask,$$

где $Average_{x,y,t}$ означает усреднение по пространственному (x,y) и временному (t) измерениям. Пространственное усреднение выполняли только в пределах области фона (фоновой маски). $I_{x,y,t+1}$ и $I_{x,y,t}$ — значения интенсивности оттенков серого на видеоизображении кальциевой динамики в координатах (x,y) и в кадрах $t+1$ и t соответственно; $BMask$ — маска фона, где $BMask(x,y) = 1$, если (x,y) находится в локальной окрестности радиуса R любого нейрона, и $BMask(x,y) = 0$ в противном случае. Величину R выбирали как радиус ограничивающей окружности нейрона типичного размера. Относительный уровень шума вычисляли путем деления уровня шума на уровень сигнала, который рассчитывался, как описано в шаге “2”. Программа C++ для вычисления относительного уровня шума представлена в дополнительных материалах.

Параметры пакета анализа “Minian”

В данной работе мы рассмотрели вопрос выбора оптимальных значений определенно-го набора параметров в Minian. Описание и

диапазон значений каждого параметра приведены в табл. 2. Мы выбрали именно эти параметры для оптимизации, поскольку они легко настраиваемы с помощью встроенных инструментов визуализации, доступных в Minian. Оптимизация параметров проводилась независимо друг от друга. При обработке модельных данных в Minian мы изменяли только один рассматриваемый параметр, оставляя остальные параметры со значениями по умолчанию. Значения параметров по умолчанию указаны в файле `pipeline.ipynb`, расположенном в репозитории Minian. Модельные данные были сгенерированы с использованием программы, написанной на языке программирования C++, реализующей модель, подробно описанную в разделе 2.1. Вычисления и визуализация модельных данных были проведены с применением библиотеки OpenCV (версия 3.X).

Используемое оборудование

Анализ параметров Minian проводили на персональном компьютере, оснащенном процессором Intel Xeon E3-1246 v3 с частотой 3.5 ГГц, 32 ГБ оперативной памяти DDR3 и операционной системой Windows 10 LTSC (версия 10.0.17763 Build 17763). В качестве программных инструментов использовали python 3.8.8, conda 4.12.0, Minian версии 1.2.1.

Статистический анализ

Для всех рассматриваемых значений параметров в Minian были обработаны 3 набора модельных данных, каждый из которых включал 10 видеофайлов. Каждый набор данных соответствовал определенному уровню шума: низкий, средний и высокий. Результаты, представленные на графиках, представлены в виде среднего значения $\pm$ стандартная ошибка среднего. Для оценки статистической значимости использовался тест Манна–Уитни. Сравнение осуществлялось внутри одного набора данных. На графиках звездочками указаны достоверные изменения F -меры относительно значения F -меры при значении параметра по умолчанию для групп данных с низким (зеленый цвет), средним (синий цвет) и высоким (красный цвет) уровнем шума соответственно.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЙ

Модельные данные с разным уровнем шума

Модельный набор данных синтезировали с тремя разными уровнями шума, следуя процедуре, описанной в разделе 2.2. Пространственно-временные характеристики активности в модельных данных соответствовали характеристикам экспериментального набора данных (https://github.com/denisecailab/minian/tree/master/demo_movies). При генерации модельных данных использовали следующие параметры модели: пространственное разрешение 752×480 пикселей, продолжительность 10 секунд, частота кадров — 20 кадров в секунду, 100 нейронов.

Координаты нейронов выбирали случайным образом из равномерного распределения с диапазонами: [0, ширина изображения), [0, высота изображения) по горизонтальной и вертикальной осям. В качестве временных компонентов модельной активности использовали динамику активности, полученную в результате обработки экспериментальных данных (https://github.com/denisecailab/minian/tree/master/demo_movies) пакетом анализа Minian с использованием настроек по умолчанию. Значение амплитуды A_g устанавливали равным 3 для каждого g -го нейрона, для обеспечения визуального сходства между модельными и экспериментальными данными. Диагональные компоненты ковариационной матрицы (уравнение (4)) выбирали случайным образом из равномерного распределения со значениями в диапазоне от 9 до 14. Такой диапазон был выбран в целях визуального соответствия размеров модельных областей активности и размеров областей активности, наблюдаемых в экспериментальных данных. Недиagonальные элементы ковариационной матрицы задавали случайным образом в диапазоне $[0, 0.25 * \min(\text{диагональные элементы})]$ для обеспечения визуального разнообразия форм областей активности. Фон в уравнении 4 моделировали одним компонентом со следующими параметрами: $K_b = 75$; $w_b = 0.001$; $\phi_b = \frac{\pi}{2}$. Диагональные компоненты ковариационной матрицы устанавливали равными 20000, а недиагональные компоненты ковариационной матрицы — равными 0. Вектор $\bar{m}_b$ представлял собой координаты центра изображения. Используемая модель фона позволяла имитировать изменения яркости от центра к периферии изображения,



Рис. 2. Примеры изображений модельного набора данных с разным уровнем шума: низким (0.09 ± 0.003), средним (0.27 ± 0.035) и высоким (0.76 ± 0.067).

Fig. 2. Simulated calcium activity images with different levels of noise: low (0.09 ± 0.003), medium (0.27 ± 0.035) and high (0.76 ± 0.067).

аналогично яркостной неоднородности, наблюдаемой в экспериментальных данных.

Далее было выполнено улучшение визуального сходства модельных данных с экспериментальными данными. Этот дополнительный этап обработки включал два Гауссовых размытия. Первое размытие реализовывали с размером ядра свертки 5×5 и стандартным отклонением 5 по каждой оси. Это размытие было применено для подавления высокой резкости синтезированных изображений, которая обычно отсутствует в реальных данных. Второе Гауссово размытие реализовывали с использованием ядра свертки размером 21×21 и стандартным отклонением 21 по каждой оси. Оно имитировало эффект свечения вокруг активных нейронов, наблюдаемый в экспериментальных данных. Результирующее изображение в градациях серого вычисляли как сумму размытых изображений.

