

УДК 004.032.26

ПОДГОТОВКА МАТЕРИАЛОВ ДЛЯ МАШИННОГО ОБУЧЕНИЯ С ПОМОЩЬЮ ВИРТУАЛЬНОЙ СТУДИИ



Л.Р. Коркин

ассистент кафедры Инженерной психологии и эргономики БГУИР, магистр техники и технологии, аспирант.



М.В. Коркина

младший научный сотрудник производственно-испытательной лаборатории РУП «Институт мясо-молочной промышленности»



Е.С.Маклак

инженер-микробиолог производственно-испытательной лаборатории РУП «Институт мясо-молочной промышленности»



К.Ф. Саевич.

профессор кафедры физикохимии материалов и производственных технологий БГЭУ, доктор биологических наук, профессор.



О.С.Медведев

магистр техники и технологии, инженер-программист, кафедра инженерной психологии и эргономики БГУИР

Л.Р. Коркин

В 2018 году окончил БГУИР, специальность «Инженерно-психологическое обеспечение информационных технологий». Квалификация «инженер-системотехник». В 2020 году окончил магистратуру по специальности «Управление безопасностью производственных процессов» с присвоением академической степени магистра техники и технологии. Читаемые курсы: безопасность жизнедеятельности человека, интерфейсы информационных систем, схемотехника, активные элементы систем безопасности.

М.В. Коркина

В 2017 году окончила Белорусский государственный университет, специальность «Микробиология». Квалификация «биолог-микробиолог». Младший научный сотрудник производственно-испытательной лаборатории РУП «Институт мясо-молочной промышленности».

Е. С. Маклак

В 2019 году окончила Белорусский государственный университет, специальность «Микробиология». Квалификация «биолог-микробиолог». Инженер-микробиолог производственно-испытательной лаборатории РУП «Институт мясо-молочной промышленности».

О.С. Медведев

В 2013 году окончил БГУИР, специальность «Инженерно-психологическое обеспечение информационных технологий». Квалификация инженер-системотехник. В 2019 году окончил магистратуру по специальности «Управление безопасностью производственных процессов» с присвоением академической степени магистра техники и технологии. Ведет курсы «CISCO».

К.Ф. Саевич

В 1984 защитил диссертацию на соискание учёной степени кандидата сельскохозяйственных наук, в 1992 — доктора биологических наук, с 1998 — профессор. Области научных интересов: экология, рациональное природопользование, безопасность жизнедеятельности человека. Достигнутые научные результаты: изобретений — 3, монографий — 6, брошюр — 20, учебных пособий — 3, учебно-методических работ — более 20, участие в научных семинарах и конференциях: Международных — более 20, Республиканских — более 15. Всего опубликовано более 350 работ.

Аннотация. В настоящее время для идентификации микроорганизмов по морфологическим признакам в учреждениях контроля качества продуктов питания используют определитель бактерий Берджи. Этот метод трудоемкий и требует большого внимания и точности выполнения. Таким образом, разработка автоматизированной системы распознавания колоний микроорганизмов на чашках Петри для последующего подсчета и предварительной идентификации микроорганизмов является актуальной задачей.

Ключевые слова. Машинное обучение, искусственный интеллект, микробиология, колонии микроорганизмов.

Введение.

Чтобы превратить распознавание объектов на изображении из идеи в реальность нам понадобится использовать такой вид машинного обучения как обучение нейронной сети (ее еще называют первыми ступенями развития искусственного интеллекта будущего), которое всегда начинается с подготовки материалов, а в нашем случае это размеченные и подготовленные должным образом изображения колоний (снимки колоний микроорганизмов (СКМ)), которые нужно будет распознавать для последующего подсчета и предварительной идентификации микроорганизмов (определение качественного и количественного состава микробиоты исследуемого продукта питания). Для этого мы создаем систему распознавания СКМ (СРСКМ), которая будет способна выявлять наличие колоний микроорганизмов (КМ) на снимках, а также проводить подсчет и предварительную идентификацию микроорганизмов, выросших на чашке Петри. Это позволит выявлять некачественные продукты питания, а также поможет повысить продовольственную безопасность и качество жизни людей.

