

ИССЛЕДОВАНИЕ И РАЗРАБОТКА МЕТОДОВ
СЕГМЕНТАЦИИ И КЛАССИФИКАЦИИ
МИКРОБИОЛОГИЧЕСКИХ ФОТОИЗОБРАЖЕНИЙ

И. В. Машечкин, И. С. Попов, М. А. Казачук, О. Е. Горохов, И. С. Лазухин

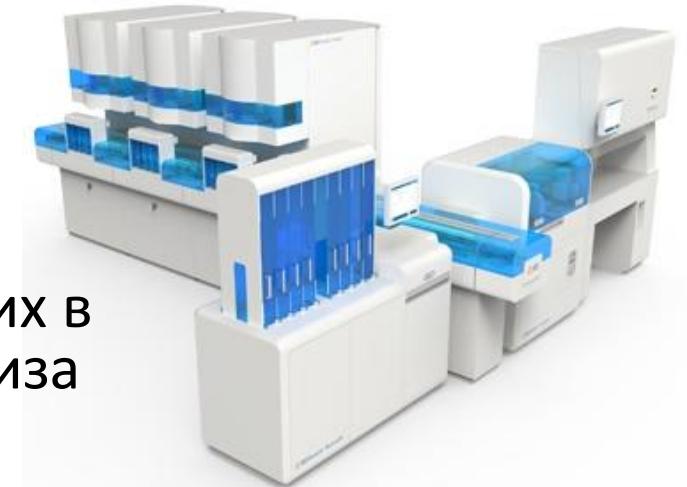
МГУ имени М. В. Ломоносова, Москва, 2025

Введение. Актуальность

Задача выявления вредоносных бактерий (Классический подход):

- Посев биоматериала в питательный раствор (в чашках Петри)
- *Визуальный осмотр чашек и обнаружение колоний сформировавшихся вредоносных микроорганизмов*
- Забор обнаруженных колоний бактерий и помещение их в масс-спектрометр для проведения спектрального анализа
- Выбор лечения в зависимости от полученного вида бактерий

Существующие программно-аппаратные комплексы позволяют полностью автоматизировать все этапы процесса, за исключением экспертного визуального осмотра



Введение. Исследуемые данные

Данные автоматического фотографирования чашек Петри



02:13



11:15



18:18

Введение. Существующие решения

- Зарекомендовавшие себя нейросетевые методы сегментации в медицине^{1,2,4} – One-Stage detector (YOLOv4, YOLOv5), Two-Stage detector (R-CNN, Faster R-CNN, Mask R-CNN), U-net
- Классические методы сегментации для медицинских изображений³ – Fast Marching Method, Multi Threshold, Watershed Segmentation

Проблемы: Недостаточная автоматизация, отсутствие открытых весов, интерпретация результатов, недостаточное качество, открытые объемные наборы данных

¹Kurumisawa T., Kawai K., Shinozuka Y. Verification of a simplified disk diffusion method for antimicrobial susceptibility testing of bovine mastitis isolates // Japanese J. Veterinary Research. 2021. 69. N 2. P. 135–143.

²Naets T. et al. A Mask R-CNN approach to counting bacterial colony forming units in pharmaceutical development // ArXiv preprint arXiv:2103.05337. 2021.

³Torelli A. et al. AutoCellSeg: robust automatic colony forming unit (CFU)/cell analysis using adaptive image segmentation and easy-to-use post-editing techniques // Scientific Reports. 2018. 8. N 1. P. 1–10.

⁴Majchrowska S. et al. AGAR a microbial colony dataset for deep learning detection // ArXiv preprint arXiv:2108.01234. 2021.

