

Классификация органоидов, дифференцирующихся в сетчатку глаза, методом глубокого обучения с предобработкой данных

Работу выполнили:

Семенов Мария, школа ОАНО “Школа “Летово”, Москва

Коломытцева Екатерина, Предуниверсарий МАИ, Москва

Наставники:

Ибрагимов Алишер, физический факультет МГУ

Сенотрусова Софья, физический факультет МГУ

При поддержке команды ИСП РАН и МФТИ:

Наумов Антон Юрьевич, Кегелес Евгений Александрович, Волчков Павел

Юрьевич, Карпулевич Евгений Андреевич

2021

Оглавление

Введение	3
Литературный обзор	4
Цели и задачи	4
Основная часть	5
Вывод	7
Список литературы	8

Введение

Глаз — часть зрительного анализатора животных, а также важная часть центральной нервной системы в целом, передающая информацию в затылочную долю коры больших полушарий по зрительному нерву.

Глаз имеет 3 оболочки: наружная — склера — плотная оболочка, она выполняет защитную функцию; сосудистая оболочка — препятствует рассеиванию света через склеру; сетчатка — внутренняя оболочка, рецепторная часть зрительного анализатора. Сетчатка состоит из пигментного слоя, фоторецепторных клеток и нейронов сетчатки.

В настоящее время широко практикуется лечение многих заболеваний сетчатки с помощью метода клеточной терапии. Данный метод позволяет лечить даже те заболевания, которые еще несколько десятков лет назад считались необратимыми и вели к потере зрения.

Идея клеточной терапии заключается в трансплантации дополнительных нейронов в орган, что позволяет восстановить его до определенного уровня.

Стволовые плюрипотентные клетки способны создавать сетчаткоподобные структуры — органоиды. Однако, дифференциация стволовых клеток в нейроны сетчатки не всегда проходит успешно. Для определения качества дифференцировки уже на 6 день после ее начала существует метод Bright-Field, реализованный с помощью машинного обучения [1].

В проекте с помощью глубокого обучения нейронной сети был реализован данный алгоритм. Глубокое обучение — метод машинного обучения, который позволяет обучить модель предсказывать результат по набору входных данных. Нейронная сеть — программное воплощение математической модели, построенной по принципу взаимодействия нервных клеток животных. Нейросеть состоит из входного слоя, получающего входные данные, какого-то количества скрытых слоев, обрабатывающих входные данные, и выходного слоя, выводящего результат.

Искусственный нейрон работает аналогично настоящему: каждый нейрон скрытого слоя имеет значение от 0 до 1. Нейроны принимают входные значения от предыдущих нейронов, суммируют их и умножают на веса. Для определения оптимальных весов и минимизирования ошибок нейронную сеть необходимо обучить на большом количестве входных данных.

Для анализа изображений был разработан специальный тип нейронных сетей — сверточная нейронная сеть (CNN). Их отличительной особенностью является сверточный слой. Данный слой представляет из себя набор карт признаков, у каждой из которых есть синаптическое ядро. Ядро представляет собой систему разделяемых весов или синапсов. Оно скользит по области предыдущей карты и находит определенные признаки у объектов. Анализ данных при помощи CNN очень распространен в области биологии и медицины.

Результатом проекта стала нейронная сеть, которая обучена отличать стволовые клетки, которые в итоге успешно дифференцируются в нейроны сетчатки, от стволовых клеток, которые не станут клетками сетчатки.

Литературный обзор

Для реализации метода дифференцировки органоидов ученым пришлось совершить несколько значимых открытий, а путь к реализации этой цели начался еще в конце 20 века. Первой значимой ступенью было открытие метода культивирования эмбриональных стволовых клеток *in vitro* [2]. В 2007 году было дано описание индуцированных человеком плюрипотентных стволовых клеток, полученных из соматических клеток [3]. Через малый промежуток времени появились исследования, в процессе которых наблюдалась способность дифференциации плюрипотентных стволовых клеток в клетки сетчатки [4].

Далее в лабораториях описывались новые методы дифференциации органоидов сетчатки, а с развитием технологий машинного обучения его методы также стали применять для работы в данной сфере.

Недавний проект, опубликованный в 2020 году, затрагивал тему предсказания качества дифференцировки органоидов на ранних этапах их развития [5]. Чтобы реализовать данную задачу была написана нейронная сеть. Для получения изображений клетки сканировали с помощью фазово-контрастного микроскопа на пятый день после начала дифференцировки. Далее продолжали поддерживать условия для дальнейшей дифференцировки клеток. Определение клеток-предшественников нейронов сетчатки происходило, основываясь на экспрессии репортерного гена Rx-GFP, который является специфичным маркером клеток-предшественников нейронов сетчатки. Если клетка являлась предшественником клетки сетчатки, то она имела яркую флуоресценцию.

Всего было получено 1209 изображений, которые были разделены на 3 группы: "обучающую", "проверочную" и "тестовую", после этого

фотографии были использованы для обучения нейронной сети распознавать правильно и неправильно дифференцирующиеся клетки. Нейросеть, созданная в проекте, определяет правильность дифференцировки органоида с точностью 84%.