Процесс усложняется тем, что в некоторых случаях метод подсчета не дает стопроцентный результат даже при применении счетчика колоний микроорганизмов. Принцип работы этого счетчика заключается в том, что оператор с помощью электропера наносит точки на дно чашки Петри в места нахождения колоний. Прибор регистрирует факт касания, а результат выводится на цифровое табло. Счетчик регистрирует только факт касания электропера, что не может застраховать конечный результат от ошибок оператора. В других случаях подсчет производится математически, т.е. проводится подсчет в определенной области, как правило, в одной четвертой чашки Петри, а после умножается на четыре. Есть и другие методы определения числа колоний микроорганизмов, но какой бы из существующих методов определения мы бы не выбрали, каждый из них имеет свои недостатки.

Проведение качественного анализа вызывает еще больше сложностей, так как у одного и того же вида микроорганизмов в зависимости от условий культивирования (температуры, pH, доступа кислорода, источников углерода и др.) колонии могут вырастать с отличающимися морфологическими признаками. Достаточно часто встречается, что колонии разных видов микроорганизмов имеют схожие морфологические признаки.

В настоящее время идентификацию микроорганизмов по морфологическим признакам проводят в основном с помощью определителя бактерий Берджи (1997 г., в двух

томах). Также выпущено множество специальных каталогов и плакатов, на которых показано, чем отличаются колонии разных видов микроорганизмов. Эти материалы могут быть использованы для обучения системы распознаванию КМ.

Методы и материалы.

В процессе обучения СРСКМ необходимо учитывать большое количество различных факторов, таких как разнообразие исследуемых продуктов питания, качество полученных снимков и многое другое. Следует учитывать, что наряду с качественными изображениями на практике придётся использовать также СКМ более низкого качества. В результате реализации проекта станет возможным идентифицировать микроорганизмы по СКМ, выращенных на чашках Петри (рисунок 1) в разных питательных средах, полученным при различном освещении.

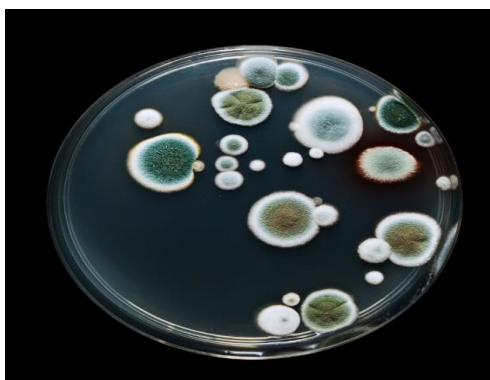


Рисунок 1. Снимок колоний плесневых грибов, выросших на чашке Петри.

Как известно, машинное обучение строится на трех основных принципах: данные, признаки, алгоритмы.

Особенности применения полносвёрточных нейронных сетей.

Полносвёрточные сети — это особый тип искусственных нейронных сетей, результатом работы которых является сегментированное изображение оригинала, где искомые элементы уже выделены требуемым образом.

Полносверточные нейронные сети используются для задач, в которых необходимо, например, определить форму и местоположение искомого объекта или нескольких объектов. Подобные задачи проблематично решать с использованием простых свёрточных нейросетей. Для общего понимания, почему и когда лучше использовать полносвёрточные сети вместо обычных свёрточных, необходимо сравнить указанные типы нейронных сетей.

Сравнение свёрточных и полносвёрточных нейронных сетей.

Самым очевидным отличием полносвёрточных сетей от других нейросетей является конечный результат работы сети. Простые сверточные сети могут использоваться для классификации, определяя к какому классу принадлежит то или иное изображение, и для локализации объекта на изображении (регрессия).

В результате применения любого из этих методов выходными данными являются числа либо массивы чисел. Иначе говоря, мы можем получить информацию о изображении (при этом очень ограниченную), но не можем преобразовать его в необходимый нам вид.

