

## АВТОМАТИЧЕСКАЯ ДЕТЕКЦИЯ И ПОДСЧЕТ МИКРОКОККОВ НА МИКРОСКОПИЧЕСКИХ СНИМКАХ С ИСПОЛЬЗОВАНИЕМ МУЛЬТИМОДАЛЬНЫХ МОДЕЛЕЙ

**Храпков А. В.<sup>1</sup>**

**Научный руководитель – канд. физ. мат. наук, доцент Самарин А. В.<sup>1</sup>**

<sup>1</sup>Университет ИТМО

andrewhrapkov@yandex.ru

### **Введение**

Микроскопический анализ является фундаментальным этапом в медицинской диагностике, экологии и контроле качества в пищевой и фармацевтической промышленности. В данной работе было проведено исследование методов детекции и подсчета микрококков [1] на микроскопических снимках. Классические детекторы, такие как YOLO [2], эффективны для микроскопии [3], но требуют значительного набора предварительно размеченных данных. В качестве альтернативного подхода были исследованы возможности мультимодальной модели Grounding DINO [4] и подобраны оптимальные параметры инференса, а также различные подходы к улучшению качества. Цель работы — оценить применимость этой модели для zero-shot детекции микрококков и сформировать практический пайплайн обработки снимков.

### **Основная часть**

Для оценки предложенного подхода вручную были размечены 10 микроскопических снимков разрешением  $3664 \times 2748$  пикселей. Эти снимки отличаются количеством бактерий (от нескольких бактерий до нескольких сотен), резкостью и контрастом, а также обладают неоднородным фоном. Сами объекты детекции маленького размера, иногда могут перекрываться.

Из-за ограничения по входному размеру модели прямой анализ кадра приводит к потере мелких объектов и большой ошибке подсчета. В качестве решения проблемы предлагается нарезка изображения на перекрывающиеся фрагменты (slicing), с последующей агрегацией результатов, также были протестированы различные размеры этой нарезки, и влияние этих параметров на финальные метрики. Помимо этого, были исследованы различные предобработки, увеличивающие контраст и различимость клеток, которые применялись перед инференсом модели. Лучший результат показал алгоритм адаптивного выравнивания гистограммы CLAHE [5].

В качестве входа модель принимает изображение и текстовый промпт. Были исследованы различные промпты, а также параметр box threshold, который отвечает за порог уверенности модели. Для постобработки использовался метод NMS (Non-Maximum Suppression) [6], который позволил удалить дублирующиеся рамки на стыках фрагментов.

Все приведенные параметры показали различное влияние на качество детекции, которое оценивалось по метрикам mAP@0.5, Precision, Recall, F1, и по ошибке подсчета: MAE, MARE. Было установлено, что повышение локального контраста, нарезка на фрагменты размером  $500 \times 500$  пикселей, использование промптов с описанием визуального признака, таких как “round bacteria” позволили существенно улучшить метрики, по сравнению с базовыми настройками модели. Кроме этого, был подобран оптимальный с точки зрения Precision/Recall порог уверенности box threshold.

### **Выводы**

Zero-shot подход на базе Grounding DINO применим как инструмент предварительной детекции при тщательном подборе предобработки, параметров инференса и постобработки. Такие методы, как нарезка изображения (slicing), метод CLANE, а также аккуратный выбор порога уверенности позволяют значительно повысить качество детекции и ощутимо снизить ошибку подсчета. В дальнейшем планируется расширение датасета, проверка устойчивости на разных условиях съемки, а также дообучение текущей модели и/или классических детекторов, (например, YOLO), на размеченных данных.

### Литература

1. Морфология бактерий [Электронный ресурс]. — URL: [https://www.allvet.ru/knowledge\\_base/microbiology/morfologiya-bakteriy/](https://www.allvet.ru/knowledge_base/microbiology/morfologiya-bakteriy/)(дата обращения: 26.02.2026).
2. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection [Электронный ресурс]. — arXiv:1506.02640v5, 2016. — URL: <https://arxiv.org/abs/1506.02640> (дата обращения: 26.02.2026).
3. López Flórez S., González-Briones A., Hernández G., Ramos C., de la Prieta F. Automatic Cell Counting With YOLOv5: A Fluorescence Microscopy Approach // International Journal of Interactive Multimedia and Artificial Intelligence. 2023. Vol. 8, № 3. P. 64–71. DOI: 10.9781/ijimai.2023.08.001.
4. Liu S., Zeng Z., Ren T., Li F., Zhang H., Yang J., Jiang Q., Li C., Yang J., Su H., Zhu J., Zhang L. Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection [Электронный ресурс]. — arXiv:2303.05499, 2023. — URL: <https://arxiv.org/abs/2303.05499> (дата обращения: 26.02.2026).
5. Namazi Nia S., Shih F. Y. Medical X-Ray Image Enhancement Using Global Contrast-Limited Adaptive Histogram Equalization // International Journal of Pattern Recognition and Artificial Intelligence. 2024. Vol. 38, № 12. Art. 2457010. DOI: 10.1142/S0218001424570106.
6. Hosang J., Benenson R., Schiele B. Learning Non-Maximum Suppression [Электронный ресурс]. — arXiv:1705.02950, 2017. — URL: <https://arxiv.org/abs/1705.02950> (дата обращения: 26.02.2026).