Для генерации трех групп модельных данных с разными уровнями шума использовали программу, написанную на языке программирования C++. Эта программа реализовывала модель, описанную в разделе 2.1. Уровень шума устанавливали равным 0.5 от уровня сигнала для группы “малый шум”, равным уровню сигнала для группы “средний шум” и 1.5 для группы “высокий шум” (раздел 2.2 “шаг 2”). Каждая группа состояла из 10 смоделированных образцов с различной динамикой активности, извлеченной из набора экспериментальных данных (https://github.com/denisecailab/minian/tree/master/demo_movies) с помощью пакета анализа Minian с использованием настроек по умолчанию и различным пространственным распределением центров нейронов. Примеры отдельных кадров синтезированных

образцов с разным уровнем шума показаны на рис. 2. Модельные данные имели такое же временное и пространственное разрешение, продолжительность и количество нейронов, что и экспериментальные данные из репозитория Minian. Увеличение относительной силы шума (на рис. 2 представлено слева направо) приводило к появлению визуальной зернистости. Для оценки относительной силы шума уровень сигнала рассчитывали из динамики активности, полученной с помощью Minian с настройками по умолчанию, после чего рассчитывали относительную силу шума в соответствии с процедурой, описанной в разделе 2.2 (шаг 6). Средние значения силы шума составляли 0.09 ± 0.003 , 0.27 ± 0.035 и 0.76 ± 0.067 для групп с низким, средним и высоким уровнем шума соответственно. Для сравнения — относительная сила шума в экспериментальных данных из репозитория Minian составляла 0.68. Выбор трех разных уровней шума, включая близкий к наблюдаемому в экспериментальных данных, позволял анализировать закономерности изменения оптимальных параметров Minian при изменении уровня шума.

Оптимальные значения параметров Minian

В этом разделе мы рассмотрим оптимальные значения параметров пакета анализа Minian. Под оптимальными значениями параметров мы понимаем те, при которых достигается максимальное значение F -меры. Для этого мы провели серию экспериментов с модельными данными разного уровня шума и различными параметрами Minian (табл. 2): *'ksize'* — параметр, управляющий медианной фильтрацией; *'noise_freq'* — параметр, управ-

ляющий частотой среза, т.е. границей между пропусканием сигнала и его подавлением фильтром; *spatial 'sparse_penal'* — параметр, управляющий общей разреженностью пространственных следов (отпечатков); *temporal 'sparse_penal'* — параметр, управляющий общей разреженностью кальциевой динамики и деконволюцией пиков; *wnd'* — параметр, управляющий морфологической операцией “tophat” для оценки фона. При использовании стандартных значений параметров Minian *F*-мера составляла 0.90 ± 0.01 при низком уровне шума, 0.77 ± 0.04 при среднем и 0.53 ± 0.03 при высоком уровне шума соответственно.

На первом этапе мы провели обработку модельных данных, изменяя только параметр *'ksize'* (рис. 3 (а)). Как показано на рис. 3 (а), *F*-мера достоверно не изменялась при изменении параметра *'ksize'* относительно значения по умолчанию для набора данных с низким уровнем шума. Однако для данных со средним и высоким уровнем шума значение *F*-меры значительно снижалось при уменьшении значения параметра *'ksize'* и возрастало при увеличении значения параметра относительно значения по умолчанию. Примечательно, что для данных с высоким уровнем шума *F*-мера значительно увеличилась при значениях параметра *'ksize'*: 9 (0.71 ± 0.02) и 11 (0.77 ± 0.01) относительно значения *F*-меры при использовании параметра *'ksize'* со значением по умолчанию, равным 7.

После оценки параметра *'ksize'* мы также проанализировали *F*-меру для параметров *'noise_freq'*, *spatial 'sparse_penal'*, *temporal 'sparse_penal'* и *'wnd'* (рис. 3 (б), 4–6). Мы провели парное сравнение между значением *F*-меры при использовании выбранного параметра по умолчанию и значением *F*-меры при использовании этого же параметра, но с другим значением внутри каждой группы данных. При низком уровне шума не наблюдалось зна-

чимых изменений *F*-меры при увеличении или уменьшении значений параметров *spatial 'sparse_penal'*, *temporal 'sparse_penal'* и *'wnd'* относительно их значений по умолчанию. Однако при значении параметра *'noise_freq'*, равном 0.01, было обнаружено значительное снижение значения *F*-меры, до 0.84 ± 0.01 .

Для данных со средним уровнем шума было обнаружено значительное увеличение *F*-меры при следующих значениях параметра *'noise_freq'*: 0.2 (0.91 ± 0.01), 0.3 (0.90 ± 0.01) и 0.5 (0.89 ± 0.01) (рис. 3 (б)). Аналогично *F*-мера существенно возрастала при значениях параметров *spatial 'sparse_penal'*: 0.05 (0.91 ± 0.01) и 0.075 (0.89 ± 0.01) и *temporal 'sparse_penal'*: 1.5 (0.89 ± 0.01) и 1.75 (0.90 ± 0.01) (рис. 4). При изменении параметра *'wnd'* значение *F*-меры достоверно не изменялось.

Для данных с высоким уровнем шума изменение значений параметра *'noise_freq'* приводило к значимым изменениям *F*-меры: 0.01 (0.77 ± 0.01), 0.02 (0.74 ± 0.02), 0.1 (0.74 ± 0.02), 0.2 (0.89 ± 0.01), 0.3 (0.86 ± 0.01), 0.45 (0.83 ± 0.01), 0.6 (0.83 ± 0.02) и 0.8 (0.83 ± 0.02). Более того, увеличение значений параметра *spatial 'sparse_penal'* приводило к увеличению *F*-меры: 0.05 (0.89 ± 0.01), 0.075 (0.1 ± 0.01) и 0.1 (0.1 ± 0.01) (рис. 4 (а)). Аналогичные результаты были получены для параметра *temporal 'sparse_penal'* (рис. 4 (б)): 1.25 (0.68 ± 0.02), 1.5 (0.84 ± 0.01) и 1.75 (0.86 ± 0.02). Также значимые изменения *F*-меры: 5 (0.78 ± 0.03), 7 (0.77 ± 0.02), 9 (0.73 ± 0.03) и 11 (0.65 ± 0.01) — наблюдались при уменьшении значений параметра *'wnd'* (рис. 5).