Введение. Проблемы данных

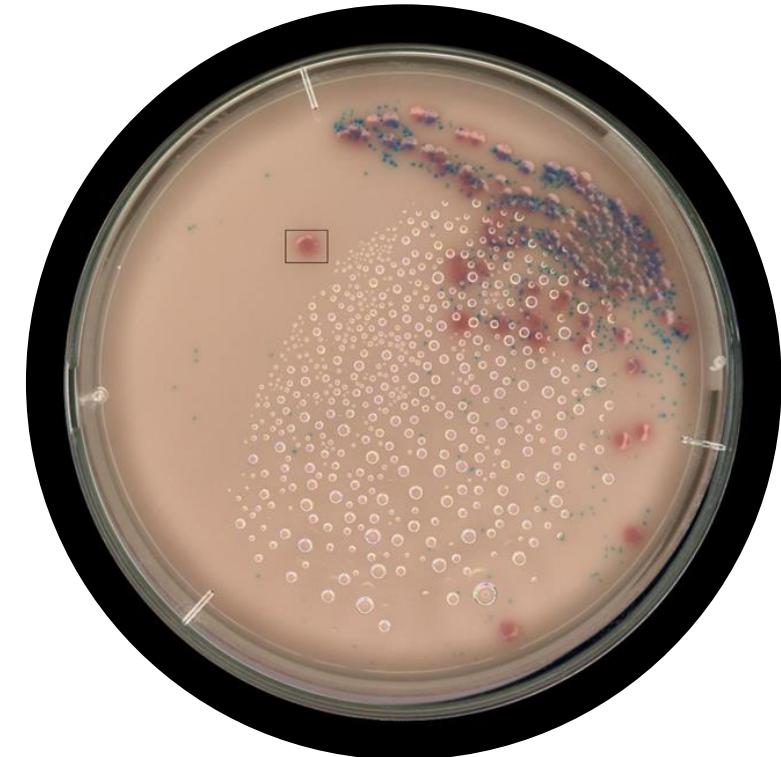
Здесь можно видеть часто встречающиеся артефакты данных, к проблемам так же стоит отнести человеческий фактор при разметке



«Засвет»



«Магнитный шарик» и «Штрих-код»



«Конденсат»

Цели исследования

- *Увеличение общего количества анализируемых образцов (изображений) в единицу времени путем минимизации участия экспертов-микробиологов*
- Уменьшение времени принятия решений по образцу
- Повышение качества обработки образца:
 - а) Для лабораторий без возможности проведения масс-спектрального анализа (*определение достоверной видовой принадлежности микроорганизмов*)
 - б) Повысить скорость и достоверность вынесения вердикта для современных лабораторий, включающих масс-спектральный анализ
- Уменьшение количества ошибок, вызванных человеческим фактором, ошибок, вызванных наличием затрудняющих исследование дефектов изображения (артефактов) и т.п.

Исследуемые данные

Были исследованы исторические данные по Москве и Московской области за 1.5 года – около 31 тыс. описаний и результатов проведения микробиологических исследований:

- Серии фотоизображений чашек Петри каждого образца (2000×2000)
- Данные об экспертной разметке регионов
- Итоговые результаты масс-спектрального анализа, характеристики исходного биоматериала, среды и пр.
- Топ-3 колоний по объему данных: *E.coli*, *Ent.faecalis*, *Kl.pne-ss-pne*

Проблема: Отсутствие возможности разметить изображения целиком! Проблемы с использованием классических метрик сегментации (mAP)

Экспертная сегментация изображений

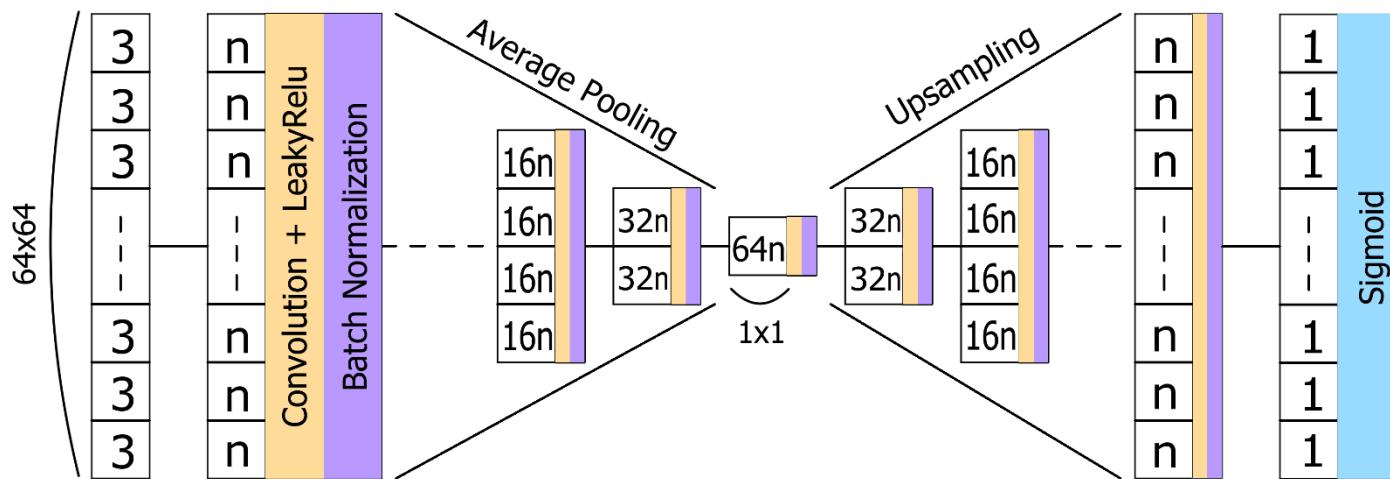
- Медицинский персонал обычно выделяет на изображениях прямоугольные регионы, содержащие потенциально интересные изолированные колонии бактерий, предпочтительно одного типа.
- Информация с координатами регионов, ожидающих исследования, используется для забора биоматериала из конкретного региона чашки Петри и дальнейшего аппаратного исследования каждого из них по-отдельности – для точного определения типа бактерии.
- Исследуемая врачебная разметка может содержать как регионы, содержащие индивидуальные колонии, так и сегменты значительного размера, содержащие несколько колоний бактерий как одного так и разных типов, что может препятствовать успешному построению моделей машинного обучения.
- Стоит так же отметить наличие ошибок, вызванных человеческим фактором, например, ошибочных меток регионов, что требует экспертных знаний для исправления

