Development of automated computer vision methods for cell counting and endometrial gland detection for medical images processing

Abstract. Current work is focused on the processing of medical images obtained by performing a pathomorphological analysis of preparation. The algorithms for processing images of nuclei of light and confocal microscopy and tissue of light microscopy were considered in particular. The application of the proposed algorithms and software for detecting pathologies was justified.

Keywords: computer vision; image processing; digital pathology; detection and classification; cell and tissue nuclei; light microscopy; confocal microscopy

For citation. Sergeev D.I., Andreev A.E., Drobitseva A.O., Cenevska S., Kukavica N., Drobitsev P.D. Development of automated computer vision methods for cell counting and endometrial gland detection for medical images processing. Trudy ISP RAN/Proc. ISP RAS, vol. 32, issue 3, 2020, pp. 119-130. DOI: 10.15514/ISPRAS–2020–32(3)–11

1 D.I. Sergeev, ORCID: 0000-0003-2503-6272 <densrv1@gmail.com>
2 A.E. Andreev, ORCID: 0000-0003-3343-2937 <alexander597@gmail.ru>
3 A.O. Drobitseva, ORCID: 0000-0002-6833-6243 <anna.drobitseva@gmail.com>
1 S. Cenevska, ORCID: 0000-0002-2272-8882 <slobodankacenevskay@yahoo.com>
1 N. Kukavica, ORCID: 0000-0001-6477-357X <nikola.kukavica.94@gmail.com>
1 P.D. Drobitsev, ORCID: 0000-0003-1116-7765 <drob@ics2.ecd.spbstu.ru>
1 Peter the Great St.Petersburg Polytechnical University, 29, Polytechnicheskaya st., St.Petersburg, 195251, Russia
2 The Research Institute of Obstetrics, Gynecology and Reproductology named after D.O. Ott, 3, Mendelevskaya line, Saint Petersburg, 199034
3 St.Petersburg State Pediatric Medical University 2 Litovskaya st., St. Petersburg, 194100, Russia

1. Introduction

The processing of medical images is an extremely important issue for biology and medicine. Pathomorphologists have to process hundreds of images of preparations per day. Their work can be automated due to computer analysis.

Medical image processing can be performed in semi-automatic and automatic modes. Semi-automatic mode is based on manual adjustment of simple, intuitive parameters for evaluating single microphotographs. Automatic mode otherwise does not require both direct operator intervention and initial settings in the processing of preparations.

Modern experts in the field of pathomorphology have an access to a wide range of technologies that make it possible to carry out various measurements depending to required tasks. A striking example of such technologies is the universal ImageJ software [1], which is capable for performing operations aimed at evaluating the geometry and color gamut of the resulting images. The main idea of this software is to write macroses that require a minimum understanding of computer technology from a specialist. This approach makes ImageJ flexible, but not user friendly. More intuitive tools are commercial software such as VideoTestMorphology [2, 3] and ImageProPlus [4]. There are various microscopic solutions. The most popular software at the moment are compared in Table 1.

Table 1. Software features

| Software | Language | Automatic mode | Machine learning | Coding | Open source |
|----------|----------|----------------|------------------|--------|-------------|
| Image J [1] | Java | macros | plugin | no | no |

DOI: 10.15514/ISPRAS–2020–32(3)–11
Сергеев Д.И., Андреев А.Е., Дробинцева А.О., Ценевска С., Кукавица Н., Дробинцев П.Д. Разработка автоматизированных алгоритмов компьютерного зрения для обработки медицинских изображений. Труды ИСП РАН, том 32, вып. 3, 2020 г., стр. 119-130

Таблица 1

| Название программного обеспечения   | Платформа     | Макросы | Ручные | Программы, требующие биоинформатики | Программы, требующие биоинформатики для пользователей без доступа к большим данным |
|-----------------------------------|--------------|---------|--------|-------------------------------------|----------------------------------|
| Cell profiler [5]                 | Python       | no      | yes    | no                                  | no                               |
| Orbit Image Analysis [6]          | Java, python | yes     | yes    | no                                  | yes                              |
| Axio vision 4.8 [7]               | -            | macros  | yes    | no                                  | no                               |
| Video test morphology 5.2 [3]     | -            | macros  | yes    | no                                  | no                               |
| CellSens [8]                      | -            | macros  | no     | no                                  | no                               |
| IMAGE-PRO-Premium [4]             | .NET         | yes     | yes    | no                                  | no                               |
| BioVision                         | -            | yes     | yes    | no                                  | no                               |

