

АЛГОРИТМ ПОДГОТОВКИ ФРАГМЕНТОВ ПОЛНОСЛАЙДОВЫХ ИЗОБРАЖЕНИЙ

Косарева А.А.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь*

*Научные руководители: Ковалев В.А. – канд.техн.наук, доцент, ведущий научный сотрудник ОИПИ НАН
Беларуси, Камлач П.В. – канд.техн.наук, доцент, доцент кафедры ЭТТ, зам. декана ФКП*

Аннотация. Подготовка изображений для обучения нейронных сетей важный и долгий процесс, ошибки которого влияют на результаты обучения и увеличение времени разработки. При работе с полнослайдовыми изображениями существует несколько проблем, обусловленных спецификой построения данных изображений и их размерами. Разработанный алгоритм подготовки фрагментов изображений позволит оптимизировать процесс создания набора данных.

Ключевые слова: подготовка данных, выделение фрагментов изображений, полнослайдовые изображения, нормализация изображений, обучение нейронных сетей.

Введение. Полнослайдовые изображения (WSI, whole slide imaging) – виртуальная микроскопия, предназначенная для удобного хранения и анализа гистопатологического или цитологического препарата. WSI позволяют медицинским специалистам работать с изображением с использованием информационных технологии, имитируя процесс обращения специалиста с микроскопом [1]. Полнослайдовые изображения можно использовать при совместном обсуждении клинических случаев специалистами или обучающимися, без привязки к лабораторным исследованиям и физическому расположению препаратов.

Особенности построения полнослайдовых изображений также позволяют использовать их при разработке интеллектуальных систем. Важно понимать, что размер таких изображений может достигать $300\,000 \times 100\,000$ пикселей, что значительно усложняет работу. Кроме того, зачастую разные анатомические признаки просматриваются на разных увеличениях, соответственно возникает необходимость выделения отдельных фрагментов изображения, определенных поставленной задачей.

Основная часть. На рисунке 1 представлен разработанный алгоритм подготовки фрагментов полнослайдовых изображений.

Для начала необходимо удалить все белые области вне препарата с изображения, так они не несут в себе никакой диагностической информации. Для этого были определены средние значения пикселей полностью белых фрагментов, частично белых фрагментов с наличием артефактов изображений, нормальных фрагментов, содержащих в себе изображения препарата. В таблице 1 представлены средние значения и стандартные отклонения для всех выделенных категорий фрагментов:

Таблица 1 – Среднее значение яркости пикселей фрагментов разного содержания

Содержание фрагмента	Среднее значение яркости пикселей фрагмента	Стандартное отклонение яркости пикселей фрагмента
Белый фон	243,15	4,23
Белый фон с артефактами	233,98	20,9
Фрагмент препарата	185,12	32,46

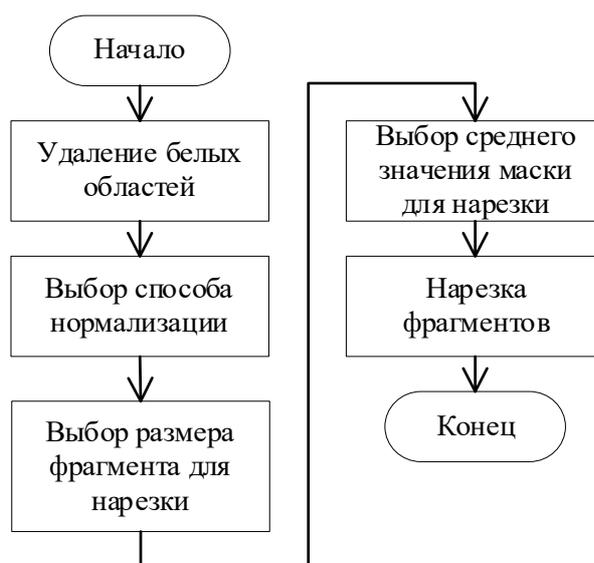


Рисунок 1 – Алгоритм подготовки фрагментов полнослайдовых изображений

Следующим шагом является выбор способа нормализации гистопатологического среза. Исследования показали, что одним из самых оптимальных вариантов является нормализация с использованием оптической плотности изображения [2,3]. На рисунке 2 представлены два фрагмента изображения. Исходное и изображение, переведённое в значения оптической плотности.

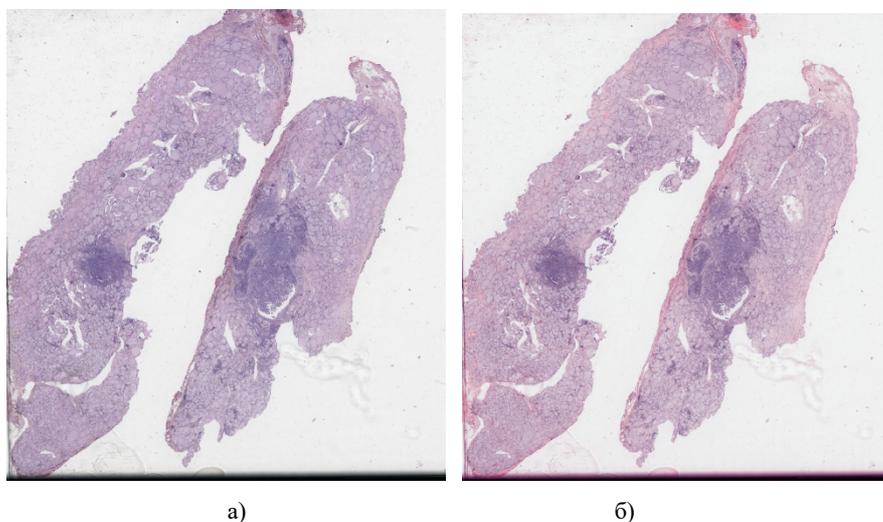


Рисунок 2 – Фрагмент гистопатологического изображения а) исходный, б) с рассчитанной оптической плотностью.

