Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 159 (2016) 58 – 65

Procedia Engineering

www.elsevier.com/locate/procedia

Humanitarian Technology: Science, Systems and Global Impact 2016, HumTech2016, 7-9 June 2016, Massachusetts, USA

Primer for image informatics in personalized medicine

Young Hwan Chang[a], Patrick Foley[b], Vahid Azimi[b], Rohan Borkar[b], and Jonathan Lefman[b]*

[a]*Biomedical Engineering, Oregon Health & Science University, Portland, Oregon USA*
[b]*Health and Life Sciences, Intel Corporation*

Abstract

Image informatics encompasses the concept of extracting and quantifying information contained in image data. Scenes, what an image contains, come from many imager devices such as consumer electronics, medical imaging systems, 3D laser scanners, microscopes, or satellites. There is a marked increase in image informatics applications as there have been simultaneous advances in imaging platforms, data availability due to social media, and big data analytics. An area ready to take advantage of these developments is personalized medicine, the concept where the goal is tailor healthcare to the individual. Patient health data is computationally profiled against a large of pool of feature-rich data from other patients to ideally optimize how a physician chooses care. One of the daunting challenges is how to effectively utilize medical image data in personalized medicine. Reliable data analytics products require as much automation as possible, which is a difficulty for data like histopathology and radiology images because we require highly trained expert physicians to interpret the information. This review targets biomedical scientists interested in getting started on tackling image analytics. We present high level discussions of sample preparation and image acquisition; data formats; storage and databases; image processing; computer vision and machine learning; and visualization and interactive programming. Examples will be covered using existing open-source software tools such as ImageJ, CellProfiler, and IPython Notebook. We discuss how difficult real-world challenges faced by image informatics and personalized medicine are being tackled with open-source biomedical data and software.

Peer-review under responsibility of the Organizing Committee of HumTech2016

Keywords: bioinformatics; image analytics; image informatics; personalized medicine;

1. Introduction

Image informatics aims to extract, quantify, and compare information contained within images. The field continues to improve efficiency, usability and reliability of image analyses. Various consumer electronics such as cell phone cameras, medical imaging scanners or microscopes are rapidly expanding the data available to image informatics. There is an increasing need for high-throughput, robust image informatics applications in part due to simultaneous advances in imaging platforms, data availability due to social media, and big data analytics.

Personalized medicine, with the goal to tailor healthcare to individuals, is preparing to take advantage of these developments. Patient health data is computationally profiled against a large pool of feature-rich data from other patients to ideally optimize how a physician chooses care. Medical imaging plays an important role in diagnostics and monitoring. Numerous challenges involved with medical images need to be tackled before this data can be used at scale.

Analytic approaches to biomedical images range from full-automation to manual, expert interpretation. Creating reliable analytics data products to be consumed by personalized medicine requires a high degree of automation. This is necessary in part to limit effects of individual decision bias; and to create standards for interpretation procedures and data processing speed to digest potentially decades of archived medical data.

* Corresponding author.
E-mail address: Jonathan.Lefman@intel.com

Peer-review under responsibility of the Organizing Committee of HumTech2016
doi:10.1016/j.proeng.2016.08.064

Automating meaningful quantitative methods for histopathology images is particularly challenging. Consider the work performed by pathologist physicians with extensive, specialized training. Their diagnostic assessments extend beyond what is seen under the microscope; often referring to medical records and experience. Replicating human expertise is particularly challenging as well as controversial. Most productive conversations tend to focus on automating what can be performed by machine, such as image acquisition and quantitative cell imaging. For example, to reduce some subjective bias and costly manual interpretation of tissue slides many algorithms and commercial tools for quantitative and objective analysis have been developed [1, 2].

Scanners such as magnetic resonance imaging (MRI) or computed tomography (CT) machines often used in diagnostic radiology output digital image data. There are numerous analytic software tools with broad usage. However, there is limited adoption by clinicians of any computer-aided diagnostic tools [3]. In fact, clinical evidence shows that such tools do not provide clinical benefit [4]. Data challenge contests continue to push for the development of tools to meet the needs of clinicians [5, 6].

