



**HAL**  
open science

## **Gland segmentation in colon histology images: The glas challenge contest**

Korsuk Sirinukunwattana, Josien P W Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al.

### ► **To cite this version:**

Korsuk Sirinukunwattana, Josien P W Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, et al.. Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis*, 2017, 35, <10.1016/j.media.2016.08.008>. <hal-04868255>

**HAL Id: hal-04868255**

**<https://hal.science/hal-04868255v1>**

Submitted on 6 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC-ND 4.0 - Attribution - Non-commercial use - No Derivative Works - International License

**Original citation:**

Sirinukunwattana, Korsuk, Pluim, Josien P. W., Chen, Hao, Qi, Xiaojuan, Heng, Pheng-Ann, Guo, Yun Bo, Wang, Li Yang, Matuszewski, Bogdan J., Brunie, Elia, Sanchez, Urko, Bohm, Anton, Ronneberger, Olaf, Ben Cheikhh, Bassem, Racoceanu, Daniel, Kainz, Philipp, Pfeiffer, Michael, Urschlerk, Martin, Snead, David R. J. and Rajpoot, Nasir M. (Nasir Mahmood). (2017) Gland segmentation in colon histology images : the GlaS challenge contest. Medical Image Analysis, 35. pp. 489-502.

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/81337>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

© 2016, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/licenses/by-nc-nd/4.0/>

**A note on versions:**

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP URL' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

# Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest

Korsuk Sirinukunwattana<sup>a,\*</sup>, Josien P. W. Pluim<sup>b</sup>, Hao Chen<sup>c</sup>, Xiaojuan Qi<sup>c</sup>, Pheng-Ann Heng<sup>c</sup>, Yun Bo Guo<sup>d</sup>, Li Yang Wang<sup>d</sup>, Bogdan J. Matuszewski<sup>d</sup>, Elia Bruni<sup>e</sup>, Urko Sanchez<sup>e</sup>, Anton Böhm<sup>f</sup>, Olaf Ronneberger<sup>f,g</sup>, Bassem Ben Cheikh<sup>h</sup>, Daniel Racoceanu<sup>h</sup>, Philipp Kainz<sup>i,j</sup>, Michael Pfeiffer<sup>j</sup>, Martin Urschler<sup>k,l</sup>, David R. J. Snead<sup>m</sup>, Nasir M. Rajpoot<sup>a,\*</sup>

<sup>a</sup>*Department of Computer Science, University of Warwick, Coventry, UK, CV4 7AL*

<sup>b</sup>*Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands*

<sup>c</sup>*Department of Computer Science and Engineering, The Chinese University of Hong Kong.*

<sup>d</sup>*School of Engineering, University of Central Lancashire, Preston, UK*

<sup>e</sup>*ExB Research and Development*

<sup>f</sup>*Computer Science Department, University of Freiburg, Germany*

<sup>g</sup>*BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany and Google-DeepMind, London, UK*

<sup>h</sup>*Sorbonne Universités, UPMC Univ Paris 06, CNRS, INSERM, Biomedical Imaging Laboratory (LIB), Paris, France*

<sup>i</sup>*Institute of Biophysics, Center for Physiological Medicine, Medical University of Graz, Graz, Austria*

<sup>j</sup>*Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich, Switzerland*

<sup>k</sup>*Institute for Computer Graphics and Vision, BioTechMed, Graz University of Technology, Graz, Austria*

<sup>l</sup>*Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria*

<sup>m</sup>*Department of Pathology, University Hospitals Coventry and Warwickshire, Walsgrave, Coventry, CV2 2DX, UK*

---

## Abstract

Colorectal adenocarcinoma originating in intestinal glandular structures is

---

\*Corresponding authors

*Email addresses:* [k.sirinukunwattana@warwick.ac.uk](mailto:k.sirinukunwattana@warwick.ac.uk) (Korsuk Sirinukunwattana), [n.m.rajpoot@warwick.ac.uk](mailto:n.m.rajpoot@warwick.ac.uk) (Nasir M. Rajpoot)

the most common form of colon cancer. In clinical practice, the morphology of intestinal glands, including architectural appearance and glandular formation, is used by pathologists to inform prognosis and plan the treatment of individual patients. However, achieving good inter-observer as well as intra-observer reproducibility of cancer grading is still a major challenge in modern pathology. An automated approach which quantifies the morphology of glands is a solution to the problem.