В нашем проекте реализуются предсказания на новой архитектуре, которая не использовалась ранее, с предварительными методами обработки для увеличения точности нейронной сети.

Цели и задачи

Цель работы: создать оптимизированную нейронную сеть, которая будет предсказывать качество дифференцировки органоидов на ранних этапах их развития с наибольшей точностью.

Задачи:

1. Изучить литературу по теме, в т.ч. основные открытия, связанные с изучением и методами применения в медицине и науке плюрипотентных стволовых клеток, а также возможностью упрощения их изучения с помощью методов машинного обучения;
2. Изучить основные методы глубокого обучения, основные библиотеки, необходимые для реализации проекта (matplotlib, pillow, numpy, scipy, pytorch, openCV);
3. Используя полученные знания, написать и обучить нейронную сеть на предобработанных данных;
4. Проверить точность работы нейронной сети на полученных из лаборатории данных, сделать соответствующие выводы.

Основная часть

Первым этапом проекта было изучение необходимых для дальнейшей работы библиотек: numpy, matplotlib, pillow, openCV. При помощи библиотеки pillow мы смогли написать программу, которая усредняет изображения попиксельно, а также совершает попиксельную дисперсию (дисперсия — это величина, которая показывает, насколько сильно остальные пиксели отличаются от среднего значения).

Используя библиотеку OpenCV, мы реализовали алгоритм, обрезающий фотографии ровно по границе органоида. Для реализации данного алгоритма лучше всего подошли методы Blob Detection [6] и метод Отцу [7].

Для того, чтобы было легче проверить корректность работы нейронной сети, было решено сначала написать алгоритм, который

различает не “сетчатку” (retina) и “не сетчатку” (non-retina), а кошек и собак. Средой программирования был выбран Google Colab — бесплатный облачный сервис на основе Jupyter Notebook. Google Colab предоставляет бесплатный доступ к мощным графическим процессорам, которые можно использовать для задач глубокого обучения.

Для проекта была использована уже обученная сверточная нейронная сеть VGG16. Данная модель нейросети была предложена К. Simonyan и А. Zisserman [8], и достигает точности 92,7% при тестировании на ImageNet, датасет которого состоит более чем из 14 миллионов изображений. Из этого вытекает минус данной нейронной сети - большой вес (нейросеть весит более 533 МБ).

```
nn_model = models.vgg16(pretrained=True)
# Freeze model weights
for param in nn_model.parameters():
    param.requires_grad = False
```

Для реализации проекта было необходимо переобучить последний слой данной нейронной сети таким образом, чтобы она определяла только 2 типа объектов (в случае проекта, retina и non-retina).

```
nn_model.classifier[6] = nn.Sequential(
    nn.Linear(4096, 256),
    nn.ReLU(),
    nn.Dropout(0.4),
    nn.Linear(256, 2),
    nn.LogSoftmax(dim=1)
)
```

Картинки для обучения нейронной сети были разделены на 3 группы: validation (исходные картинки), train (измененные, например, перевернутые, картинки из validation), test (часть картинок, которые не участвуют в коде и на которых проверяется корректность работы нейронной сети после обучения). В процессе обучения в основном участвовали картинки из validation, но точность обучения проверялась также и на картинках из train. Картинки из test были задействованы только 1 раз в самом конце программы, когда нейросеть уже обучилась, чтобы проверить ее точность. Это позволило обучить нейросеть до высокой

точности определения органоидов, несмотря на не слишком большое количество имеющихся фотографий.

```
data = {
    'train':
        myDataset(trainindir, transform=image_transforms['train']),
    'valid':
        myDataset(trainindir, transform=image_transforms['valid']),
    'test':
        myDataset(trainindir, transform=image_transforms['test']),
}
```