Таким образом, существует много современных программ, предназначенных для морфометрии. Однако на практике существуют ряд фактов, которые значительно ограничивают энтузиастов в выборе своих инструментов. Во-первых, большинство бесплатных и полуавтоматизированных программ являются неполными и требуют биоинформатики в команде. Во-вторых, уже существующие и полуавтоматизированные программы в большинстве случаев универсальные, что означает отсутствие учета параметров микроскопа, типа ткани и освещения в процессе работы. Однако, наиболее конвейерные и удобные программы для пользователей с низкими требованиями по цене не доступны в малых лабораториях.

**3. Popular algorithms**

Разработка программного обеспечения включает в себя следующие этапы:

1. **Image pre-processing**

   Это измерение условий, которые увеличивают эффективность и качество идентификации и определения клеток в медицинских препаратах. Включает в себя операции обработки, такие как преобразование, анализ, анализ клеточных параметров, и т. д. [9]. Хотя существует не так много методов, которые способны в полной мере учесть все особенности ими создаваемые препятствия, они играют значительную роль в биологической съемке.

2. **Typical tasks**

   2.1 **Processing of cell and tissue preparations**

   Процесс обработки препаратов включает в себя обработку клеточных и тканевых уровней. В обоих случаях, клетки обычно содержат в большинстве своем специфические для них белки и молекулы, которые они используют для реакции с маркерами. В этой работе, были исследованы маркеры, такие как эстроген рецептор (ER) и прогестерон (PR). Затем, были применены механизмы, моделирование, такт анализ, и т.д. [9]. Хотя есть методы, которые позволяют учитывать все особенности клеток, они играют значительную роль в биологической съемке.

   **3.2 Detection of objects of interest**

   В этой стадии, X и Y координаты центра интересующего объекта (нуклеуса) определяются. В результате, для каждого объекта, которые являются интересующими, могут быть определены. Это позволяет учитывать различные центры интереса, которые могут быть определены. Влияние реальной оси не является возможным, но при этом в целом подобные алгоритмы включают в себя активный контурный алгоритм, алгоритм, позволяющий измерять классы, и т.д. [12].

   **3.3 Selection of characteristics of objects of interest, classification and arrival at a decision**

   В заключении, при разработке алгоритмов, которые позволяют учитывать различные центры интереса, могут быть применены для цветового нуклеуса и их цветов. Это позволяет автоматически определять количество маркеров и их цвета. Это очень важный этап для патоморфологов.
3.4 Methods for the separation of overlapping nuclei

A specific task is the separation of overlapping nuclei. This can be caused by cell division, the camera’s viewing angle in the process of shooting the preparation, and also, the location of the nuclei on top of each other in the depth of the examined tissue. The following approaches can be used to separate the fused nuclei and accurately determine their number:

- the method of active contours with the preliminary use of erosion;
- classifiers (convolutional neural network), previously trained in classes that determine the number of cells in the area of a given size;
- watersheds algorithms;
- segmentation algorithms focused on topological features of objects [14].

4. Image dataset

Many images of cell structures and tissues preparations of light and confocal microscopy were collected and labeled (see Table 2).

Material and equipment for shooting images was provided by the Institute of Obstetrics, Gynecology and Reproductology Ott. The shooting of individual classes of images was carried out with a fixed scale of the microscope. Image preparations of various types of tissues with different lighting conditions and marker colors were collected and labelled.

To accelerate nuclei and glands labelling on images, a software was developed that allows a specialist to set a marker on an object of interest in the image. Subsequently, the coordinates of the centers of these markers (x, y), as well as the length (Size Y) and width (Size X) were recorded in a csv file (see Table 3). Thus, a numerical data of the location and shape of the investigated structures were obtained. The markup was carried out by an employee from the laboratory by a cell biologist at the Ott Research Institute.