This paper provides an overview to the Gland Segmentation in Colon Histology Images Challenge Contest (GlaS) held at MICCAI'2015. Details of the challenge, including organization, dataset and evaluation criteria, are presented, along with the method descriptions and evaluation results from the top performing methods.

*Keywords:* Histology Image Analysis, Segmentation, Colon Cancer, Intestinal Gland, Digital Pathology

---

## 1. Introduction

Cancer grading is the process of determining the extent of malignancy and is one of the primary criteria used in clinical practice to inform prognosis and plan the treatment of individual patients. However, achieving good reproducibility in grading most cancers remains one of the challenges in pathology practice (Cross et al., 2000; Komuta et al., 2004; Fanshawe et al., 2008). With digitized images of histology slides becoming increasingly ubiquitous, digital pathology offers a viable solution to this problem (May, 2010). Analysis of histology images enables extraction of quantitative morphological features, which can be used for computer-assisted grading of cancer making the grading process more objective and reproducible than it currently is (Gurcan et al., 2009). This has led to the recent surge in development of algorithms for histology image analysis.

In colorectal cancer, morphology of intestinal glands including architectural appearance and gland formation is a key criterion for cancer grading (Compton, 2000; Bosman et al., 2010; Washington et al., 2009). Glands are important histological structures that are present in most organ systems as the main mechanism for secreting proteins and carbohydrates. An intestinal gland (colonic crypt) found in the epithelial layer of the colon, is made up of a single sheet of columnar epithelium, forming a finger-like tubular structure that extends from the inner surface of the colon into the underlying

22 connective tissue (Rubin et al., 2008; Humphries and Wright, 2008). There  
23 are millions of glands in the human colon. Intestinal glands are responsible  
24 for absorption of water and nutrients, secretion of mucus to protect the ep-  
25 ithelium from a hostile chemical and mechanical environment (Gibson et al.,  
26 1996), as well as being a niche for epithelial cells to regenerate (Shanmu-  
27 gathasan and Jothy, 2000; Humphries and Wright, 2008). Due to the hostile  
28 environment, the epithelial layer is continuously regenerating and is one of  
29 the fastest regenerating surface in human body (Crosnier et al., 2006; Barker,  
30 2014). This renewal process requires coordination between cell proliferation,  
31 differentiation, and apoptosis. The loss of integrity in the epithelial cell re-  
32 generation, through a mechanism that is not yet clearly understood, results  
33 in colorectal adenocarcinoma, the most common type of colon cancer.

34 Manual segmentation of glands is a laborious process. Automated gland  
35 segmentation will allow extraction of quantitative features associated with  
36 gland morphology from digitized images of CRC tissue slides. Good quality  
37 gland segmentation will pave the way for computer-assisted grading of CRC  
38 and increase the reproducibility of cancer grading. However, consistent good  
39 quality gland segmentation for all the differentiation grades of cancer has  
40 remained a challenge. This was a main reason for organizing this challenge  
41 contest.

42 The Gland Segmentation in Colon Histology Images (GlaS) challenge<sup>1</sup>  
43 brought together computer vision and medical image computing researchers  
44 to solve the problem of gland segmentation in digitized images of Hema-  
45 toxylin and Eosin (H&E) stained tissue slides. Participants developed gland  
46 segmentation algorithms, which were applied to benign tissue and to colonic  
47 carcinomas. A training dataset was provided, together with ground truth  
48 annotations by an expert pathologist. The participants developed and op-  
49 timized their algorithms on this dataset. The results were judged on the  
50 performance of the algorithms on test datasets. Success was measured by  
51 how closely the automated segmentation matched the pathologist’s.

## 52 2. Related Work

53 Recent papers (Wu et al., 2005a,b; Gunduz-Demir et al., 2010; Fu et al.,  
54 2014; Sirinukunwattana et al., 2015; Cohen et al., 2015) indicate the increas-

---

<sup>1</sup><http://www.warwick.ac.uk/bialab/GlaScontest>

55 ing interest in histology image analysis applied to intestinal gland segmenta-  
56 tion. In this section, we review some of these methods.