In this primer, we present common, initial approaches to biomedical and clinical image data analytics. Concepts including data preparation and data formats; data access and storage systems; image processing; visualization and interactive programming; and computer vision and machine learning. Example workflows demonstrate how researchers new to image analytics can get started using popular open-source software tools such as CellProfiler, ImageJ, and IPython Notebook. These tools represent entry points to complex quantification tasks for personalized medicine.

2. Parameters Affecting Data Acquisition

Two pervasive factors that limit large-scale image analytics in personalized medicine arise from differences involved in operating imaging devices and specimen preparation. To illustrate by example, MRI and CT scanners produced by different manufacturers output varying intensity readings despite using the same scanning protocol [7, 8]. Techniques, such as phantom calibrations, have been developed to address these sorts of issues [9], however they are not yet standards and not used in clinical systems. Inconsistencies also arise in pathology tissue sample preparation where technicians from different laboratories may vary techniques for embedding and staining.

When considering large scale radiology and histology imaging analysis across different patients or many institutions, we need to consider standards image including required illumination or intensity, resolution, compression, file format, pixel depth. Although a wide range of requirements and parameters such as human factors or non-imaging parameters makes universal standards difficult, we need to define these parameters first and develop a systematic approach for standardization. Techniques to perform these sorts of tasks are specific to the imaging modality. Establishing guidelines for these cases require experimental systems to test algorithms and workflows.

3. Storage and Formats

There is momentum toward digital image acquisition and storage in clinical radiology. Common diagnostic sonography scanners and X-ray machines are moving toward digital recording directly to electronic health record (EHR) systems. There is rapid feedback for clinicians, who may not be onsite with the patient. X-ray radiation dosages have been reduced using digital scanner systems while maintaining or improving image quality [10]. Pathology is also adopting digital microscopy systems to allow integration with consultation services and hospital EHRs.

The accepted standard medical image format is Digital Imaging and Communications in Medicine (DICOM), which also includes definitions for data transmission [11]. The consortium for updating these standards are led by medical imaging experts and device manufacturers. DICOM images are outputted from scanner to hospital systems, Picture Archiving and Communication Systems (PACS). PACS allow storage and retrieval of images and are often integrated as a component of EHRs.

Digital systems have been in place for decades for MRI and CT systems. More recently digital pathology is relying on whole slide imagers (WSI), automated microscopy systems which record images from slides, possibly at multiple resolutions. These virtual slides are stored and cataloged. However, there is not yet a consensus format such as DICOM for digital pathology. Common formats include TIFF, JPEG, and formats specific to scanner manufacturers.

Many biomedical image formats accommodate compression to reduce on-disk storage requirements and to increase data transmission efficiency. Popular lossless compression techniques include run-length encoding and DEFLATE [12]. Lossy techniques such as JPEG effectively reduce data size while tending to preserve salient visual features. Lossless compression, which tend to be less effective for data size reduction, is preferred for maintaining the original content often desired in analytic applications without introducing compression artifacts observed in lossy compression, as artifacts can affect accuracy of techniques such as segmentation.

Biomedical research groups working on large image datasets often rely on networked file systems which provide extended storage and often backup. Many groups rely on internal information technology departments for management. The technologies that provide this functionality are storage area network (SAN), network attached storage (NAS), and distributed file systems.

SAN and NAS both allow remote storage volumes to be mounted as a system drive. Distributed file systems such as Lustre, Gluster, and Ceph are utilized for redundancy and reliability within larger storage systems.