Построение решения. Сегментация (1)

1. Применение медианного фильтра
2. Оконная параллельная обработка изображения **предложенной нейросетью класса энкодер-декодер** для выделения бинарной маски объектов
3. Алгоритм марширующих кубов для обнаружения грубых сегментов-регионов, содержащих колонии бактерий
4. Отбор **регионов, наиболее пригодных** для исследования

Построение решения. Сегментация (2)

Предложенная архитектура – преобразует входное трехканальное изображение в одноканальную маску:



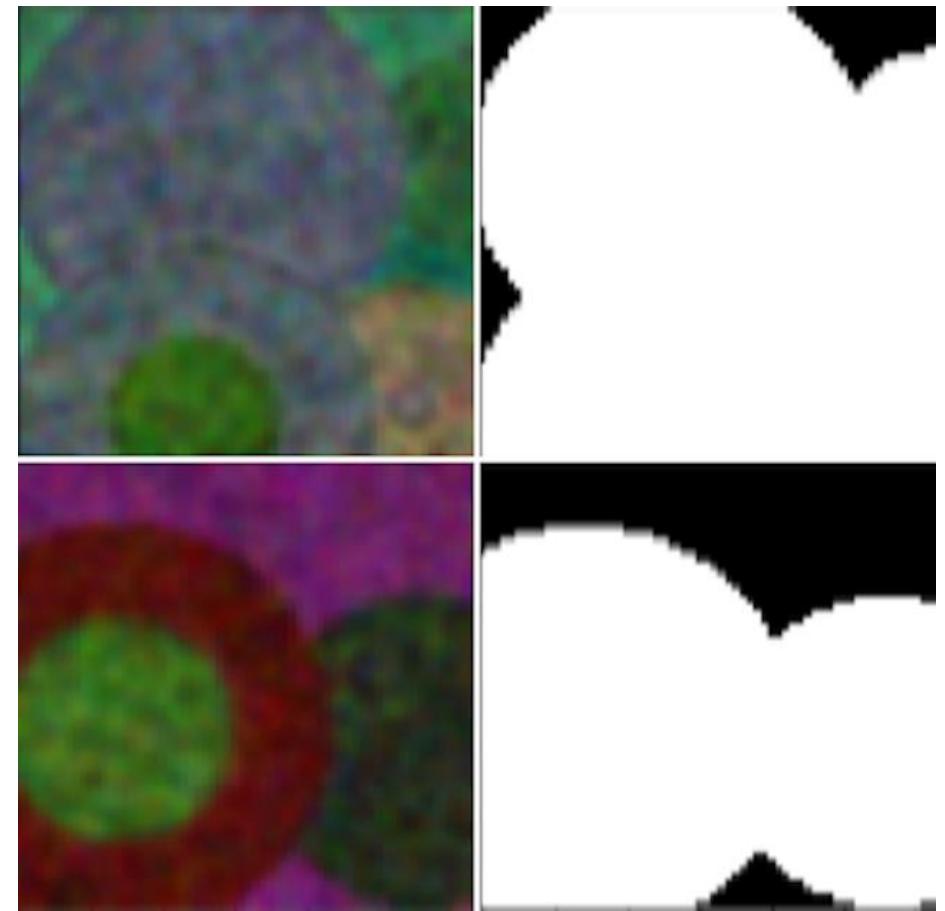
Проблема обучения нейросети для сегментации – колонии на изображении невозможно разметить полностью