Таким образом, можно сделать вывод, что для данных с низким уровнем шума значения по умолчанию параметров, таких как *'ksize'*, *'noise_freq'*, *spatial 'sparse_penal'*, *temporal 'sparse_penal'* и *'wnd'* являются оптимальными. Однако для данных со средним и высоким уровнем шума эти параметры требуется изменять как в сторону увеличения, так и в сторо-

Таблица 3. Рекомендации по выбору параметров Minian в зависимости от уровня шума исходных данных
Table 3. Minian parameter selection guide

Название параметра	Значение по умолчанию	Уровень шума		
		низкий	средний	высокий
<i>'ksize'</i>	7	По умолчанию	По умолчанию	11 (↑)
<i>'noise_freq'</i>	0.06	По умолчанию	0.2 (↑)	0.2 (↑)
<i>spatial 'sparse_penal'</i>	0.01	По умолчанию	0.05 (↑)	0.05 (↑)
<i>temporal 'sparse_penal'</i>	1	По умолчанию	1.75 (↑)	1.75 (↑)
<i>'wnd'</i>	15	По умолчанию	По умолчанию	5 (↓)

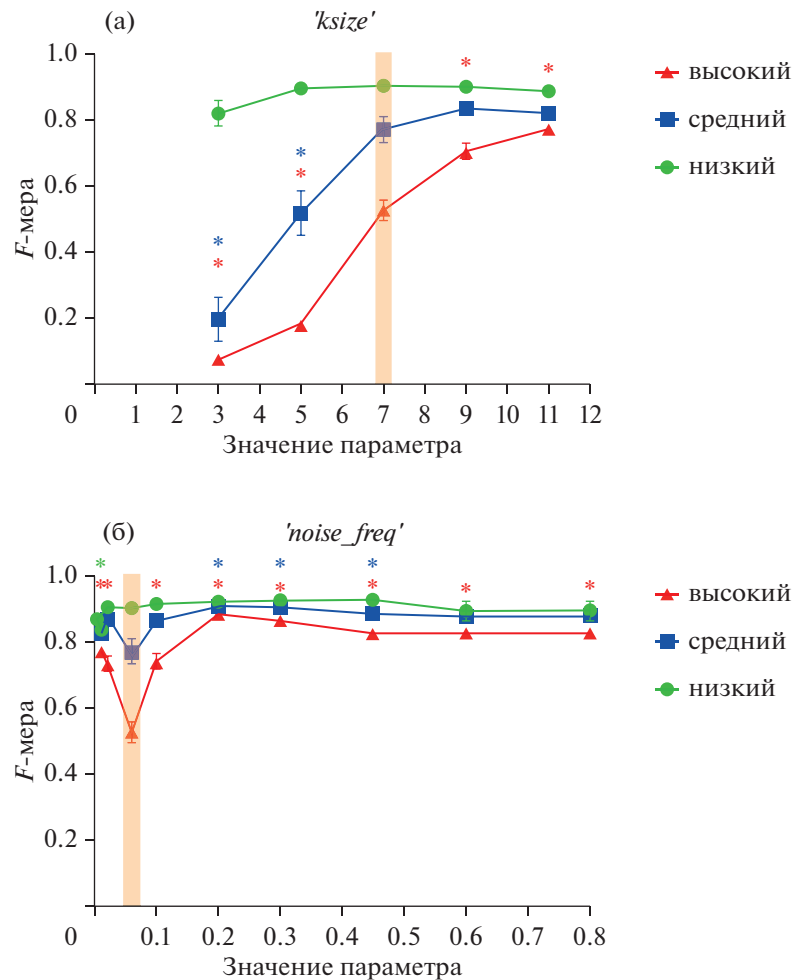


Рис. 3. График изменения значения F -меры в зависимости от значения параметра: (а) — ' $ksize$ ': 3, 5, 9, 11; (б) — ' $noise_freq$ ': 0.005, 0.01, 0.02, 0.06, 0.1, 0.2, 0.3, 0.45, 0.6, 0.8; для модельного набора данных с разным уровнем шума (низкий, средний, высокий). Оранжевым прямоугольником отображен диапазон значений F -меры для значения параметра по умолчанию при разных уровнях шума. Данные представлены как среднее $\pm$ стандартная ошибка среднего. Зелеными, синими и красными "*" обозначены статистически значимые изменения F -меры относительно значения F -меры при значении параметра по умолчанию для групп данных с низким, средним и высоким уровнем шума соответственно.

Fig. 3. The $F1$ score is plotted as a function of parameter values: (a) — ' $ksize$ ': 3, 5, 9, 11; (б) — ' $noise_freq$ ': 0.005, 0.01, 0.02, 0.06, 0.1, 0.2, 0.3, 0.45, 0.6, 0.8; for simulated datasets with different noise levels (low, medium, high) processed through Minian. The orange rectangle on the graph indicates the range of $F1$ score at the default value of Minian parameters. The data are presented as mean $\pm$ SEM. Green, blue and red "*" symbols indicate significant changes in the $F1$ score compared to the default $F1$ score value for the data groups characterized by low, medium and high noise levels, respectively.

ну уменьшения. На основании полученных результатов мы можем рекомендовать оптимальные значения параметров Minian в зависимости от уровня шума (низкий, средний и высокий) для модельного набора данных (табл. 3). Результаты данного исследования могут служить предварительным руководством при выборе соответствующих значений параметров для обработки экспериментальных данных с использованием пакета анализа Minian.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В данном исследовании мы сосредоточились на проблеме выбора оптимальных параметров в пакете анализа Minian. Мы исследовали, как изменение значений параметров Minian влияет на точность обнаружения нейронов в зависимости от уровня шума в исходных данных.