### Table 2. Collected images

| Confocal image class                  | Number of images |
|--------------------------------------|------------------|
| Confocal microscopy cell preparations| 30               |
| Confocal microscopy tissue preparations| 92              |
| Lightin microscopy cell preparations  | 100              |

### Table 3. A dataset part for the cells

| Image name | sizeX | sizeY | x   | y   |
|------------|-------|-------|-----|-----|
| 1(5).jpg   | 963   | 963   | 278 | 100 |
| 1(5).jpg   | 963   | 963   | 441 | 201 |
| 1(5).jpg   | 963   | 963   | 795 | 246 |

5. Formulation of the problem

The aim of the work is the research and development of the following algorithms:

- the cell nuclei number estimation with and without researched marker expression on light microscopy image preparations;
- the cell nuclei number estimation with and without researched marker expression on confocal microscopy image preparations;
- highlighting the internal and external glands borders in the confocal microscopy image preparations.

For all algorithms, the following requirements are established:

- work without an operator;
- resistance to changes in the brightness of the preparation;
- resistance to various colors of markers;
- image scale is an input parameter of the algorithms.

6. Suggested algorithms

The algorithms were developed in the PyCharm environment in Python 3.7 using the OpenCV-python 4.0.0.21 library. The source code of developed algorithms is available on github.com [15].

6.1 Counting the number of cell nuclei in confocal microscopy images

- Enter the scaling parameter expectedPixelsPer100Nm – the number of pixels per 100 nanometers;
- Read color image I in RGB format, depth 8 bits per channel;
- Bring the image I to a scale of 1.5 nanometers per pixel;
- Convert I to HSV format, write the V component to the variable V;
- Apply the contrast limited adaptive histogram equalization method with clipLimit = 2 and titleGridSize = 8 on image V;
- Perform erosion on image V with an ellipse core of size 3;
- Calculate mean as the average value of pixels V;
- For each pixel: V_{ij}; if V_{ij} > mean + 20, assign V_{ij} = min (V_{ij} + 100, 255) if V_{ij} < mean − 20, assign V_{ij} = max (V_{ij} − 100, 0);
- Apply a median filter with a core of size 5 to the image V;
- Perform threshold binarization of image V with a threshold 127. Write the result to variable B;
- Perform a contour search on image B, leaving contours that do not have nested paths;

Calculate the centers of mass (C_x, C_y) for each contour using formulas (1):

\[
C_x = \frac{m_{10}}{m_{00}}; C_y = \frac{m_{01}}{m_{00}} \\
\]

\[
m_{p,q} = \sum_{x,y} x^p y^q 
\]

(1)
Leave only those contours for which $|C_x - C_y| < 1$;

- The number of contours received will be the total number of nuclei.

6.2 Detection of the internal contours of the glands in confocal microscopy images

- Read color image $I$ in RGB format, depth 8 bits per channel;
- Bring the image to a scale of 1.5 nanometers per pixel;
- Convert image $I$ to HSV format, write the $V$ component to the variable $V$;
- Use the contrast limited adaptive histogram equalization method with parameters $clipLimit = 2$ and $tileGridSize = 8$ on image $V$;
- Perform threshold binarization on $V$ component of HSV with a threshold of 127;
- Perform 27 erosion steps on the image $V$ with the ellipse core of size 3;
- Perform a contour detection on image $B$, leaving paths that do not have nested paths. Write the result to the contours variable;
- For each contour: calculate the area, count the number of pixels $V_i$ falling into this contour, taking into account the $V_i$ exceeding 15, and write to the variable nonZeroPixelsArea;
  - suppose that the contour is the inner border of the gland if the nonzero pixels of the region exceed the product of the contour of the region 0.4;
- Recognized glands boundaries will be in the contours list on the output image.