Решение: Генерация искусственных данных для обучения

Построение решения. Сегментация (3)

Предложенная процедурная генерация искусственных данных для обучения:

- Размер данных соответствует размеру входа сети (64×64)
- Случайный выбор цвета, текстуры «питательной среды»
- Диапазон возможного количества «колоний», что случайным образом размещаются в сегменте
- Случайные цвет, структура и текстура колоний
- Сглаживание полученного сегмента фильтром Гаусса со случайно выбранным σ
- Одновременно генерируется соответствующая маска сегмента



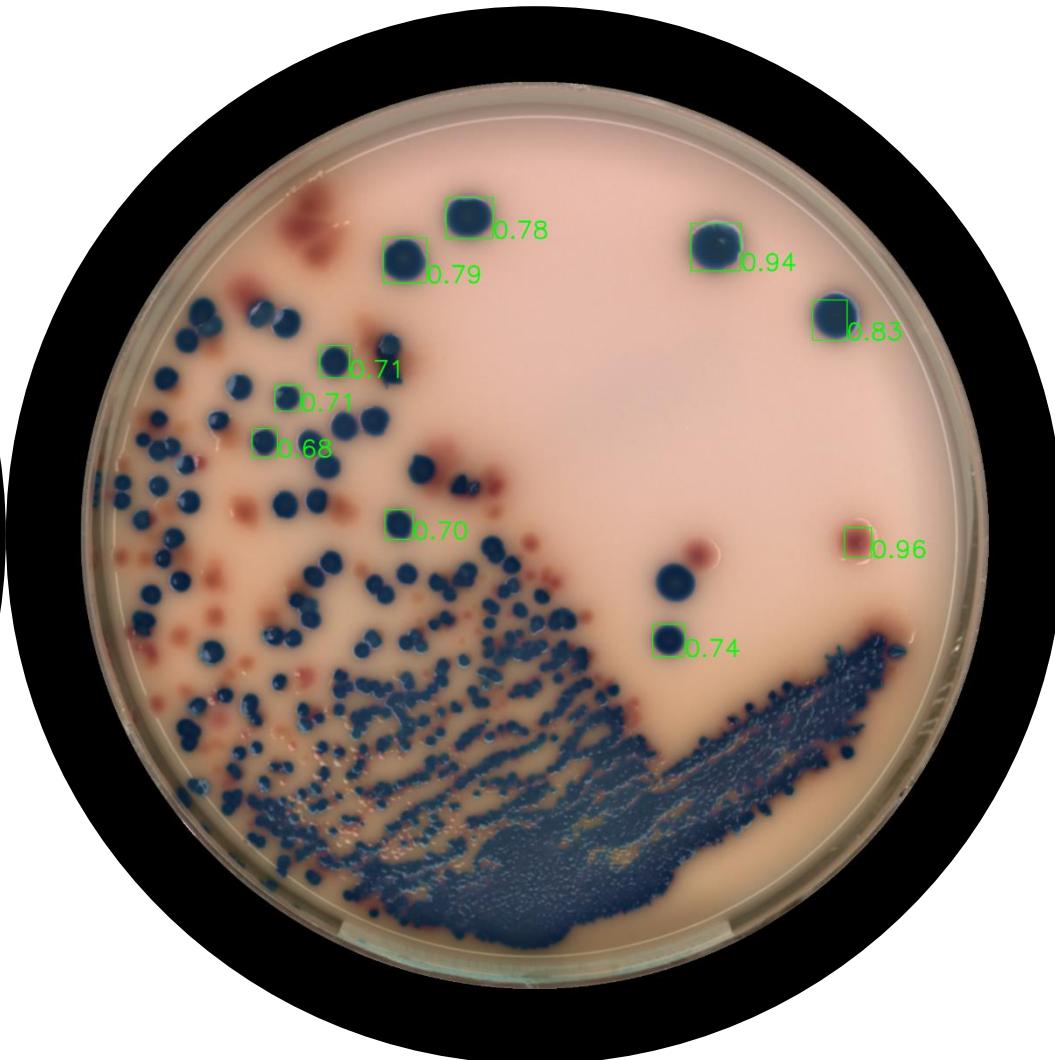
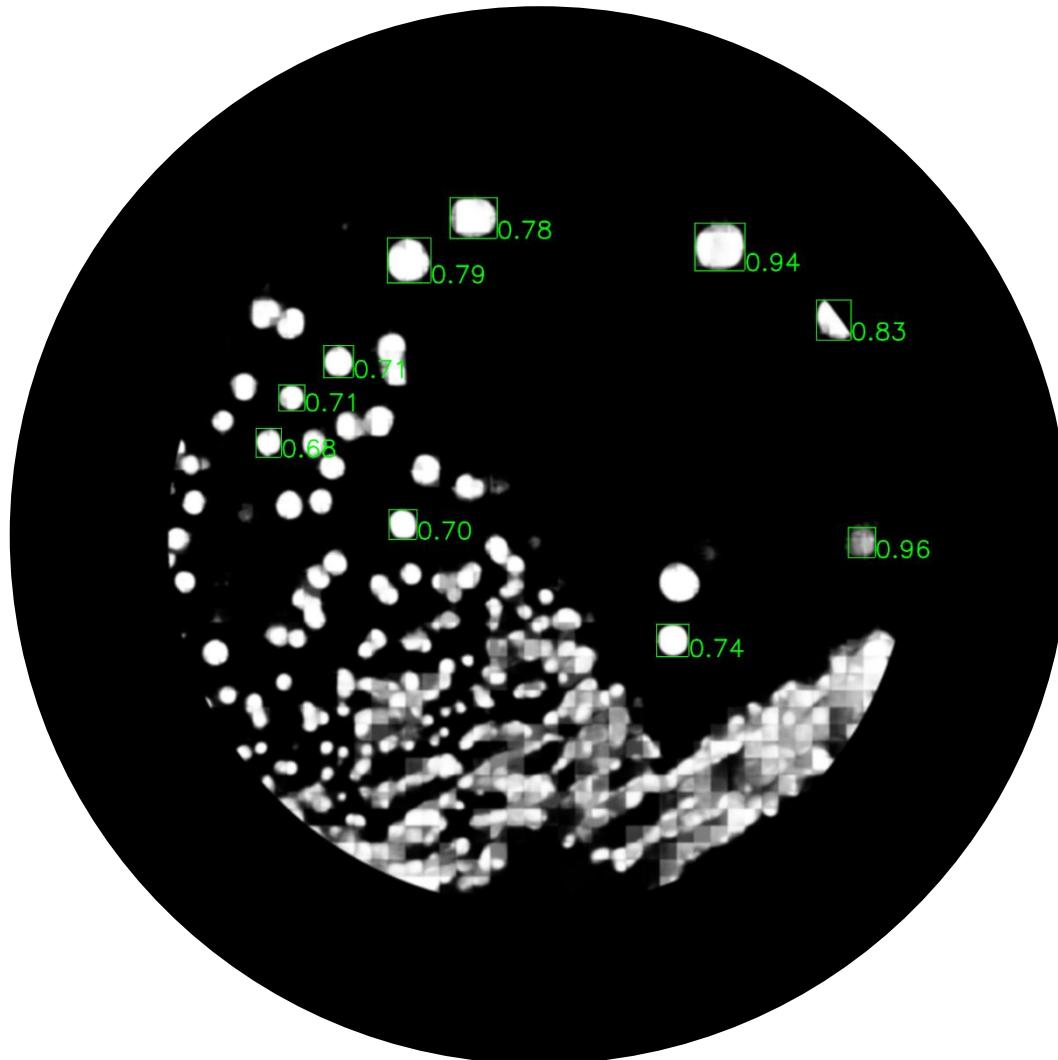
Построение решения. Сегментация (4)

Отбор регионов, наиболее *пригодных для исследования*:

1. Маска изображения размывается по Гауссу (постобработка результата сети)
2. По заданному порогу яркости выделяются контуры и им соответствующие прямоугольные сегменты-регионы R_i
3. Отбрасываются регионы, не соответствующие тривиальным критериям размера
4. Степень подобия окружности R_i : $c_i = 1 - \frac{std(dist)}{mean(dist)}$
dist соответствует множеству расстояний от точек региона до его центра
5. Локальная обособленность региона p_i – доля нулевых значений маски в выбранной малой окрестности контура (также отсекаем регионы, замыкающие невыпуклые контуры)
6. Глобальная обособленность региона d_i как нормализованное минимальное значение в матрице расстояний между центрами всех найденных регионов
7. На основе финальной метрики $score(R_i) = w_c c_i + w_p p_i + w_d d_i$ производится построение ранжированного списка найденных регионов