Полносвёрточные нейронные сети на выходе выдают сегментированное изображение, соответствующее по размерности входному. Поэтому их второе название — сегментационные нейронные сети.

Сегментация — это объединение объектов в группы по общим признакам. Таким образом, мы получаем от сети намного больше информации, а достигнутый результат можем обработать простыми эвристическими методами.

Данные (Data) — это базовая информация. Сюда входят необходимые выборки данных, работе с которыми требуется обучить систему. Чем больше данных загружается в СРСКМ, тем точнее и лучше она будет работать. Данные напрямую зависят от задачи, которая стоит перед нейронной сетью.

Выборка данных и ее обработка занимает большой промежуток времени. Это повысит точность оценки и прогнозирования, которую мы получаем в итоге. При этом очень важно предоставлять максимум разнообразной информации, поскольку система может увидеть связь там, где человек ее не заметит.

Изучение признаков (свойства, метрики, фичи, характеристики, features и другие признаки). Для разработки СРСКМ совместно с сотрудниками Института мясо-молочной промышленности будут определены характеристики и свойства, которые система будет отслеживать в ходе обучения.

Поскольку правильность определения свойств напрямую влияет на результат, который необходимо получить, их отбор занимает зачастую больше времени, чем сам процесс обучения. Здесь главное — не торопиться и не ограничивать набор характеристик, так как можно исказить восприятие системы, а вместе с ним и конечный результат.

Исходя из анализа свойств компонентов можно определить признаки и в зависимости от задач установить основные свойства.

Разрабатывается алгоритм, который предоставляет системе набор последовательных операций и методы для решения поставленной задачи. Именно от выбранного метода напрямую зависит скорость и точность результата обработки исходных данных.

Очень важно подготовить множество возможных вариантов изображений колоний определенного вида микроорганизмов, отличающихся по морфологическим признакам, выращенных в различных средах, сфотографированных из разных положений. Каждый СКМ должен быть размечен, а колонии должны быть выделены, чтобы СРСКМ понимала, где на СКМ объекты, которые необходимо научиться распознавать. Если делать это всё вручную, понадобится очень много времени.

В нашей работе было принято решение проводить обучение в двух направлениях: проведение количественного анализа и создание виртуальной студии на основе Blender, с помощью которой можно будет сгенерировать размеченные изображения для обучения системы с использованием данных литературных источников.

Первый этап обучения планируется проводить около 5 месяцев. За это время СРСКМ должна научиться распознавать наличие КМ на чашке Петри, а также подсчитывать количество выросших колоний. Результаты работы могут быть неудовлетворительными, так как система будет распознавать колонии микроорганизмов только со снимков, сделанных в идеальных условиях, а тень, свет, искажение могут привести к критической потере точности определения.

На следующем этапе обучения СРСКМ будут учтены особенности получения снимков. Для этих целей на основе языка программирования python и библиотек машинного зрения таких как Open CV, matplotlib, numpy, imutils и других система будет разбивать свои действия на шаги:

1. распознать чашку Петри;
2. сконвертировать изображение из RGB (на языке python это BGR) в черно-белое (или двухцветное) изображение (рис 2 а, б, в, г.);
3. распознать контуры МО по разности контрастов;
4. подсчитать количество КМ после выделения контуров.

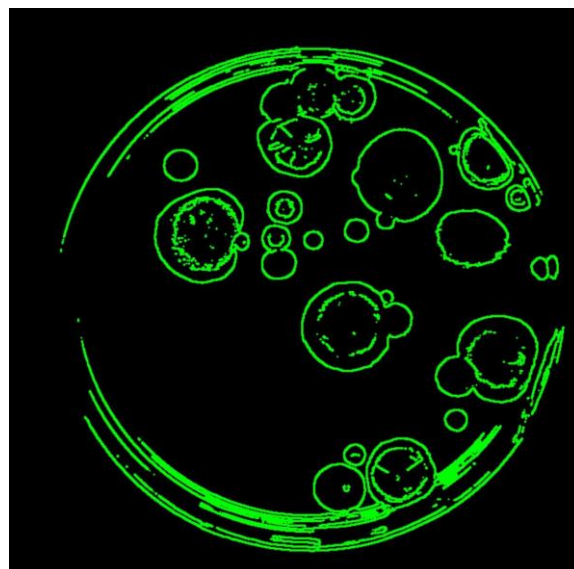
В лаборатории свои особенности помещения, отличные от тех, что в фотостудии, где возможно получение СКМ высокого качества. Также стоит учитывать, что пользователь будет делать снимки в основном при помощи фотокамеры мобильного телефона, что не относится к оборудованию для