In recent years, public cloud providers such as Amazon Web Services (AWS) and Google Cloud have gained popularity with biomedical research groups. The popularity of these storage services can be attributed to the low cost and horizontally scalability of these solutions. In the context of image informatics, labs may produce more data than is locally available, making cloud storage attractive. A major caveats of cloud storage systems in clinical settings are Protected Health Information and Health Insurance Portability and Accountability Act (HIPAA) privacy enforcement [13]. Some providers have demonstrated capabilities to go beyond HIPAA requirements, yet there is continued reluctance to utilize cloud storage by hospital systems.

4. Applications of Image Processing

Image processing techniques arise from signal processing. Operations are most frequently used to enhance and extract features or to suppress unwanted information like noise or background. Here example applications of image processing often applied to 2D pathology and 3D radiology images are discussed. This is a cursory overview of possible approaches.

4.1. Stain normalization for histopathology

Staining histology specimens often vary. Stain normalization is the process which measures staining variation using histogram equalization or the rank statistics of the input image and removes this variation in each color channel [14]. In situations where either hematoxylin or eosin stain components have unequal representation, normalization approaches like color deconvolution is used [15].

4.2. MRI dataset registration

Aligning and warping data, known as registration, is frequently used in longitudinal or population comparison radiological studies. Bringing together multiple 3D radiology images generally requires calculating the offsets to align. Offsets correspond to rigid and non-rigid transformations. Rigid transformations affect scale, translation, and rotation while non-rigid transformations deal with affine and non-linear deformations [16].

5. Applications of Computer Vision and Machine Learning

Approaches to recognizing what is happening in images along with understanding trends within large datasets have arisen from combinations of algorithms in two distinct sub-fields of computer science: computer vision and machine learning. Computer vision algorithms often perform tasks such as segmentation, detection, and tracking. Machine learning algorithms use image analysis to develop models or representations for predicting or categorizing image content. Recent trends in deep learning using convolutional neural networks are blurring boundaries between computer vision and machine learning [17].

Methods to discover useful information biomedical data like histopathology microscopy images tends to be specific to which questions are being asked. For example, automated cancer grading techniques require detecting suspected tumor regions, segmenting structures like nuclei, and deciding which cancer grade the structures represent. To do so, first we need to extract features, the components of the data contributing a statistically significant signal which can be used for identifying different categories. Using histopathology as an example, computed features like morphological profiles and texture are extensively used for defining nuclei types [18]. From extracted features, nuclei can be classified to categories tumor-like or normal [19].

Complex classification tasks, such as segmenting normal versus tumor cells, require a combination of unsupervised and supervised learning techniques in order to be effective. Unsupervised learning uses concepts like principal component analysis and clustering, which group statistically similar data features. Supervised learning relies on training data containing annotated or correctly identified observations. Supervised approaches tend to require large amounts of reliable training data. In personalized medicine, preparing training data can be both expensive and time consuming. Training data preparation for histopathology or radiology requires a pathologist or radiologist, respectively, to visually examine and correctly annotate images. Depending on the machine learning goal, these tasks may require both precise hand-drawn polygon labels and consensus from several specialists annotating the data to limit individual bias in training. Recent semi-supervised learning approaches aim to unify the prior techniques, requiring a small training dataset while progressively augmenting its size using a best-guess approach. There is an ongoing focus on engineering feature selection procedures to determine optimal approaches for machine learning techniques such as Bayesian networks, Markov models, K-nearest neighbors and support vector machines [20].

6. Visualization and Interactive Programming

Image visualization tools are needed to both aid researchers in experimental workflow development and present clinicians

with images needed for decision making. Presentation of images may be original "raw" data or transformed data, depending on the application. Pathology images are most often presented as large 2D microscopy slide panels with capabilities such as zoom and region of interest annotations. MRI and CT instruments capture data in three dimensions, but are presented to clinicians in 2D to simplify analysis. This 2D presentation of a 3D scan across sagittal, coronal, and transverse views is frequently referred to as 2.5D visualization.

Various system architecture styles may be used with visualizations. Viewers may reside as a stand-alone desktop application. Clinical radiology images are most frequently presented by online viewers retrieve which images from credentialed data server while rendering occurs at the local computer system. There are also efforts to bring clinical viewers to smart phones and tablets where rendering may be occur at the client or server side depending on devices and bandwidth.