6.3 Counting the number of cell nuclei in light microscopy images

- Enter image $I$ in RGB format and bring it to a scale of 1.5 nanometers per pixel (similar to algorithms A and B);
- Convert $I$ to HSV format, write the $V$ component to the variable $V$ (similar to algorithms A and B);
  - Equalize the histogram of the $V$ component in intensity. To do this,
    - obtain the hist distribution vector of colors $(0..255)$;
    - accumulate the histogram values in $hist$, where $hist[n] = \sum_{i=0}^{n} hist[i] \cdot hist[i]$;
    - calculate the $V_{max}$ as: $V_{max} = hist[len(hist)] - 1 \times \frac{100 - \text{thresh}}{100}$, where thresh is a threshold value equal to 1;
    - replace $V$ pixels with a value greater than $V_{max}$ with a value of $V_{max}$;
  - Perform Laplass transforms on image $V$ with sigma equal to 3, 6 and 9, write the results in $L_1, L_2, L_3$, respectively;
    - calculate the sum of the images $L_1, L_2, L_3$ by pixels and write to laplaceSum; normalize $L_{sum}$ in the range from 0 to 255;
    - perform Otsu binarization on $L_{sum}$; write the result to variable $B$;
  - Search for contours in image $B$, leaving contours that do not have nested paths (similar to algorithm A);
- The task of scaling the image to 1.5 nanometers per pixel was solved by manual measuring the length of the scale bar on the image in pixels. However, we plan to develop automated recognition of the scale bar’s length on images.

7. Results

7.1 The algorithm for counting the number of cell nuclei in confocal microscopy images

The percent of successful recognition in the researched images was 84%. An example of image recognition is presented in Fig. 1.

For the task of counting nuclei, the percentage of successful recognition was presented as the average value for the analyzed images. For each image $I$, there were the number of detected nuclei $N_d$ and the number of labeled nuclei $N_l$. The percent of recognition $p$ for image $I$ was calculated by the formula (2):

$$
p = \frac{\min(N_d, N_l)}{\max(N_d, N_l)} \times 100\%$$

Fig. 1. Highlighting of nuclei on the preparations of confocal microscopy.

Initial image (left), image with selected nuclei (right)

7.2 Highlighting of the internal contours of the glands on confocal microscopy images.

The percentage of successful recognition in the researched images was 70%. An example of image recognition is shown in Fig. 2.
For the problem of counting glands, the percentage of successful recognition was given as the average value for the analyzed images. For each image $I$, there were internal contours of the glands detected by the algorithm – DetectedContours and internal contours of the glands marked by the pathomorphologist – labelledContours. The intersection area of the marked and detected contours $A_d$ and the total area of the marked out contours – $A_I$ were calculated. The percent $p$ of recognition for image $I$ was calculated by the formula (3):

$$P = \frac{A_d}{A_I} \cdot 100\%$$

(3)

7.3 The estimation of the number of cell nuclei in light microscopy images

The percentage of successful recognition was 90% in the researched images. An example of image recognition is presented in Fig. 3.

For the task of counting nuclei on light microscopy preparations, the percentage of successful recognition was analyzed by the formula (2) similarly to the percentage of recognition for confocal microscopy.

8. Comparison with existed software

The purpose of last stage of this work was to compare the obtained data with the results of other applications used to evaluate microphotographs.

Since most of the software intended for cytological studies are either expensive or require a long study of the manual and programming languages, it was decided to compare the performance of the created software with Fiji – the ImageJ plugin for evaluating microscopy images, which is the most balanced among its plugins. The initial task was to estimate the number of cells in 30 microphotographs in the jpg format. The comparison results between proposed algorithm and Fiji for confocal imagery task is presented in Table 4.