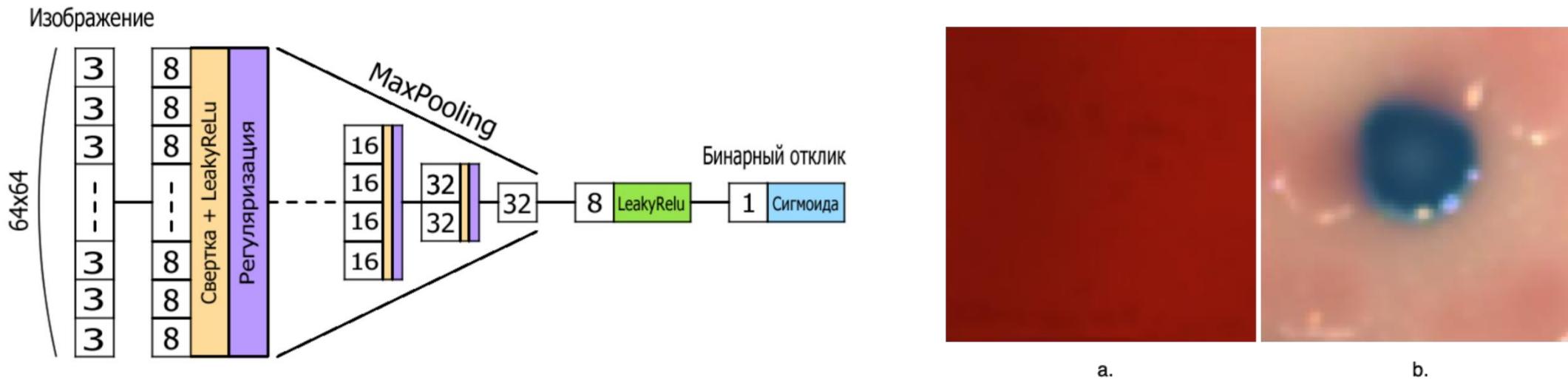
Задача обнаружения роста на основе предложенной метрики – определение порога

Построение решения. Сегментация (5)



Построение решения. Классификация (1)

- Классификация как использование *ансамбля независимых бинарных моделей* – масштабируемость, независимость, ожидаемая реакция для неизвестных классов объектов
- Обучение: размеченные данные фиксированной колонии против размеченных данных колоний других видов
- Проблемы: Недостаток размеченных данных по конкретным колониям, ошибки в врачебной разметке (см. ниже, справа)
- В связи с этим в качестве основной модели была предложена легковесная сверточная архитектура (см. ниже, слева)

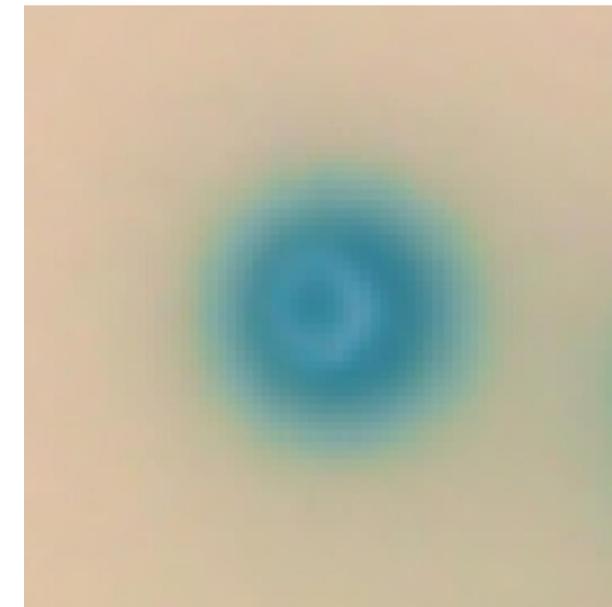


Построение решения. Классификация (2)

Пример классификации (в ансамбле участвуют 3 модели):

<i>E.Coli</i>	<i>Ent.faecalis</i>	<i>Kl.pne-ss-pne</i>
0.024	0.881	0.001

Результат классификации – *Ent.Faecalis*



Экспериментальное исследование (1)