The Open Health Imaging Foundation (OHIF) is a non-profit organization focused on developing open source clinical radiology data tracking and viewer [21]. The OHIF viewer runs in a web browser with capabilities to support views in standard desktop systems as well as some viewer support tablet computers and smart phones. By developing open-source software the foundation hopes to gain support and co-development by the community. Rendering occurs on the client system while data is served remotely.

OMERO is an open source system for handling microscopy data, providing data visualization, data management and some functionality for data processing [22]. The Bio-Formats library is a Java library which reads and writes over 140 proprietary microscopy formats. This library is a core component of OMERO's capability to work with many types of microscopy systems. Software in the OMERO distribution includes a desktop client, web client, and API, which allows developers to build custom applications on top of OMERO. The Open Microscopy Environment team along with the biomedical imaging community develop image analysis scripts that can be executed from the OMERO client.

Cytomine is an open source web application solution for remote pathology image viewing and collaborative annotations [23]. By utilizing modern web frameworks and building on the efforts of other open source image pathology tools, Pathology scientists often use the tool find regions of interest using built-in computer vision algorithms. Summaries can be exported directly to email.

Developing algorithms and workflows to transform images into measurable units is aided by rapid feedback from integrated visualization tools. One of the most popular tools for biomedical imaging is ImageJ [24]. The application originally created at the NIH started in 1987 has since become one of the most widely used image processing application in the biological microscopy community. The reason for its popularity among scientists can be attributed to its cross platform execution, ease to use, and wide file format support. The biological imaging community has contributed over 400 image algorithms back to the code base. These plugins ship with Fiji, an enhanced variant of ImageJ [25]. ImageJ and Fiji provide 2D, 2.5D, 3D, and volume rendering views.

The wide variety of algorithms and workflow packages provided with Fiji allow quick experimentation with rapid visual feedback. As an example of Fiji utility, we show a method to segment tumor masses from MRI images in the glioblastoma component of the Cancer Imaging Archive [26]. Each dataset contains at least one likely tumor mass per image providing the possibility to extract features such as size, shape, or texture across the entire cohort. Figure 1 shows how the "level sets" segmentation plugin was used to perform 3D segmentation on a large, distinctive tumor. The resulting segmentation may be visualized as well as quantified. Greyscale value intensities with the segmented region present a narrower distribution with larger values.

7. Illustrations

All figures should be numbered with Arabic numerals (1,2,3,….). Every figure should have a caption. All photographs, schemas, graphs and diagrams are to be referred to as figures. Line drawings should be good quality scans or true electronic output. Low-quality scans are not acceptable. Figures must be embedded into the text and not supplied separately. In MS word input the figures must be properly coded. Preferred format of figures are PNG, JPEG, GIF etc. Lettering and symbols should be clearly defined either in the caption or in a legend provided as part of the figure. Figures should be placed at the top or bottom of a page wherever possible, as close as possible to the first reference to them in the paper. Please ensure that all the figures are of 300 DPI resolutions as this will facilitate good output.

The figure number and caption should be typed below the illustration in 8 pt and left justified [***Note:*** one-line captions of length less than column width (or full typesetting width or oblong) centered]. For more guidelines and information to help you submit high quality artwork please visit: http://www.elsevier.com/artworkinstructions Artwork has no text along the side of it in the main body of the text. However, if two images fit next to each other, these may be placed next to each other to save space. For example, see Fig. 1.

CellProfiler is an open-source, cross platform biological image analysis software package developed by the Broad Institute [27]. The software is targeted for use in high-throughput in vitro cell phenotype quantification. The user may use the graphical interface to construct analytical pipelines involving image processing, computer vision, and statistical analysis without requiring knowledge of programming nor principles of signal and image processing. Those with training may access methods to programmatically operate the software, allowing finer control over processes. There are over 70 processing modules, each with

accompanying tutorials, which can be used to generate processing pipelines. If desired functionality is not available, additional functionality can be programmed and imported through APIs.