профессиональной фотосъемки. Поэтому при подготовке и обучении системы необходимо учесть такие факторы, как:

- неудачный ракурс;
- слишком яркое или тусклое освещение;
- плохой фон;
- фото с перспективой.



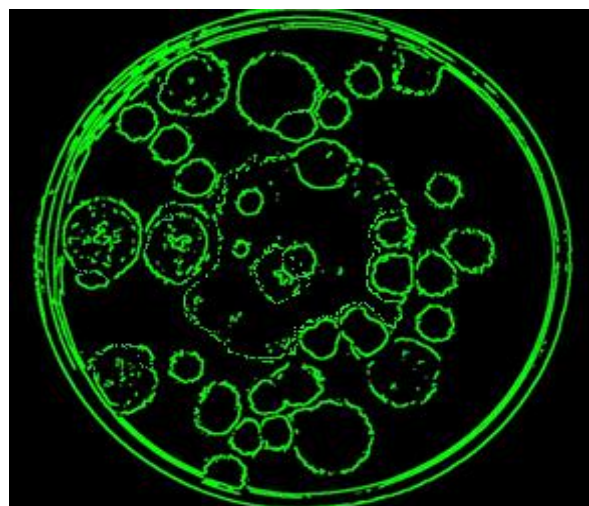
А



Б



В



Г

Рисунок 1. Начальное (а, в) и сконвертированное (б, г) изображения из BGR в двухцветное изображение

Для подготовки высокоэффективной и точной СРСКМ на первом этапе работы было принято решение изготовить (оборудовать) «мобильная мини фотостудию», которая будет состоять из цифрового фотоаппарата большого разрешения (в формате JPEG 1:1 и разрешение не ниже 2656 x 2656 с полноразмерной не менее 24-мегапиксельной матрицей), установленного на штатив, и подставки с внутренней подсветкой, которая будет

подсвечивать чашки Петри. Дополнительно можно установить светорассеивающие экраны, а при необходимости и дополнительное внешнее освещение.

На втором этапе работы после того, как система научится выявлять колонии МО на чашке Петри и подсчитывать их, планируется обучить систему сравнивать полученные СКМ с имеющимися в базе данных изображениями со СКМ. Для этих целей необходимо сделать по одному снимку колоний каждого исследуемого вида микроорганизмов и подобрать некоторое количество разных фонов. Далее с помощью скрипта, подключенного к Blender, изображение колоний микроорганизмов будет подставлено на каждый фон.

В созданной нами мини фотостудии имеется виртуальная камера, которая позволяет создать анимацию при изменении положения. В процессе анимации скрипт делает рендеры каждого кадра, и у нас появляется база изображений (материалов) для дальнейшего обучения. Этот процесс получения изображений с помощью 3D-программы называется рендерингом. В процессе рендеринга скрипт делает автоматическую разметку материалов (боксинг). В результате получается база рендеров и соответствующие каждому рендеру xml файлы с указанием точных координат объекта на изображении, а в нашем случае КМ (рис 3). Использование рендеров автоматизирует и существенно ускоряет процесс подготовки необходимых данных для обучения СРСКМ.



Рисунок 3. Подготовленные и размеченные СКМ

Все полученные материалы виртуальная студия сразу распределяет в две папки: test и train. Папка train будет использоваться для обучения СРСКМ, а в папку test будет попадать каждый пятый рендер, что будет использоваться для самотестирования системы.