Как и ожидалось, при высоком уровне шума исходных данных значение параметра

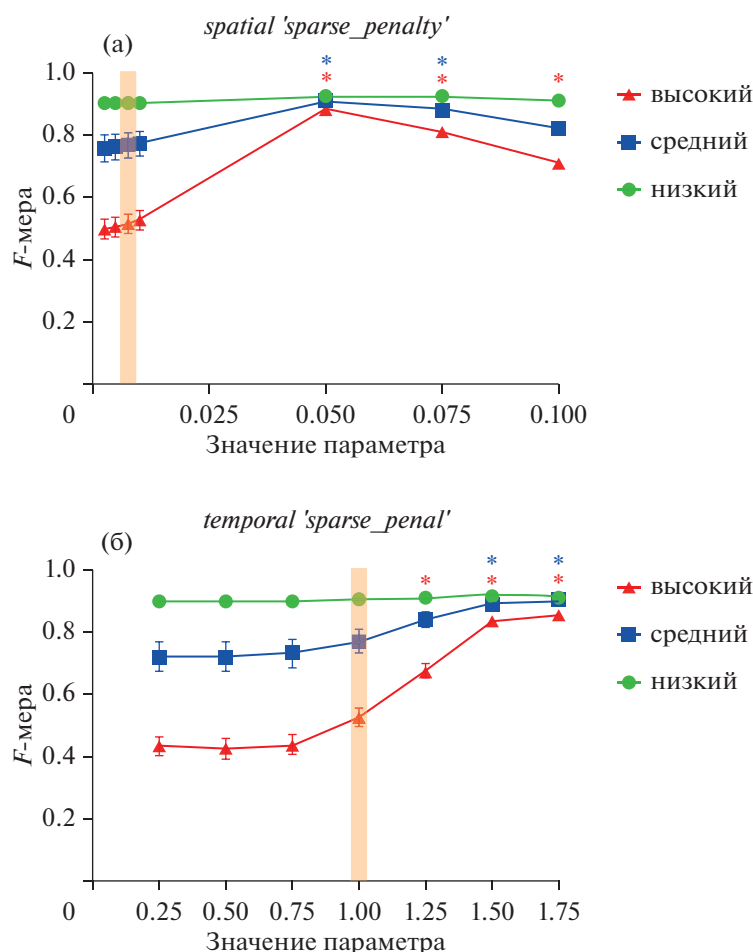


Рис. 4. График изменения значения F -меры в зависимости от значения параметра: (а) — *spatial 'sparse_penalty'*: 0.0025, 0.0050, 0.0075, 0.01, 0.05, 0.075, 0.1; (б) — *temporal 'sparse_penalty'*: 0.25, 0.50, 0.75, 1, 1.25, 1.5, 1.75; для модельного набора данных с разным уровнем шума (низкий, средний, высокий). Оранжевым прямоугольником отображен диапазон значений F -меры для значений параметров по умолчанию при разных уровнях шума. Данные представлены как среднее $\pm$ стандартная ошибка среднего. Зелеными, синими и красными “*” обозначены статистически значимые изменения F -меры относительно значения F -меры при значении параметра по умолчанию для групп данных с низким, средним и высоким уровнем шума соответственно.

Fig. 4. The $F1$ score is plotted as a function of parameter values: (a) — *spatial 'sparse_penalty'*: 0.0025, 0.0050, 0.0075, 0.01, 0.05, 0.075, 0.1; (б) — *temporal 'sparse_penalty'*: 0.25, 0.50, 0.75, 1, 1.25, 1.5, 1.75; for simulated datasets with different noise levels (low, medium, high) processed through Minian. The orange rectangle on the graph shows the range of $F1$ score at the default value of Minian parameters. Data are presented as mean $\pm$ SEM. Green, blue and red “*” symbols indicate significant changes in the $F1$ score compared to the default $F1$ score value for the data groups characterized by low, medium and high noise levels, respectively.

'*ksize*', контролирующего размер медианного фильтра, используемого для подавления шума, должно быть увеличено с 7 до 11. Медианный фильтр является широко применяемым инструментом в цифровой обработке сигналов и изображений для снижения шума. Значение параметра медианного фильтра следует увеличивать с увеличением уровня шума.

Интересно отметить, что значение параметра '*noise_freq*' (частота среза) по умолчанию, равное 0.06, имело наименьшее значение F -меры по сравнению с другими значениями. Воз-

можно, это связано с тем, что список значений параметра '*noise_freq*' является приблизительным и должен быть изменен под конкретный набор данных. Параметр '*noise_freq*' отвечает за фильтрацию полезного сигнала от шума на основе частоты. Как и ожидалось, значение этого параметра должно увеличиваться с 0.06 до 0.2 по мере увеличения уровня шума. В документации к Minian на сайте разработчика, в частности в описании исходного кода, авторы указывают значение 0.25 в качестве значения по умолчанию для параметра '*noise_freq*',

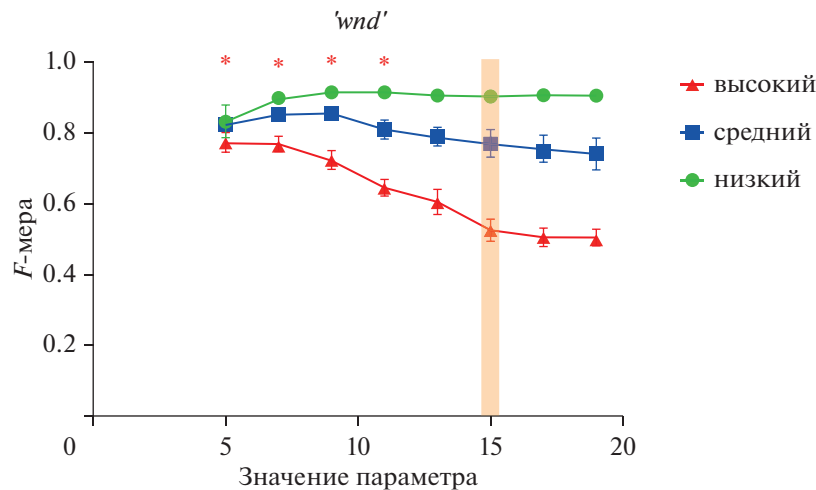


Рис. 5. График изменения значения F -меры в зависимости от значения параметра 'wnd' (5, 7, 9, 11, 13, 17, 19) для модельного набора данных с разным уровнем шума (низкий, средний, высокий). Оранжевым прямоугольником отображен диапазон значений F -меры для значения параметра 'wnd' по умолчанию при разных уровнях шума. Данные представлены как среднее $\pm$ стандартная ошибка среднего. Синими и красными "*" обозначены статистически значимые изменения F -меры относительно значения F -меры при значении параметра по умолчанию для групп данных со средним и высоким уровнем шума соответственно.

Fig. 5. The $F1$ score is plotted as a function of the 'wnd' parameter values (5, 7, 9, 11, 13, 17, 19) for simulated datasets with different noise levels (low, medium, high) processed through Minian. The orange rectangle on the graph shows the range of $F1$ score at the default value of the 'wnd' parameter. The data are presented as mean $\pm$ SEM. Blue and red "*" symbols indicate significant changes in the $F1$ score compared to the default $F1$ score value for the data groups characterized by medium and high noise levels, respectively.