57 Wu et al. (2005a) presented a region growing method, which first thresh-  
58 olds an image, in order to separate nuclei from other tissue components.  
59 Large empty regions, which potentially correspond to lumen found in the  
60 middle of glands, are then used to initialize the seed points for region grow-  
61 ing. The expanding process for each seed is terminated when a surround-  
62 ing chain of epithelial nuclei is reached, and subsequently false regions are  
63 removed. Although this method performs well in segmenting healthy and  
64 benign glands, it is less applicable to cancer cases, where the morphology of  
65 glands can be substantially deformed.

66 In contrast to the above method, which mainly uses pixel-level informa-  
67 tion, Gunduz-Demir et al. (2010) represented each tissue component as a  
68 disk. Each disk is represented by a vertex of a graph, with nearby disks  
69 joined by an edge between the corresponding vertices. They proposed an al-  
70 gorithm, using graph connectivity to identify initial seeds for region growing.  
71 To avoid an excessive expansion beyond the glandular region, caused, for ex-  
72 ample, by large gaps in the surrounding epithelial boundary, edges between  
73 nuclear objects are used as a barrier to halt region growing. Those regions  
74 that do not show glandular characteristics are eliminated at the last step.  
75 The validation of this method was limited only to the dataset with healthy  
76 and benign cases.

77 Fu et al. (2014) introduced a segmentation algorithm based on polar  
78 coordinates. A neighborhood of each gland and a center chosen inside the  
79 gland were considered. Using this center to define polar coordinates, the  
80 neighborhood is displayed in  $(r, \theta)$  coordinates with the  $r$ -axis horizontal  
81 and the  $\theta$ -axis vertical. One obtains a vertical strip, periodic with period  
82  $2\pi$  in the vertical direction. As a result, the closed glandular boundary  
83 is transformed into an approximately vertical periodic path, allowing fast  
84 inference of the boundary through a conditional random field model. Support  
85 vector regression is later deployed to verify whether the estimated boundary  
86 corresponds to the true boundary. The algorithm performs well in both  
87 benign and malignant cases stained by Hematoxylin and DAB. However, the  
88 validation on routine H&E stained images was limited only to healthy cases.

89 Sirinukunwattana et al. (2015) recently formulated a segmentation ap-  
90 proach based on Bayesian inference, which allows prior knowledge of the  
91 spatial connectivity and the arrangement of neighboring nuclei on the ep-  
92 ithelial boundary to be taken into account. This approach treats each gland-

Table 1: Details of the dataset.

Histologic Grade	Number of Images (Width x Height in Pixels)		
	Training Part	Test Part A	Test Part B
Benign	37 $\left\{ \begin{array}{l} 1 \quad (574 \times 433) \\ 1 \quad (589 \times 453) \\ 35 \quad (775 \times 522) \end{array} \right.$	33 $\left\{ \begin{array}{l} 1 \quad (574 \times 433) \\ 4 \quad (589 \times 453) \\ 28 \quad (775 \times 522) \end{array} \right.$	4 (775 × 522)
Malignant	48 $\left\{ \begin{array}{l} 1 \quad (567 \times 430) \\ 3 \quad (589 \times 453) \\ 44 \quad (775 \times 522) \end{array} \right.$	27 $\left\{ \begin{array}{l} 1 \quad (578 \times 433) \\ 2 \quad (581 \times 442) \\ 24 \quad (775 \times 522) \end{array} \right.$	16 (775 × 522)

93 dular structure as a polygon made of a random number of vertices. The  
 94 idea is based on the observation that a glandular boundary is formed from  
 95 closely arranged epithelial nuclei. Connecting edges between these epithelial  
 96 nuclei gives a polygon that encapsulates the glandular structure. Inference of  
 97 the polygon is made via Reversible-Jump Markov Chain Monte Carlo. The  
 98 approach shows favorable segmentation results across all histologic grades  
 99 (except for the undifferentiated grade) of colorectal cancers in H&E stained  
 100 images. This method is slow but effective.