Table 4. Comparison of created algorithm and Fiji for confocal imagery

| No | Labelled | Fiji % | Created algorithm % | % of created algorithm |
|----|----------|--------|----------------------|------------------------|
| 1(5) | 11 | 15 | 73,3 | 14 | 78,5 |
| 1(6) | 11 | 17 | 64,7 | 18 | 61,1 |
| 1(7) | 95 | 109 | 87,2 | 94 | 98,9 |
| 1(9) | 12 | 25 | 48,0 | 21 | 54,2 |
| 1(13) | 360 | 424 | 84,9 | 195 | 98,9 |
| 1(30) | 145 | 261 | 55,6 | 165 | 87,8 |
| 1(31) | 128 | 161 | 79,5 | 128 | 100 |
| 1(38) | 45 | 59 | 76,3 | 57 | 78,9 |
| 1(41) | 189 | 192 | 98,4 | 182 | 96,2 |
| 1(42) | 123 | 121 | 98,4 | 117 | 95,1 |

Result of Fiji 76% +/- 16.1%

Result of created algorithm 83,8% +/- 11.8%

For the task of counting nuclei on light microscopy preparations, the percentage of successful recognition was analyzed by the formula (2) similarly to the percentage of recognition for confocal microscopy.

Thus, the developed algorithm exceeds the Fiji accuracy by 7.8%. It should be noted that in order to achieve maximum accuracy in Fiji, the threshold parameter is required and the estimated radius of the object of interest should be introduced as the lower limit, whereas in the written code the determination of the size of objects takes place automatically, which excludes the element of subjectivity from the study and the necessity for preliminary processing of photographs. For example, to evaluate the image, the created algorithm does not require preliminary removal of the scale bar from the image. Also, for the various color markers expression in the nuclei classification task using Fiji, it is necessary to set the Color Threshold value for each type of marker in each photo separately, which not only significantly increases the time of analyze carried out by the pathomorphologist, but also greatly reduces the quality of this analysis, since nuclei having weak expression are most likely will not be included in the corresponding group. Unlike Fiji, the created algorithm equally effectively copes with the task of counting the total number of cores as well as with the task of classifying them.

6. Further research

In the future, it is planned to continue research in this area, improving the reliability of the algorithms, particularly:
use CNN and U-net networks for better segmentation of core images [16, 17];
detect the contours of the scale bar for automatic scaling of the drug;
automatically recognize marker colors (support more than two colors).

References / Список литературы

[1]. Kuznet S.M., Panteleev V.G. The application the Hardware-Software Complex «VideoTest – Morphology» for the differentiation of tumor cells. Clinic. N.1. 2011, pp. 122-123 (in Russian) / С.М. Кузнецов, В.Г. Пантелеев. Использование аппаратно-программного комплекса «ВидеоТест – Морфология» для дифференциации опухолевых клеток. Пеликлиника, no 1, 2011 г., стр. 122-123.
Сергеев Д.И., Андреев А.Е., Дробинцева А.О., Ценевска С., Кукавица Н., Дробинцев П.Д. Разработка автоматизированных алгоритмов компьютерного зрения для обработки медицинских изображений. Труды ИСП РАН, том 32, вып. 3, 2020 г., стр. 119-130