которое близко к оптимальному значению, полученному на нашем наборе данных. Однако в репозитории с кодом указано значение 0.06 в качестве значения по умолчанию. Исходя из полученных результатов, можно сделать вывод, что значение параметра 'noise_freq' должно быть уменьшено с 0.06 до 0.02 при среднем уровне шума и с 0.06 до 0.01 при высоком уровне шума исходных данных.

Параметр *spatial 'sparse_penal'* регулирует разреженность получаемых пространственных следов¹. При увеличении значения этого параметра будет удалено больше пространственных следов, что является важным при высоком уровне шума из-за наличия большого количества артефактов. Значение параметра *spatial 'sparse_penal'* следует увеличивать с ростом уровня шума, от 0.01 до 0.05. Важно отметить, что чрезмерное увеличение значения этого параметра может привести к удалению полезных пространственных отпечатков.

¹ Пространственный след (или пространственный отпечаток) относится к представлению пространственной информации в сигнале.

Неожиданным оказалось то, что изменение значений параметра *temporal 'sparse_penal'* привело к изменению F -меры. Этот параметр контролирует разреженность результирующей кальциевой динамики² и деконволюции пиков. Предполагается, что это может быть связано с тем, что нейроны идентифицируются на основе флуоресцентных сигналов. Если полученный сигнал превышает определенный пороговый уровень, то пространственная область считается полезной и определяется как нейрон. Согласно полученным результатам, рекомендуется увеличить значение этого параметра от 1 до 1.75 по мере увеличения уровня шума.

² Разреженность в данном контексте относится к свойству результирующего кальциевого сигнала или динамики, где большинство значений или пиков имеют нулевые или очень низкие значения, в то время как некоторые из них имеют более высокие значения. Такая разреженность может быть результатом физиологических особенностей клеток или биологических процессов. В анализе кальциевых сигналов разреженность может быть использована для выделения и анализа релевантных событий или активаций.

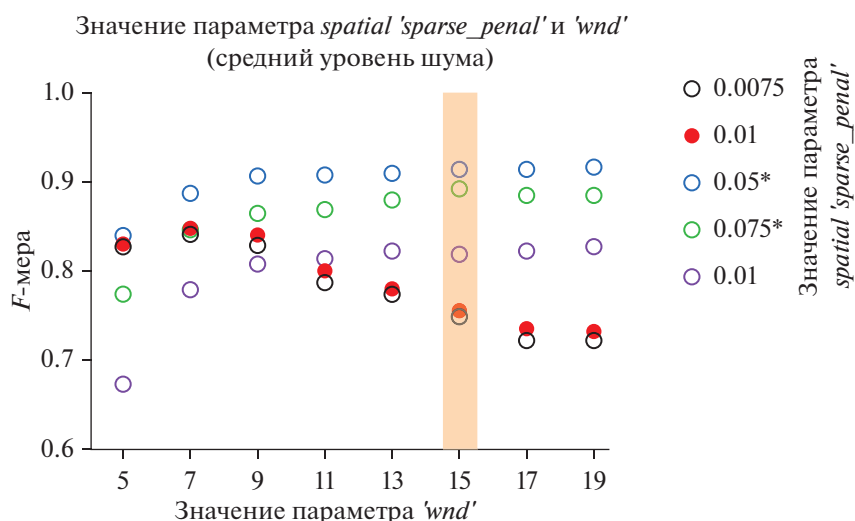


Рис. 6. График изменения значения F_1 -меры при парном изменении значений параметров *'wnd'* (0.005, 0.01, 0.02, 0.06, 0.1, 0.2, 0.3, 0.45, 0.6, 0.8) и *spatial 'sparse_penal'* (0.0075, 0.01, 0.05, 0.075, 0.1) для модельного набора данных с разным уровнем шума (низкий, средний, высокий). Оранжевым прямоугольником отображен диапазон значений F_1 -меры для значения параметра *'wnd'* по умолчанию при разных уровнях шума. Красными окружностями обозначены значения F_1 -меры при стандартном значении параметра *spatial 'sparse_penal'*. Данные представлены как среднее $\pm$ стандартная ошибка среднего. (*) — указывает на значения параметров, при которых наблюдались достоверные изменения F_1 -меры относительно значения F_1 -меры при значении параметра по умолчанию.

Fig. 6. The F_1 score plotted as a function of *'wnd'* (0.005, 0.01, 0.02, 0.06, 0.1, 0.2, 0.3, 0.45, 0.6, 0.8) and *spatial 'sparse_penal'* (0.0075, 0.01, 0.05, 0.075, 0.1) parameters for simulated datasets with medium noise level. The orange rectangle on the graph indicates the range of F_1 score at the default value of *'wnd'* parameter. Red circles indicate the F_1 score at the default value of *spatial 'sparse_penal'* parameter. The data are presented as mean $\pm$ SEM. (*) — indicates parameter values at which significant changes in the F_1 score compared to the F_1 score at the default parameter value.

Наконец, параметр *'wnd'*, который управляет морфологической операцией “tophat”³ для оценки фонового сигнала, следует оставить по умолчанию, то есть 15, при низком и среднем уровнях шума. Однако при высоком уровне шума рекомендуется уменьшить значение параметра с 15 до 5.

Полученные оптимальные значения параметров могут служить начальной отправной точкой для обработки экспериментальных данных. Для этой цели мы предлагаем оценить уровень шума в экспериментальных данных с помощью программы, написанной на языке программирования C++ (см. дополнительные материалы), и инициализировать обработку этих данных в пакете анализа Minian с использованием оптимальных значений

параметров, полученных на модельных данных с аналогичным уровнем шума.

Следует отметить, что в нашем исследовании не было рассмотрено влияние одновременного изменения нескольких параметров Minian на F_1 -меру. В качестве предварительного исследования мы провели оценку влияния одновременного изменения двух параметров: *'wnd'* и *spatial 'sparse_penal'* — на изменение F_1 -меры при среднем уровне шума (рис. 6). Полученные результаты согласуются с предыдущими: значение параметра *spatial 'sparse_penal'* следует увеличить с 0.01 до 0.05, в то время как значение параметра *'wnd'* должно оставаться неизменным.

В дальнейшем мы планируем расширить наше исследование, используя общедоступные наборы данных для однофотонной кальциевой визуализации, а также оценить эффективность предложенного метода при одновременной настройке двух или более параметров.

³ Морфологическая операция “tophat” (преобразование в виде цилиндра) — это операция, используемая в математической морфологии и обработке цифровых изображений, для извлечения мелких элементов и деталей из заданных изображений.