Также для увеличения точности, в алгоритме было поставлено условие, что нейронная сеть может получать одни и те же картинки на вход без увеличения точности 10 раз. Когда нейросеть достигает наилучшей точности, она сохраняется. Таким образом возможно “натренировать” нейронную сеть до наиболее высокой точности узнавания объектов, избежав риска неправильного обучения и потери точности.

```
n_epochs_stop = 10
max_val_accuracy = -np.Inf
epochs_no_improve = 0

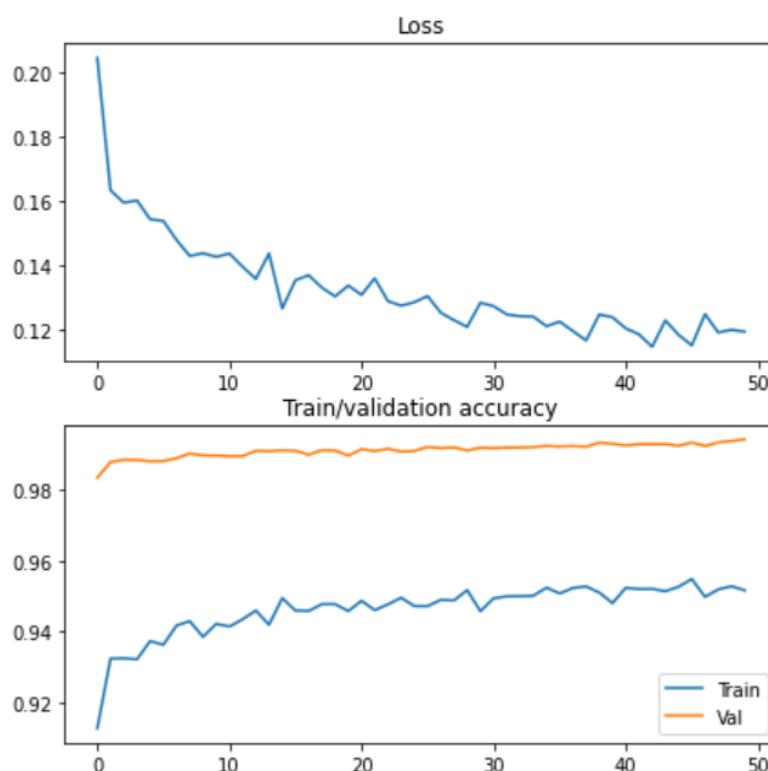
if val_accuracy > max_val_accuracy:
    # Save the model
    torch.save(model, path_to_save)
    epochs_no_improve = 0
    max_val_accuracy = val_accuracy
else:
    epochs_no_improve += 1
    # Check early stopping condition
    if epochs_no_improve == n_epochs_stop:
        print('Early stopping!')
        break
```

В итоге при тестовом запуске нейронной сети, определяющей котов и собак, точность достигала 99.3%, что является очень высоким результатом.

```
test_accuracy = compute_accuracy(nn_model, test_loader)
test_accuracy
```

0.993

С помощью библиотеки matplotlib по результатам обучения нейросети были построены графики, иллюстрирующие весь процесс обучения.



На графиках видно, как изменяется точность нашей сети с количеством эпох тренировки. Обратите внимание, что точность на валидационных данных выше, чем на тренировочных, значит у нас не происходит так называемого переобучения (когда сеть просто запоминает картинки, но не признаки на них).

Далее картинки котов и собак были заменены на фотографии органоидов. Для достижения большей точности определения органоидов был введен новый метод предобработки изображений, не использующийся в других работах на эту тему. Тем не менее, точность определения класса объекта в случае классов retina и non-retina оказалась ниже, чем при определении котов и собак. Это связано с тем, что в случае определения класса органоида все объекты похожи больше, чем коты и собаки, следовательно, нейронная сеть способна выделить меньше отличительных черт каждого класса. В данном случае точность достигла 85%.

Вывод

В ходе работы мы смогли достичь поставленной цели, а именно создали нейронную сеть, определяющую качество дифференцировки органоидов на ранних этапах их развития. Была проделана большая работа: с начала учебного года мы работали над этим проектом, тренируясь на совершенно разных материалах, чтобы в итоге получить высокий результат для более сложных изображений сетчаток.

Мы смогли получить точность, равную 85%, на нейронной сети, которая раньше не использовалась в проекте, добавив автоматическую обрезку изображений от лишних деталей, что позволяет сделать акцент только на органоиде, и как следствие, приводит к увеличению точности.

В заключение хочется выразить слова благодарности наставникам проекта Сириус.Лето: Сенотрусовой Софье и Ибрагимову Алишеру за плодотворные занятия по программированию и работе над проектом. Также, пользуясь возможностью, отдельно хочется поблагодарить команду из ИСП РАН за предоставленные возможности по проекту, советы по реализации и научные наставления, а именно:

1. Наумов Антон Юрьевич,
2. Кегелес Евгений Александрович
3. Волчков Павел Юрьевич
4. Карпулевич Евгений Андреевич

Список литературы

1. Buggenthin F, Marr C, Schwarzfischer M, Hoppe PS, Hilsenbeck Schroeder T, et al. An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. BMC Bioinformatics. 2013;14: 297.
2. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, et al. Embryonic stem cell lines derived from human blastocysts. Science. 1998;282: 1145–1147.
3. Osakada F, Jin Z-B, Hiram Y, Ikeda H, Danjyo T, Watanabe K, et al. In vitro differentiation of retinal cells from human pluripotent stem cells by small-molecule induction. J Cell Sci. 2009;122: 3169–3179.
4. Fligor CM, Huang K-C, Lavekar SS, VanderWall KB, Meyer JS. Differentiation of retinal organoids from human pluripotent stem cells. Methods Cell Biol. 2020;159: 279–302.
5. Kegeles E, Naumov A, Karpulevich EA, Volchkov P, Baranov P. Convolutional Neural Networks Can Predict Retinal Differentiation in Retinal Organoids. Front Cell Neurosci. 2020;14: 171.
6. <https://learnopencv.com/blob-detection-using-opencv-python-c/>
7. https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html
8. VGG16 - сверточная сеть для выделения признаков изображений (neurohive.io)