Data processing pipelines, which chain customized processing and analysis functions can be reused and distributed as text files. These allow sharing of the steps needed to achieve results, which is an increasingly common requirement of scientific journals [28]. Pipelines can operate in computing cluster systems. Results may be outputted to flat files or MySQL databases, which provides data format consistency and a rudimentary setup for data provenance.

A common use case is quantifying in vitro cancer cell responses to drug molecules. In this example dataset, microscopy images were acquired of cancer cell lines in three experimental conditions: control, wortmannin, and drug compound LY294002 [29]. While the biological significance is beyond this text, we show how CellProfiler can be used to quantify biochemical processes that is captured as image snapshots. The steps illustrated in show the transformation from images to meaningful results (figure 2). After importing images, processing modules are added to the pipeline to run serial operations. Here, the first module applies illumination correction which normalizes brightness and contrast within the image. Next, a watershed algorithm segments cells against the background. The segmented image results are then used to extract features like size, shape, intensity, and texture; lastly, the export module writes this data to common file formats, which may be used in downstream analyses.

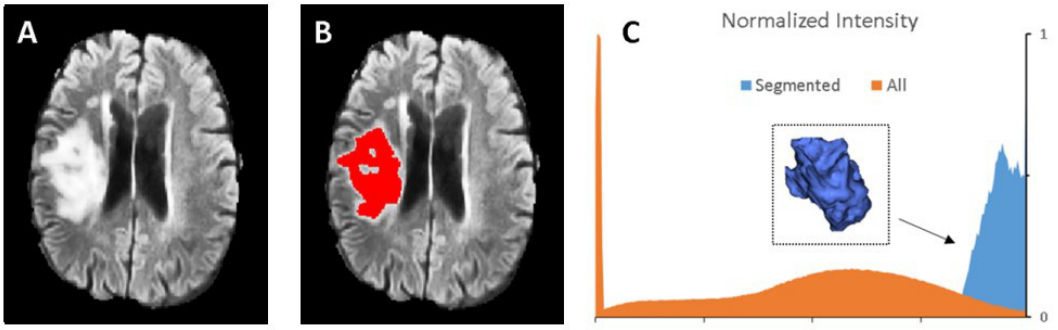

Fig. 1. The enhanced distribution of ImageJ, Fiji, was used to segment and quantify a candidate glioblastoma tumor. (A) The site of the tumor can be identified against apparent normal tissue background. (B) Applying level set segmentation with a single seed point yields a discrete 3D sub-volume. (C) A normalized histogram of pixel intensities of the tumor component, rendered as an isosurface in the boxed sub-panel, is compared to the normalized histogram of intensity values of the original image volume.

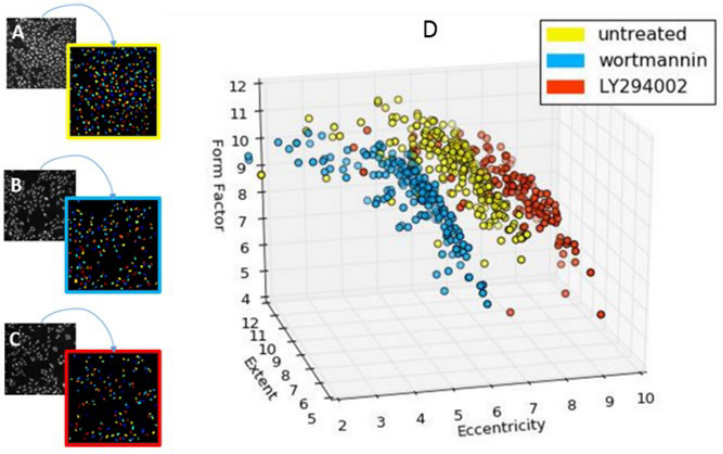

Fig. 2. Images of cancer cells having various treatments. In each panel, the original microscopy image is transformed to a representation of the segmented results. Conditions represented are (A) the control, (B) wortmannin, and (C) compound LY294002. (D) Shows a summary graphical view of the three different conditions. Clusters form based on three extracted features from the images (all three features are normalized to unit variance).