Ограничения исследования

Как было отмечено ранее, в нашем исследовании использовался модельный набор данных, который имитирует пространственно-временные характеристики нейронной активности экспериментальных данных однофотонной микроскопии (образцы из репозитория Minian). Относительный уровень шума в этих экспериментальных данных сопоставим с высоким уровнем шума в модельном наборе данных. Результаты нашего исследования показывают, что изменение параметров Minian относительно значений по умолчанию может повысить точность обнаружения нейронов в случае высокого уровня шума в данных. Однако в случае низкого уровня шума в модельных данных мы не наблюдали существенных изменений в точности обнаружения нейронов при изменении параметров Minian относительно значений по умолчанию.

Рассматриваемый нами подход представляется перспективным для улучшения анализа данных однофотонной микроскопии. Однако следует отметить, что в данных, полученных с помощью двухфотонной и трехфотонной микроскопии, отношение сигнал-шум (SNR) значительно выше, чем в однофотонных данных, поэтому наш подход может быть менее эффективным для улучшения анализа таких данных. В нескольких исследованиях проводилась количественная оценка увеличения SNR при переходе от однофотонной к двухфотонной и трехфотонной микроскопии. Например, исследование (Zhou et al., 2018) указывает на увеличение SNR более чем на 40% при использовании двухфотонного возбуждения по сравнению с однофотонным возбуждением с использованием того же минискапа. Трехфотонная микроскопия позволяет получать изображения с еще более высоким отношением сигнал/шум. Например, исследование (Escobet-Montalban et al., 2018) показывает, что на глубине 450 мкм SNR снижается всего на 15% в трехфотонном режиме, в то время как в двухфотонном режиме SNR снижается на 71%.

В табл. 1 приведены примеры исследований, в которых использовались модельные наборы данных для однофотонной миниатюрной микроскопии. В этих исследованиях моделированные наборы данных применялись для оценки точности обнаружения нейронов и паттернов их активности в различных

пакетах анализа. В нашей работе модельные данные использовались для определения оптимальных параметров Minian. Создание собственного модельного набора данных предоставило нам ряд преимуществ по сравнению с использованием существующих модельных данных. Мы могли гибко настраивать уровень шума, фон и пространственные характеристики нейронов, чтобы обеспечить большую схожесть с экспериментальными данными, полученными из определенной области мозга с использованием определенного минискапа.

Далее мы сравниваем нашу методику генерации модельных данных с другими подходами. Процесс генерации модельных данных описан в статье (Zhou et al., 2018), где рассматривается метод CNMF-e для определения положения и динамики активности нейронов. В последующих исследованиях (Dong et al., 2022) применялась та же методика, предложенная (Zhou et al., 2018). В нашем исследовании мы использовали аналогичный подход для моделирования пространственного распределения нейронов, используя 2D-Гауссиан. В упомянутых работах использовалась модель, в которой временная динамика моделировалась в виде всплесков, генерируемых процессом Бернулли, за которыми следовал экспоненциальный спад. В нашей работе мы не моделировали временную динамику, а напрямую используем динамику активности, полученную в результате обработки экспериментальных данных с параметрами по умолчанию в Minian. Этот подход сводит к минимуму количество параметров генератора, которые необходимо настроить для точного представления реальных данных. Фон можно смоделировать с помощью модели случайного блуждания (Zhou et al., 2018) или просто используя статическое реальное изображение в качестве фона (Lu et al., 2018a). В нашем случае мы использовали однокомпонентный низкочастотный фон, который, в отличие от ранее упомянутых подходов, создает изображения, визуально похожие на реальные данные, и не оказывает сильного влияния на сигнал или шум. В части моделирования шума наша работа не имеет существенных отличий от существующих подходов — шум моделируется как аддитивный случайный компонент с равномерным распределением.

Наше исследование подтверждает целесообразность и предлагает способ настройки параметров Minian при среднем или высоком

уровне шума данных. Однако обобщение этих выводов и рекомендаций по корректровке параметров требует дальнейших исследований. Для улучшения обобщаемости этих рекомендаций целесообразно проверить полученные результаты на расширенной серии модельных данных. Синтезированный набор данных можно расширить, включив более сложное моделирование фона и варьируя количество активных нейронов.

В качестве альтернативы для улучшения обобщаемости рекомендаций по настройке параметров Minian для произвольных данных, дальнейшее развитие может быть сосредоточено на применении предлагаемого подхода к конкретным экспериментальным данным. Раздел 2.2 посвящен этой теме и описывает метод создания модельного набора данных, который воспроизводит пространственно-временные характеристики реальных данных. Однако этот процесс включает несколько этапов, требующих более детальной формализации. Например, в нашем исследовании фон был моделирован упрощенно, в то время как реалистическое моделирование фона требует отдельного исследования. Один из возможных способов избежать необходимости реалистичного моделирования фона — проверить устойчивость оптимальных параметров Minian на синтетических данных с различными статическими и динамическими фонами.

Первое направление состоит в повышении обобщаемости выводов и формулировании универсальных рекомендаций, включающих следующие шаги: 1) измерение уровня шума с использованием процедуры, описанной в разделе 2.2, и 2) настройка параметров Minian на рекомендуемые значения для измеренного уровня шума на шаге 6 процедуры. Второе направление связано с уточнением неопределенных аспектов процедуры генерации модельных данных, которые имитируют реальные данные, а затем — с применением оптимальных параметров Minian к реальным данным. Например, одним из таких аспектов является реалистичное моделирование фона. Первое направление предполагает широкомасштабную проверку полученных результатов на синтетических данных, включая различные модели фона и шума, разное количество нейронов, а также разные временные и пространственные разрешения. Второе направление включает уточнение неопределенных шагов в процедуре создания смоделированных данных,

особенно в отношении реалистичного моделирования фона.

Результаты нашего текущего исследования представляют интерес для исследовательского сообщества по следующим причинам:

1. Они демонстрируют целесообразность настройки параметров в соответствии с уровнем шума на основе наших синтетических данных.
2. Они позволяют оценить влияние изменений конкретных параметров на точность обнаружения нейронов.

ЗАКЛЮЧЕНИЕ

В нашей работе была исследована зависимость между точностью обнаружения нейронов и значениями параметров пакета анализа Minian при обработке модельных данных с разным уровнем шума. В результате исследования мы предлагаем рекомендации по настройке этих параметров в зависимости от уровня шума исходных данных. При низком уровне шума и высоком отношении сигнал/шум значения параметров Minian по умолчанию являются оптимальными. Однако при увеличении уровня шума исходных данных наблюдаются существенные изменения в точности обнаружения нейронов при варьировании значений параметров Minian относительно значений по умолчанию.