В рамках задачи сегментации было исследовано четыре алгоритма:

- R-CNN (обучен на сгенерированных данных)
- OnePetri (YOLOv5)
- Auto Cell Segmentation*
- Предложенный метод сегментации (обозн. GenCNN) – порог 0.5

Постановка эксперимента:

- Сегментация заданного микробиологического изображения
- Оценка роста на изображении на основе региона с наибольшим доверием

Набор данных:

- Фиксированные параметры съемки
- 5200 изображений чашек Петри, обработанных масс-спектрометром – «рост обнаружен»
- 6200 изображений чашек Петри, отмеченных «роста нет»

*Модель AutoCellSeg является полуавтоматической – необходима предварительная экспертная разметка регионов, потому для ее оценки использовался поднабор, содержащий по 100 изображений каждого класса.

Экспериментальное исследование (2)

Набор данных задачи классификации:

- Фиксированные параметры съемки
- Виды колоний бактерий: *E.coli*, *Ent.faecalis*, *Kl.pne-ss-pne*
- Тренировочных образцов (соответственно): 750, 1250, 1000
- В качестве несоответствующих классу меток также использовались остальные виды колоний с меньшим объемом данных (суммарно 2000)

Экспериментальное исследование (3)

Метрика	RCNN	OnePetri (YOLO)	AutoCellSeg*	GenCNN
Accuracy	0.584	0.432	0.449	0.897
+Precision	0.981	0.101	0.889	0.952
+Recall	0.341	0.910	0.294	0.879
-Precision	0.474	0.983	0.306	0.822
-Recall	0.989	0.390	0.895	0.927
ROC-AUC	0.665	0.700	0.594	0.926

Вид	Точность	Полнота	Точность	ROC-AUC	Образцов
E.coli	0.952	0.882	0.970	0.978	150
Ent.faecalis	0.841	0.932	0.950	0.980	250
Kl.pne-ss-pne	0.716	0.773	0.931	0.935	200

*Модель AutoCellSeg является полуавтоматической – необходима предварительная экспертная разметка регионов, потому для ее оценки использовался поднабор, содержащий по 100 изображений каждого класса.

Выводы

Научная новизна работы заключается в

- Разработанном нейросетевом алгоритме построения бинарной маски объектов на изображениях чашек Петри на основе процедурно сгенерированных данных
- Предложенной процедуре отбора регионов, наиболее пригодных для исследования

Данные алгоритмы обладают интерпретируемыми результатами работы и способны значительно уменьшить участие человека в процессе проведения микробиологического исследования

Заключение

- Были предложены новые интеллектуальные технологии обнаружения и классификации колоний микроорганизмов на микробиологических фотоизображениях чашек Петри
- Предложенные алгоритмы обладают интерпретируемыми результатами работы и уменьшают необходимость участия человека в процессе проведения микробиологического исследования
- Для оценки качества работы предлагаемых методов была исследована независимая выборка данных
- По результатам экспериментов, предложенные алгоритмы превосходят по качеству существующие решения по данной тематике и могут быть использованы в определенных ситуациях как в качестве полностью автоматического средства, так и в качестве средства поддержки принятия решения экспертом-микробиологом в автоматизированном режиме

Публикации

- Bacterial colony detection method for microbiological photographic images / O. Gorokhov, R. Fazylov, M. Kazachuk et al. // 2022 International Joint Conference on Neural Networks (IJCNN). — IEEE: 2022. — P. 1–8.
- Применение технологий искусственного интеллекта для обнаружения и классификации колоний микроорганизмов при проведении микробиологических анализов / О. Е. Горохов, М. А. Казачук, И. С. Лазухин и др. // *Высокотехнологическая медицина*. — 2023. — Т. 10, № 1. — С. 28–38.
- Интеллектуальные технологии сегментации и классификации микробиологических фотоизображений / О. Е. Горохов, М. А. Казачук, И. С. Лазухин и др. // *Вестник Московского университета. Серия 15: Вычислительная математика и кибернетика*. — 2023. — № 4. — С. 21–32.
- Intelligent technologies for the segmentation and classification of microbiological photographic images / O. Gorokhov, M. Kazachuk, I. Lazukhin et al. // *Moscow University Computational Mathematics and Cybernetics*. — 2023. — Vol. 47, no. 4. — P. 201–210.