8. IPython Notebook

IPython Notebook (shortened to Notebook) is a web browser based interactive coding and experimentation environment [30]. Code, documentation, and results are in files which may be shared. Notebook evolved from the IPython shell, an interactive Python command-line interface. Features of IPython and Notebook include macros called magic commands, completion, and execution history. Notebook was designed around concepts in other graphical notebook environments, like Mathematica and Sage for the Python language. Its creators aimed to make it a natural extension of the shell. Notebook's intuitive development model has become popular for data scientists and coders alike, so much so that there is now kernel support for over 40 other languages under Project Jupyter.

Code may be executed in cells, which are isolated coding sections sharing a global space. Running individual cells allows for rapid experimentation without running all code, as would be the case in typical file-based development. Cell execution results are displayed inline after each cell, which gives a unified place for feedback and code instead of performing command line operations.

Illustrated here is an interactive session for segmenting cell nuclei from an H&E prepared tissue section (figure 3). The view_image function take a threshold argument and displays an overlay of colorized regions above this number. The browse_images function calls view_image using Notebook's interact functionality - which makes the interface responsive to user input without requiring page reloads. Until Notebook abstracted and generalized this type of interactivity, graphical feedback depended on Python libraries like wx, larger graphical libraries like Qt, or domain specific tools such as user interface tools in the computer vision library, OpenCV. Multiple interactive tools can be applied, in this case to apply size filters for segmentation results. The ultimate analysis product here is a matplotlib scatter plot which compares segmented cell area to H&E stain intensity.

9. Conclusion

Personalized medicine is preparing to utilize large and varied clinical data. Bioinformatics efforts are rapidly expanding the use and availability of genomic sequencing data. In practice, whole genomes can be compared across individuals and species as linear sequences. However, genomic data is fundamentally limited to the four bases of DNA but spatial arrangement and organization of cells, tissues, and structure are generally not reflected in genomics. Thus, integrating imaging informatics can refine and complement genomic analysis. In general, biomedical images have a larger degree of complexity, as image data is multi-dimensional and frequently composed of multiple channels with a wide range of pixel values.

To uncover the salient information in images, we must understand how to extract these signals or reduce less-meaningful background. This large parameter space frequently results from data acquisition and scanner systems; and biological differences in specimens and patients. This limits general solutions to extract clinically actionable information for most types of biomedical images, resulting in custom or manually tuned approaches for each application. Recent efforts in deep learning, celebrated for high accuracy in classification tasks, are at the intersection of computer vision and machine learning. Deep learning tends to require an abundance of data for training, which is challenging for clinical applications because not enough examples of a condition exist, or medical privacy restrictions.

Ongoing efforts are working to meet these challenges; the National Cancer Institute hosts open source datasets, such as The Cancer Genome Atlas and The Cancer Imaging Archive data portals, as well as open data competitions. These public initiatives, coupled with open source software development, promote global, cross-disciplinary efforts. Simultaneously, public cloud computing platforms are transforming the scale of these approaches. Where medical privacy restrictions remain in place, large hospital systems are able to pool internal data sources for systemic big data approaches [31].

Utilizing concepts presented here, researchers new to image analytics can develop novel approaches to image analytics. Software tools providing visual feedback are most effective for rapidly creating sophisticated workflows. Those with programming experience have greater flexibility to investigate new methods.