БЛАГОДАРНОСТИ

Мы благодарны Александру Митиневу за помощь в реализации программного обеспечения, Анастасии Большаковой за административную помощь, а также сотрудникам лаборатории молекулярной нейродегенерации и лаборатории синаптической биологии за полезные обсуждения и предложения.

ФИНАНСИРОВАНИЕ

Работа выполнена при поддержке гранта Российского научного фонда № 22-75-00028 (А.И.Е.) и программы стратегического академического лидерства Южного федерального университета (“Приоритет 2030”) (И.Б.). Финансовая поддержка была распределена следующим образом: результаты, изображенные на рис. 1–2, были поддержаны Российским научным фондом 22-75-00028, а результаты, изображенные на рис. 3–6, были поддержаны Программой стратегического академического лидерства Южного федерального университета (“Приоритет 2030”).

ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Набор модельных данных, использованных в представленной работе, дополнительные таблицы, а также программу, написанную на языке программирования C++, предназначенную для определения уровня шума, можно найти по ссылке https://mega.nz/folder/KSZT0TqC#CFWoZGxluxnvFJZF_i_U-w5w, <https://disk.yandex.ru/d/WnDxW-spr3VUnw> и <https://www.jvnd.ru/supplemental-materials/>.

СПИСОК ЛИТЕРАТУРЫ

- Aharoni D., Khakh B.S., Silva A.J., Golshani P. All the light that we can see: a new era in miniaturized microscopy. *Nature Methods*. 2019. 16 (1): 11–13.
- Bao Y., Soltanian-Zadeh S., Farsiu S., Gong Y. Segmentation of neurons from fluorescence calcium recordings beyond real time. *Nature Machine Intelligence*. 2021. 3 (7): 590–600.
- Barbera G., Jun R., Zhang Y., Liang B., Li Y., Lin D.-T. A miniature fluorescence microscope for multi-plane imaging. *Scientific Reports*. 2022. 12 (1): 16686.
- Cantu D.A., Wang B., Gongwer M.W., He C.X., Goel A., Suresh A., Kourdougli N., Arroyo E.D., Zeiger W., Portera-Cailliau C. EZcalcium: Open-Source Toolbox for Analysis of Calcium Imaging Data. *Frontiers in Neural Circuits*. 2020. 14.
- de Kraker L., Seignette K., Thamizharasu P., van den Boom B.J.G., Ferreira Pica I., Willuhn I., Levelt C.N., Tögt C.V.D. SpecSeg is a versatile toolbox that segments neurons and neurites in chronic calcium imaging datasets based on low-frequency cross-spectral power. *Cell Reports Methods*. 2022. 2 (10): 100299.
- Denk W., Strickler J.H., Webb W.W. Two-Photon Laser Scanning Fluorescence Microscopy. *Science*. 1990. 248 (4951): 73–76.
- Dong Z., Mau W., Feng Y., Pennington Z.T., Chen L., Zaki Y., Rajan K., Shuman T., Aharoni D., Cai D.J. Minian, an open-source miniscope analysis pipeline. *eLife*. 2022. 11: e70661.
- Engelbrecht C.J., Johnston R.S., Seibel E.J., Helmchen F. Ultra-compact fiber-optic two-photon microscope for functional fluorescence imaging *in vivo*. *Optics Express*. 2008. 16 (8): 5556–5564.
- Erofeev A.I., Barinov D.S., Gerasimov E.I., Pchitskaya E.I., Bolsunovskaja M.V., Vlasova O.L., Bezprozvanny I.B. NeuroInfoViewer: A Software Package for Analysis of Miniscope Data. *Neuroscience and Behavioral Physiology*. 2021. 51 (8): 1199–1205.
- Escobet-Montalban A., Gasparoli F.M., Nyk J., Liu P., Yang Z., Dholakia K. Three-photon light-sheet fluorescence microscopy. *Opt Lett*. 2018. 43 (21): 5484–5487.
- Friedrich J., Giovannucci A., Pnevmatikakis E.A. Online analysis of microendoscopic 1-photon calcium imaging data streams. *PLoS Comput Biol*. 2021. 17 (1): e1008565.
- Ghosh K.K., Burns L.D., Cocker E.D., Nimmerjahn A., Ziv Y., Gamal A.E., Schnitzer M.J. Miniaturized integration of a fluorescence microscope. *Nature Methods*. 2011. 8 (10): 871–878.
- Giovannucci A., Friedrich J., Gunn P., Kalton J., Brown B.L., Koay S.A., Taxidis J., Najafi F., Gauthier J.L., Zhou P., Khakh B.S., Tank D.W., Chklovskii D.B., Pnevmatikakis E.A. CaImAn an open source tool for scalable calcium imaging data analysis. *eLife*. 2019. 8: e38173.
- Grienberger C., Konnerth A. Imaging calcium in neurons. *Neuron*. 2012. 73 (5): 862–885.
- Iwasaki S., Ikegaya Y. *In vivo* one-photon confocal calcium imaging of neuronal activity from the mouse neocortex. *J. Integr. Neurosci.* 2018. 17 (3–4): 671–678.
- Li M., Liu C., Cui X., Jung H., You H., Feng L., Zhang S. An Open-Source Real-Time Motion Correction Plug-In for Single-Photon Calcium Imaging of Head-Mounted Microscopy. *Frontiers in Neural Circuits*. 2022. 16.
- Liberti W.A., Perkins L.N., Leman D.P., Gardner T.J. An open source, wireless capable miniature microscope system. *J. Neural Engineering*. 2017. 14 (4): 045001.
- Lu J., Li C., Singh-Alvarado J., Zhou Z.C., Fröhlich F., Mooney R., Wang F. MIN1PIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline. *Cell Rep*. 2018a. 23 (12): 3673–3684.
- Lu J., Li C., Singh-Alvarado J., Zhou Z.C., Fröhlich F., Mooney R., Wang F. MIN1PIPE: A Miniscope 1-Photon-Based Calcium Imaging Signal Extraction Pipeline. *Cell Reports*. 2018b. 23 (12): 3673–3684.
- Mukamel E.A., Nimmerjahn A., Schnitzer M.J. Automated Analysis of Cellular Signals from Large-Scale Calcium Imaging Data. *Neuron*. 2009. 63 (6): 747–760.
- Pachitariu M., Stringer C., Dipoppa M., Schröder S., Rossi L.F., Dagleish H., Carandini M., Harris K.D. Suite2p: beyond 10000 neurons with standard two-photon microscopy. *bioRxiv*. 2017. 061507.
- Pnevmatikakis E.A., Giovannucci A. NoRMCorr: An online algorithm for piecewise rigid motion correction of calcium imaging data. *J. Neuroscience Methods*. 2017. 291: 83–94.
- Pnevmatikakis Eftychios A., Soudry D., Gao Y., Machado T.A., Merel J., Pfau D. et al. Simultaneous Denoising, Deconvolution, Demixing of Calcium Imaging Data. *Neuron*. 2016. 89 (2): 285–299.
- Pochechuev M.S., Fedotov I.V., Ivashkina O.I., Roshchina M.A., Meshchankin D.V., Sidorov-Biryukov D.A., Fedotov A.B., Anokhin K.V., Zheltikov A.M. Reconnectable fiberscopes for chronic *in vivo* deep-brain imaging. *J. Biophotonics*. 2018. 11 (4): e201700106.
- Radstake F.D.W., Raaijmakers E.A.L., Luttge R., Zinger S., Frimat J.P. CALIMA: The semi-automated open-source calcium imaging analyzer. *Computer Methods and Programs in Biomedicine*. 2019. 179: 104991.
- Resendez S.L., Stuber G.D. *In vivo* Calcium Imaging to Illuminate Neurocircuit Activity Dynamics Underlying Naturalistic Behavior. *Neuropsychopharmacology*. 2015. 40 (1): 238–239.
- Robbins M., Christensen C., Kaminski C., Zlatić M. Calcium imaging analysis ? how far have we come?