101 Most of the works for intestinal gland segmentation have used differ-  
 102 ent datasets and/or criteria to assess their algorithms, making it difficult to  
 103 objectively compare their performance. There have been many previous ini-  
 104 tiatives that provided common datasets and evaluation measures to validate  
 105 algorithms on various medical imaging modalities (Murphy et al., 2011; Gur-  
 106 can et al., 2010; Roux et al., 2013; Veta et al., 2015). This not only allows a  
 107 meaningful comparison of different algorithms but also allows the algorithms  
 108 to be implemented and configured thoroughly to obtain optimal performance  
 109 (Murphy et al., 2011). Following these successful initiatives, we therefore or-  
 110 ganized the Gland Segmentation in Colon Histology Images (GlaS) challenge.  
 111 This challenge was a first attempt to address the issues of reproducibility and  
 112 comparability of the results of intestinal gland segmentation algorithms. It  
 113 was also aimed at speeding up even further the development of algorithms for  
 114 gland segmentation. Note that none of above methods for intestinal gland  
 115 segmentation participated in this competition.

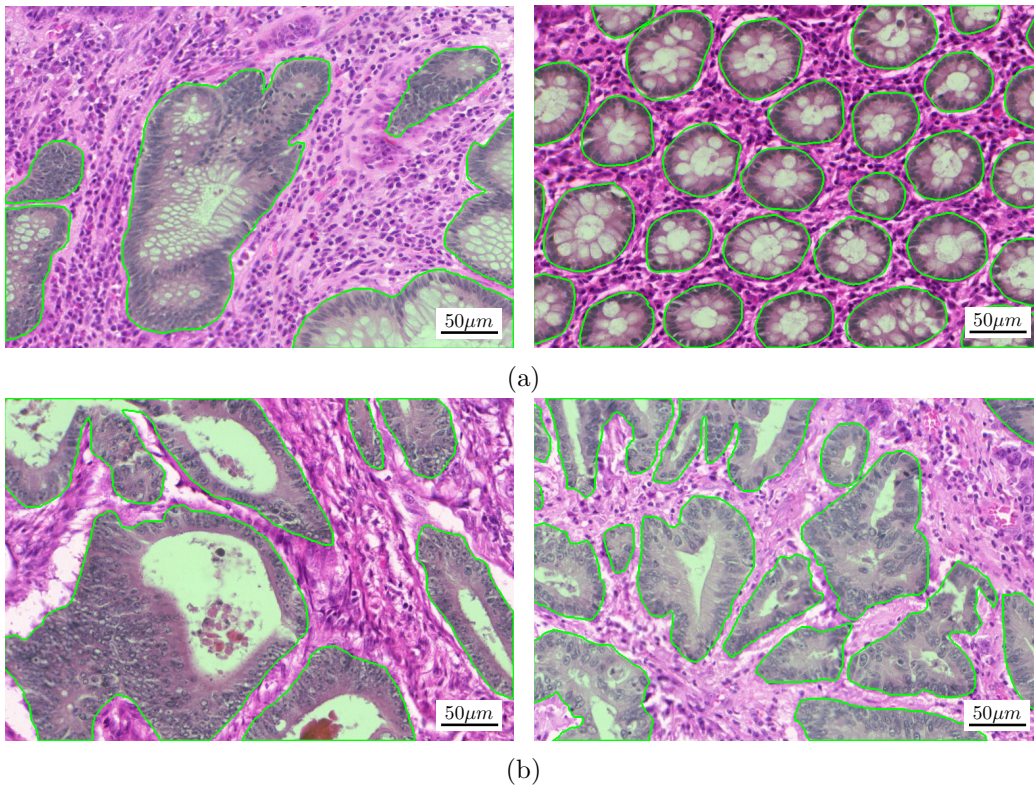


Figure 1: Example images of different histologic grades in the dataset: (a) benign and (b) malignant.

### 116 3. Materials

117 The dataset used in this challenge consists of 165 images derived from  
118 16 H&E stained histological sections of stage T3 or T4<sup>2</sup> colorectal adenocar-  
119 cinoma. Each section belongs to a different patient, and sections were pro-  
120 cessed in the laboratory on different occasions. Thus, the dataset exhibits  
121 high inter-subject variability in both stain distribution and tissue architec-  
122 ture. The digitization of these histological sections into whole-slide images  
123 (WSIs) was accomplished using a Zeiss MIRAX MIDI Slide Scanner with a  
124 pixel resolution of  $0.465\mu m$ . The WSIs were subsequently rescaled to a pixel  
125 resolution of  $0.620\mu m$  (equivalent to  $20\times$  objective magnification).