[2]. ImageJ software homepage (2020). Available at https://imagej.net.
[3]. VideoTextMorphology software homepage (2020). Available at http://www.digitalimagingsystems.co.uk/pdfs/morpho_en.pdf.
[4]. IMAGE-PRO Premium software homepage (2020). Available at https://www.mediacy.com/imagepro.
[5]. CellProfiler software homepage (2020). Available at https://cellprofiler.org.
[6]. OrthoImage/Analysis software homepage (2020). Available at https://www.orbit.bio.
[7]. AxiomVision 4.8 software homepage (2020). Available at https://www.micro- shop.zeiss.com/en/us/system/software-axiovision+software-products/1007/.
[8]. CellSens software homepage (2020). Available at https://www.olympus-lifesience.com/en/software/cellsens/.
[9]. Bereszky O.N., Melnik G.N. Information technology for the analysis and synthesis of histological images in automated microcopy systems. Control systems and computers, no. 4, 2013, pp. 26-32 (in Russian) / О.Н. Березский, Г.Н. Мельник. Информационная технология анализа и синтеза гистологических изображений в системах автоматизированной микроскопии. Управляющие системы и машины, no. 4, 2013 г., стр. 26-32.
[10]. Carrarelli P., Rocha A. Increased expression of antimullerian hormone and its receptor in endometriosis. Fertility and sterility, vol. 101, no. 5, pp. 1353-1358.
[11]. Drobintseva A.O., Polyakova V.O., Masing D.S., Matyushkin L.B. Confocal microscopy. Role and importance in the study of the reproductive system. Tutorial. SPb., TsOP "Nevsky", 2015, .18 p. (in Russian) / А.О. Дробинцева, В.О. Полякова, Д.С. Мазин, Л.Б. Матюшкин. Конфокальная микроскопия. роль и значение в исследовании репродуктивной системы: учебное пособие. СПб., ЦОП «Невский», 2015 г., 18 стр.
[12]. Chen, S., Zhao, M., Wu, G., Yao, C., Zhang, J. Recent Advances in Morphological Cell Image Analysis. Computational and Mathematical Methods in Medicine, vol. 2012, Article ID 101536, 10 p.
[13]. Gurcan M. N., Boucheron L., Can A., Madabushi A., Rapoot N., & Yener B. Histopathological Image Analysis: A Review. IEEE Reviews in Biomedical Engineering, vol. 2, 2009, pp. 147–171.
[14]. Kovrigin A.V. Application of the principles of constructing machine vision systems in the task of analyzing images of cellular structures. Scientific journal of KubSAU, no. 29 (5), 2007, pp. 1-10 (in Russian) / Ковригин А.В. Применение принципов построения систем машинного зрения в задаче анализа изображений клеточных структур. Научный журнал Кубанского государственного аграрного университета, no. 29(5), 2007 г., стр. 1-10.
[15]. Xue Y., Ray N. (Cell Detection with Deep Convolutional Neural Network and Compressed Sensing. arXiv:1708.03307, 2007.
[16]. Falk, T., Mai, D., Bensch, R. et al. U-Net: deep learning for cell counting, detection, and morphometry. Nature Methods, vol. 16, 2019, pp. 67–70 (2019).
[17]. Sergreev D.I., Drobintseva A.O. et al. Biomedicine diagnostic repository. Source code of this paper. Available at https://github.com/densvr/biomedicine-diagnostic.

Information about authors / Информация об авторах

Daniel Igorevich SERGEEV – PhD student of the Institute of Computer Science and Technology. Scientific interests: computer vision, machine learning, development in the Android environment.

Sergey D.I., Andreev A.E., Drobintseva A.O., Cenevskis S., Kukavicsa N., Drobintsev P.D. Development of automated computer vision methods for cell counting and endometrial gland detection for medical images processing. Tруды ИСП РАН, том 32, вып. 3, 2020 г., стр. 119-130

Анна Олеговна ДРОБИНЦЕВА – доцент, кандидат биологических наук, доцент кафедры медицинской биологии. Научные интересы: биомедицинская диагностика, молекулярные маркеры, пептидные гормоны, бесплодие, микроскопия.

Nikola KUKAVITSA is a graduate student of the Institute of Computer Science and Technology. Scientific interests: computer vision, machine learning, python.

Alexander Evgenievich ANDREEV – graduate of the magistracy of SPbPU, researcher. Research interests: biomedical diagnostics, computer vision, machine learning.

Pavel Dmitrievich DROBINTSEV – Associate Professor, Candidate of Technical Sciences, Director of the Higher School of Software Engineering at the Institute of Computer Science and Technology. Research interests: computer vision, machine learning, python.

Information about authors / Информация об авторах

Daniel Igorevich SERGEEV – Ph.D. student of the Institute of Computer Science and Technology. Scientific interests: computer vision, machine learning, development in the Android environment.

Alexander Evgenievich ANDREEV – graduate of the magistracy of SPbPU, researcher. Research interests: biomedical diagnostics, computer vision, machine learning.

Information about authors / Информация об авторах

Daniel Igorevich SERGEEV – PhD student of the Institute of Computer Science and Technology. Scientific interests: computer vision, machine learning, development in the Android environment.

Nikola KUKAVITSA is a graduate student of the Institute of Computer Science and Technology. Scientific interests: computer vision, machine learning, python.