- [version 2; peer review: 3 approved]. F1000Research. 2021. 10 (258).
- Russell J.T. Imaging calcium signals in vivo: a powerful tool in physiology and pharmacology. *British J. Pharmacology*. 2011. 163 (8): 1605–1625.
- Sità L., Brondi M., Lagomarsino de Leon Roig P., Curreli S., Panniello M., Vecchia D., Fellin T. A deep-learning approach for online cell identification and trace extraction in functional two-photon calcium imaging. *Nature Communications*. 2022. 13 (1): 1529.
- ter Veer M.J.T., Pfeiffer T., Nägerl U.V. Two-Photon STED Microscopy for Nanoscale Imaging of Neural Morphology In Vivo. *Super-Resolution Microscopy: Methods and Protocols*. 2017. 45–64.
- Vogt N. Two-photon imaging in freely behaving mice. *Nature Methods*. 2022. 19 (5): 518–518.
- Wu X., Yang X., Song L., Wang Y., Li Y., Liu Y., Yang X., Wang Y., Pei W., Li W. A Modified Miniscope System for Simultaneous Electrophysiology and Calcium Imaging *in vivo*. *Frontiers in Integrative Neuroscience*. 2021. 15.
- Zátonyi A., Madarász M., Szabó Á., Lőrincz T., Hodován R., Rózsa B., Fekete Z. Transparent, low-autofluorescence microECOG device for simultaneous Ca^{2+} imaging and cortical electrophysiology *in vivo*. *J. Neural Engineering*. 2020. 17 (1): 016062.
- Zhou P., Resendez S.L., Rodriguez-Romaguera J., Jimenez J.C., Neufeld S.Q., Giovannucci A., Friedrich J., Pnevmatikakis E.A., Stuber G.D., Hen R., Kheirbek M.A., Sabatini B.L., Kass R.E., Paninski L. Efficient and accurate extraction of *in vivo* calcium signals from microendoscopic video data. *eLife*. 2018. 7: e28728.
- Zong W., Obenhaus H.A., Skytøen E.R., Eneqvist H., de Jong N.L., Vale R., Jorge M.R., Moser M.-B., Moser E.I. Large-scale two-photon calcium imaging in freely moving mice. *Cell*. 2022. 185 (7): 1240–1256.e1230.
- Zong W., Wu R., Chen S., Wu J., Wang H., Zhao Z. et al. Miniature two-photon microscopy for enlarged field-of-view, multi-plane and long-term brain imaging. *Nature Methods*. 2021. 18 (1): 46–49.

ON OPTIMIZING MINISCOPE DATA ANALYSIS WITH SIMULATED DATA: A STUDY OF PARAMETER OPTIMIZATION IN THE MINIAN ANALYSIS PIPELINE

A. I. Erofeev^{a, #, †}, M. V. Petrushan^{b, †}, L. V. Lysenko^b,
E. K. Vinokurov^a, O. L. Vlasova^a, and I. B. Bezprozvanny^{a, b, c, ##}

^aLaboratory of molecular neurodegeneration, Peter the Great St. Petersburg polytechnic university, St. Petersburg, Russia

^bLaboratory of Synaptic Biology, Southern Federal University, Rostov-On-Don, Russia

^cDepartment of Physiology, University of Texas Southwestern Medical Center at Dallas, Dallas, TX, United States of America

[#]e-mail: alexander.erofeev@gmail.com

^{##}e-mail: ilya.bezprozvanny@utsouthwestern.edu

In vivo calcium imaging is widely used technique in neuroscience to evaluate the activity of neuronal networks. The miniscope, a single-photon miniature fluorescent microscope, has made it possible to conduct *in vivo* calcium imaging in freely moving animals. Various algorithms and software packages have been developed for the analysis of miniscope data. This study investigates the relationship between the sensitivity of neuron detection and the processing parameters utilized in the Minian analysis pipeline at different noise levels. To achieve this objective, we generated simulated data possessing certain attributes of an experimentally derived dataset. Simulated data was generated with various noise levels and processed through to the Minian analysis pipeline. Based on our findings, we provide recommendations for optimal values of Minian pipeline parameters depending on different noise levels. The results obtained in this study may serve as a preliminary guide for selecting appropriate parameter values during the processing of experimental data using the Minian analysis pipeline. The findings of this study are expected to be relevant to neuroscientists involved in the acquisition and processing of miniscope data.

Keywords: miniscope, fluorescence, calcium, Minian, modeling, parameter optimization

[†] These authors have contributed equally to this paper and share first authorship.