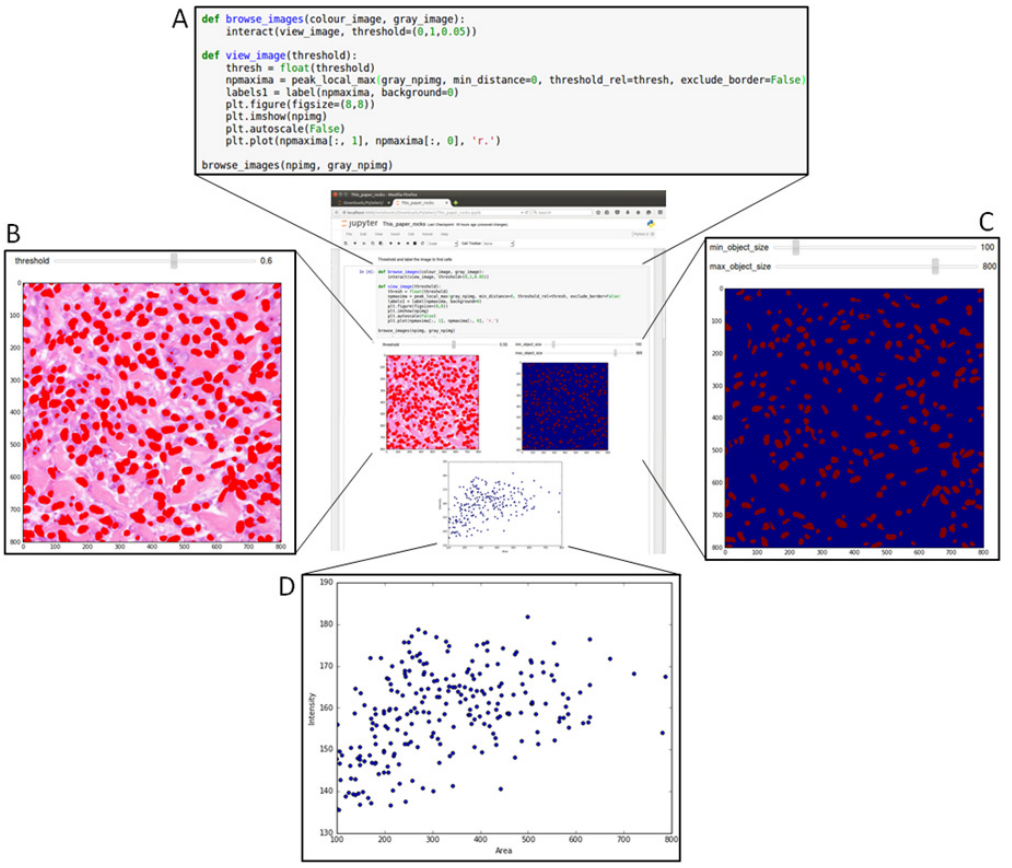

Fig. 3. Illustration of example workflow. (A) Python code is edited in the cell shown here. (B) Image display with the original image with an interactive overlay of segmentation results. The slider above the image changes threshold level. (C) Segmented objects are shown in red on a blue background. The sliders above the image determine minimum and maximum object size. Objects outside that range are omitted from the image and resulting list. (D) Scatter plot showing results of segmentation and cell sorting. Each point represents a cell object where x-axis is object area and y-axis is mean object intensity.