Подобный способ перевода изображения из RGB-формата в формат оптической плотности позволяет лучше визуализировать различия плотности изображенного вещества. Данное изображение также можно использовать для искусственного выделения двух каналов окрашивания (гематоксилин и эозин) [4].

Следующий этап предлагаемого алгоритма - выбор представительного размера фрагмента изображения для нарезки. Этот процесс требует отдельного рассмотрения, так как при разном размере представительного фрагмента на изображении просматриваются разные структуры. Выбор представительного размера зависит от поставленной задачи и необходимого смыслового наполнения фрагмента.

После выбора размера, необходимо определить среднее значение яркости пикселей маски в области фрагмента при нарезке. В нашем исследовании при работе с фрагментами

размера 2048×2048 пикселей среднее значение маски размеченного класса выбиралось равным 64, а при работе с фрагментами 1024×1024 пикселей равным 128. Увеличение среднего значения яркости пикселей маски в области фрагмента при нарезке улучшало смысловое наполнение фрагмента, но значительно влияло на их количество. Следовательно, при ограниченном количестве исходных полнослайдовых изображений разных пациентов выбор среднего значения яркости пикселей маски в области фрагмента следует осуществлять, соблюдая баланс между количеством нарезанных фрагментов и их содержанием.

Последний этап алгоритма, нарезка фрагментов изображения, была реализована программно с использованием библиотеки *OpenSlide*, позволяющей считывать полнослайдовые изображения на разных увеличениях, а также выделять отдельные их регионы [5,6].

Заключение. В данном докладе был представлен алгоритм подготовки фрагментов полнослайдового изображения. Рассматривались вопросы нормализации изображений, размера фрагментов, выбора среднего значения яркости пикселей маски в области фрагмента. Были определены средние значения и стандартные отклонения яркости пикселей полностью и частично белых фрагментов. Рассчитанные значения позволят ускорить процесс подготовки набора данных, при включении их в условия отсеивания лишних белых областей фона при нарезке.

Правильная подготовка данных важна для дальнейшего автоматического анализа изображения и разработки интеллектуальных систем. Следующим этапом исследования планируется проведение экспериментов по определению оптимального размера фрагментов гистопатологического изображения и разработка системы поиска схожих фрагментов, содержащих признаки наличия раковой опухоли.

Список литературы

1. Kumar N, Gupta R, Gupta S. Whole Slide Imaging (WSI) in Pathology: Current Perspectives and Future Directions. *J Digit Imaging*. 2020 Aug;33(4):1034-1040. doi: 10.1007/s10278-020-00351-z. PMID: 32468487; PMCID: PMC7522141.2.
2. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Schmitt C, et al. A method for normalizing histology slides for quantitative analysis. 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. p.1107–1110, 2009.
3. Drozdal, Michal & Chartrand, Gabriel & Vorontsov, Eugene & Jorio, Lisa & Tang, An & Romero, Adriana & Bengio, Y. & Pal, Chris & Kadoury, Samuel & Shakeri, Mahsa. (2017). Learning Normalized Inputs for Iterative Estimation in Medical Image Segmentation. *Medical Image Analysis*. 44. 10.1016/j.media.2017.11.005.
4. Alsubaie N, Trahearn N, Raza SEA, Snead D, Rajpoot NM (2017) Stain Deconvolution Using Statistical Analysis of Multi-Resolution Stain Colour Representation. *PLoS ONE* 12(1): e0169875. <https://doi.org/10.1371/journal.pone.0169875>
5. OpenSlide: A Vendor-Neutral Software Foundation for Digital Pathology Adam Goode, Benjamin Gilbert, Jan Harkes, Drazen Jukic, M. Satyanarayanan *Journal of Pathology Informatics* 2013, 4:27
6. A Vendor-Neutral Library and Viewer for Whole-Slide Images Adam Goode, M. Satyanarayanan *Technical Report CMU-CS-08-136, June 2008 Computer Science Department, Carnegie Mellon University*

UDC 528.854.2

ALGORITHM FOR THE PREPARATION OF FRAGMENTS OF WHOLE-SLIDE IMAGING

Kosareva A.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Kovalev V.A. – PhD, leading researcher of United Institute of Informatics Problems

Kamlach P.V. – PhD, associate professor, associate professor of the Department of ETT, Deputy Dean of FCD

Annotation. Preparation of images for teaching neural networks is an important and long process, the mistakes of which affect the learning outcomes and increase the development time. When working with whole-slide imaging, there are several problems due to the specifics of constructing these images and their sizes. The developed algorithm for the preparation of images will allow you to optimize the process of creating a set of data.

Keywords: data preparation, selecting images, whole-slide imaging, normalization of images, training of neural networks.