Во время обучения система делает так называемые чекпоинты — проверяет насколько правильно идет процесс. Для этого нужен материал, похожий на изучаемый. Такой метод обучения универсален для любых объектов, которые должна будет научиться распознавать система. Этот метод называется обучением СРСКМ на основе синтетических материалов.

Для выполнения работы нами планируется подготовить более сорока тысяч изображений колоний различных видов микроорганизмов, выросших на чашках Петри. Часть СКМ будет выполнена вручную, но большинство — сгенерировано в виртуальной студии в полуавтоматическом режиме.

Если на снимке изображено несколько колоний, то система будет распознавать каждую колонию по отдельности. По завершении работы система будет выдавать информацию о проведенном анализе в виде итоговой таблицы.

На заключительном этапе исследований нами будет подготовлен ряд материалов более низкого качества для обучения. В результате система должна будет пройти ещё 3000000 шагов и в итоге выдать результат, который должен оказаться гораздо менее чувствительным к качеству входящего изображения, а значит и к качеству СКМ.

С помощью СРСКМ станет возможным проводить качественный и количественный анализ микробиоты исследуемого продукта питания. Система будет подсчитывать количество КМ, выросших на чашке Петри, а также сортировать и классифицировать КМ по тем или иным признакам. Полученные результаты будут представлены в виде таблицы.

Заключение.

В результате выполнения данной работы станет возможным использование современных методов интерпретации полученных СКМ в виде таблицы, в которой будут содержаться данные о качественном и количественном составе микробиоты исследуемого продукта питания. Это позволит повысить качество оценки безопасности продуктов питания, устранив влияние человеческого фактора.

Список использованных источников

- [1] Коркин Л.Р., Коркина М.В., Саевич К.Ф., Ящин К.Д., Подготовка материалов для машинного обучения с помощью виртуальной студии. Седьмая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск. – 2021 год.
- [2] Коркин Л.Р. Использование системы распознавания снимков колоний микроорганизмов на чашках Петри для оценки качества продуктов питания. 57-я Научная Конференция Аспирантов, Магистрантов и Студентов БГУИР; Минск. – 2021.
- [3] Хоулт Дж., Криг Н. Определитель бактерий Берджи. В 2-х т, 1997 год
- [4] Тимур Казанцев. Искусственный интеллект и Машинное обучение. Основы программирования на Python. 2020.
- [5] Тонни Гэддис. Начинаем программировать на Python, 4-е издание, 2019.
- [6] Глубокое обучение на Python Автор: Шолле Франсуа Год: 2018 Издательство: Питер ISBN 978-5-4461-0770-4 Страниц: 400.
- [7] Применение нейронных сетей для задач классификации [Электронный ресурс]. – Режим доступа: <https://basegroup.ru/community/articles/classification>(дата обращения 08.05.2018).

PREPARATION OF MATERIALS FOR MACHINE LEARNING USING THE VIRTUAL STUDIO

L. KORKIN

Assistant of the Department of Engineering Psychology and Ergonomics of BSUIR, Master of Engineering and Technology

M. KORKINA

Researcher of The Production and Testing Laboratory, Institute for the Meat and Dairy Industry

K. MAKLAK

An engineer microbiologist of The Production and Testing Laboratory, Institute for the Meat and Dairy Industry

O. MEDVEDEV

Assistant of the Department of Engineering Psychology and Ergonomics of BSUIR, Master of Engineering and Technology

K. SAEVICH

Doctor of Biological Sciences, Professor

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus, Minsk
Institute for the Meat and Dairy Industry, Republic of Belarus, Minsk
E-mail: meat-dairy@tut.by*

Abstract. At present, in food quality control institutions the identification of microorganisms by morphological features is done by The Bergi bacteria determinant. This method is time-consuming and requires a lot of attention and precision of execution. Thus, the development of an automated system for recognizing colonies of microorganisms on Petri dishes for subsequent counting and preliminary identification of microorganisms is an urgent task.

Keywords. Machine learning, artificial intelligence, microbiology, colonies of microorganisms.