126 A total of 52 visual fields from both malignant and benign areas across  
127 the entire set of the WSIs were selected in order to cover as wide a vari-  
128 ety of tissue architectures as possible. An expert pathologist (DRJS) then  
129 graded each visual field as either ‘benign’ or ‘malignant’, according to the  
130 overall glandular architecture. The pathologist also delineated the boundary  
131 of each individual glandular object on that visual field. We used this manual  
132 annotation as ground truth for automatic segmentation. Note that different  
133 glandular objects in an image may be part of the same gland. This is because  
134 a gland is a 3-dimensional structure that can appear as separated objects on  
135 a single tissue section. The visual fields were further separated into smaller,  
136 non-overlapping images, whose histologic grades (i.e. benign or malignant)  
137 were assigned the same value as the larger visual field. Representative exam-  
138 ple images of the two grades can be seen in Figure 1. This dataset was also  
139 previously used in the gland segmentation study by Sirinukunwattana et al.  
140 (2015).

141 In the challenge, the dataset was separated into **Training Part**, **Test**  
142 **Part A**, and **Test Part B**. Note that the data were stratified according  
143 to the histologic grade and the visual field before splitting. This was done  
144 to ensure that none of the images from the same visual field appears in  
145 different parts of the dataset (i.e. Training, Test Part A, or Test Part B).  
146 However, since the data were not stratified based on patient, different visual

---

<sup>2</sup>The T in TNM cancer staging refers to the spread of the primary tumour. In colorectal cancer, stage T3 means the tumour has grown into the outer lining of the bowel wall, whereas stage T4 means the tumour has grown through the outer lining of the bowel wall. The cancer stage is different from the tumour histologic grade, as the latter indicates the aggressiveness of the tumour.

147 fields from the same slide can appear in different parts of the dataset. A  
148 breakdown of the details of the dataset is shown in Table 1. The ground  
149 truth as well as the histologic grade which reflects morphology of glandular  
150 structures were provided for every image in the Training Part at the time of  
151 release. We used Test Part A and Test Part B as off-site and on-site test  
152 datasets respectively. Furthermore, to ensure blindness of evaluation, the  
153 ground truth and histologic grade of each image in the test parts were not  
154 released to the participants.

#### 155 4. Challenge Organization

156 The GlaS challenge contest was officially launched by the co-organizers  
157 (KS, JPWP, DRJS, NMR) on April 21<sup>st</sup>, 2015, and was widely publicized  
158 through several channels. At the same point, a challenge website<sup>3</sup> was set up  
159 to disseminate challenge-related information and to serve as a site for reg-  
160 istration, submission of results, and communication between the organizers  
161 and contestants. The challenge involved 4 stages, as detailed below:

162 *Stage 1: Registration and Release of the Training Data.* The registration  
163 was open for a period of about two months (April 21<sup>st</sup> to June 30<sup>th</sup>, 2015).  
164 Interested individuals or groups of up to 3 people that were affiliated with an  
165 academic institute or an industrial organization could register and download  
166 the training data (Training Part, see Section 3 for details) to start developing  
167 their gland segmentation algorithms. From this point forward, we will refer  
168 to a separate individual or a group of registrants as a ‘team’.

169 *Stage 2: Submission of a Short Paper.* In order to gain access to the first  
170 part of the test data, each registered team was required to submit a 2-page  
171 document containing a general description of their segmentation algorithms  
172 and some preliminary results obtained from running each algorithm on the  
173 training data. Each team could submit up to 3 different methods. The  
174 intention of this requirement was for the organizers to identify teams who  
175 were serious about participating in the challenge. The organizers based their  
176 reviews on two criteria: clarity of the method description and soundness of  
177 the validation strategy. Segmentation performance was not considered in this  
178 review. The submission of this document was due by July 17<sup>th</sup>, 2015.

---

<sup>3</sup><http://www.warwick.ac.uk/bialab/GlaScontest>

179 *Stage 3: Release of the Test Data Part A and Submission of Segmentation*  
180 *Results.* The first part of the test data (Test Part A, see Section 3 for de-  
181 tails) was released on August 14<sup>th</sup>, 2015 to those teams selected from the  
182 previous stage which also agreed to participate in the GlaS contest. The  
183 teams were given a month to further adjust and optimize their segmentation  
184 algorithms, and carry out segmentation on Part A of the test data. Each  
185 team could hand-in up to 3 sets of results per method submitted in Stage  
186 2. The submission of the segmentation results was due by September 14<sup>th</sup>,  
187 2015. Evaluation of the submitted results was not disclosed to the teams  
188 until after the challenge event.