References

[1] Kothari, S., et al., Pathology imaging informatics for quantitative analysis of whole-slide images. J Am Med Inform Assoc, 2013. 20(6): p. 1099-108.
[2] Ertosun, M.G. and D.L. Rubin, Automated Grading of Gliomas using Deep Learning in Digital Pathology Images: A modular approach with ensemble of convolutional neural networks. AMIA Annu Symp Proc, 2015. 2015: p. 1899-908.
[3] van Ginneken, B., C.M. Schaefer-Prokop, and M. Prokop, Computer-aided diagnosis: how to move from the laboratory to the clinic. Radiology, 2011. 261(3): p. 719-32.
[4] Lehman, C.D., et al., Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection. JAMA Intern Med, 2015. 175(11): p. 1828-37.
[5] Kistler, M., et al., The virtual skeleton database: an open access repository for biomedical research and collaboration. J Med Internet Res, 2013. 15(11): p. e245.
[6] Kaggle. Diabetic Retinopathy Detection. 2015; Available from: https://www.kaggle.com/c/diabetic-retinopathy-detection.
[7] Plante, E. and L. Turkstra, Sources of error in the quantitative analysis of MRI scans. Magn Reson Imaging, 1991. 9(4): p. 589-95.
[8] Gulliksrud, K., C. Stokke, and A.C. Martinsen, How to measure CT image quality: variations in CT-numbers, uniformity and low contrast resolution for a CT quality assurance phantom. Phys Med, 2014. 30(4): p. 521-6.
[9] O'Callaghan, J., et al., Is your system calibrated? MRI gradient system calibration for pre-clinical, high-resolution imaging. PLoS One, 2014. 9(5): p. e96568.
[10] Uffmann, M. and C. Schaefer-Prokop, Digital radiography: the balance between image quality and required radiation dose. Eur J Radiol, 2009. 72(2): p. 202-8.
[11] Mustra, M., K. Delac, and M. Grgic. Overview of the DICOM standard. in ELMAR, 2008. 50th International Symposium. 2008.
[12] Katz, P.W., String searcher, and compressor using same. 1991, Google Patents.
[13] Office of the Secretary, H.H.S., HIPAA administrative simplification: modifications to medical data code set standards to adopt ID-10-CM and ICD-10-PCS. Final rule. Fed Regist, 2009. 74(11): p. 3328-62.
[14] Niethammer, M., et al., Appearance Normalization of Histology Slides. Mach Learn Med Imaging, 2010. 6357: p. 58-66.

[15] Khan, A.M., et al., A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. IEEE Trans Biomed Eng, 2014. 61(6): p. 1729-38.
[16] Fedorov, A., et al., Evaluation of brain MRI alignment with the robust Hausdorff distance measures, in Advances in Visual Computing. 2008, Springer. p. 594-603.
[17] Roux, L., et al., Mitosis detection in breast cancer histological images An ICPR 2012 contest. Journal of pathology informatics, 2013. 4.
[18] Liu, S., P.A. Mundra, and J.C. Rajapakse. Features for cells and nuclei classification. in Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE. 2011. IEEE.
[19] Nayak, N., et al. Classification of tumor histopathology via sparse feature learning. in Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on. 2013. IEEE.
[20] Naik, S., et al. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. in Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on. 2008. IEEE.
[21] Open Health Imaging Foundation. 2016; Available from: http://ohif.org/.
[22] The Open Microscopy Environment. 2016; Available from: https://www.openmicroscopy.org.
[23] Maree, R., et al., Collaborative analysis of multi-gigapixel imaging data using Cytomine. Bioinformatics, 2016.
[24] Schneider, C.A., W.S. Rasband, and K.W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. Nat Methods, 2012. 9(7): p. 671-5.
[25] Schindelin, J., et al., Fiji: an open-source platform for biological-image analysis. Nat Methods, 2012. 9(7): p. 676-82.
[26] Clark, K., et al., The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging, 2013. 26(6): p. 1045-57.
[27] Carpenter, A.E., et al., CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol, 2006. 7(10): p. R100.
[28] Journals unite for reproducibility. Nature, 2014. 515(7525): p. 7.
[29] Ljosa, V., K.L. Sokolnicki, and A.E. Carpenter, Annotated high-throughput microscopy image sets for validation. Nat Methods, 2012. 9(7): p. 637.
[30] Pérez, F. and B.E. Granger, IPython: a system for interactive scientific computing. Computing in Science & Engineering, 2007. 9(3): p. 21-29.
[31] Gainer, V.S., et al., The Biobank Portal for Partners Personalized Medicine: A Query Tool for Working with Consented Biobank Samples, Genotypes, and Phenotypes Using i2b2. J Pers Med, 2016. 6(1).