189 *Stage 4: GlaS'2015 Challenge Event.* The event was held in conjunction  
190 with MICCAI'2015 on October 5<sup>th</sup>, 2015. All teams were asked to produce  
191 segmentation results on the second part of the test data (Test Part B, see Sec-  
192 tion 3) within 45 minutes. The teams could either bring their own machines  
193 or conduct an experiment remotely. There was no restriction on the num-  
194 ber of machines that the teams could use to produce results. Those teams  
195 that could not be present at the event provided implementations of their  
196 algorithms with which the organizers carried out the segmentation on their  
197 behalf. Each team was also asked to give a short presentation, discussing  
198 their work. At the end of the event, the complete evaluation of segmentation  
199 results across both parts of the test data was announced, which included a  
200 final ranking of the submitted methods. This information is also available  
201 on the challenge website.

#### 202 *4.1. Challenge Statistics*

203 By the end of Stage 1, a total of 110 teams from different academic and  
204 industrial institutes had registered. A total of 21 teams submitted the 2-page  
205 document for review in Stage 2, and 20 teams were invited to participate in  
206 the GlaS competition event. In Stage 3, only 13 teams submitted results  
207 on Part A of the test data in time. Late entries were neither evaluated nor  
208 considered in the next stage of the competition. On the day of the challenge  
209 event, 11 of the 13 teams that submitted the results on time in Stage 3  
210 attended the on-site competition and presented their work. The organizers  
211 carried out the segmentation on behalf of the other two teams that could not  
212 be present.

## 213 5. Evaluation

214 The performance of each segmentation algorithm was evaluated based on  
215 three criteria: 1) accuracy of the detection of individual glands; 2) volume-  
216 based accuracy of the segmentation of individual glands; and 3) boundary-  
217 based similarity between glands and their corresponding segmentation. It  
218 may seem that volume-based segmentation accuracy would entail boundary-  
219 based segmentation accuracy between a gland and its segmentation. How-  
220 ever, in practice, this is not always the case. The volume-based metric for  
221 segmentation accuracy used in this challenge, was defined and calculated us-  
222 ing the label that the algorithm had assigned to each pixel, but the boundary-  
223 based metric used the position assigned by the algorithm to the boundary  
224 of each gland. Pixels labels may be fairly accurate, while the boundary  
225 curves are very different. The remainder of this section describes all metrics  
226 employed in the evaluation.

227 We use the concept of a pair of corresponding segmented and ground  
228 truth objects as proposed in Sirinukunwattana et al. (2015). Let  $\mathcal{S}$  denote a  
229 set of all segmented objects and  $\mathcal{G}$  denote a set of all ground truth objects.  
230 We also include in each of these sets the empty object  $\emptyset$ . We define a function  
231  $G_* : \mathcal{S} \rightarrow \mathcal{G}$ , by setting, for each segmented object  $S \in \mathcal{S}$ ,  $G_*(S) = G \in \mathcal{G}$   
232 where  $G$  has the largest possible overlapping area with  $S$ . Although there  
233 could be more than one  $G \in \mathcal{G}$  that maximally overlaps  $S$ , this in practice  
234 is extremely rare, and it is good enough to consider one of these  $G$  as the  
235 value of  $G_*(S)$ . If there is no overlapping  $G$ , we set  $G_*(S) = \emptyset$ . (However, in  
236 the context of Hausdorff distance – see Section 5.3 –  $G_*$  will be extended in  
237 a different way.) Similarly, we define  $S_* : \mathcal{G} \rightarrow \mathcal{S}$ , by setting, for each  $G \in \mathcal{G}$ ,  
238  $S_*(G) = S \in \mathcal{S}$ , where  $S$  has the largest possible overlapping area with  $G$ .  
239 Note that  $G_*$  and  $S_*$  are, in general, neither injective, nor surjective. Nor  
240 are they inverse to each other, in general. They do, however, assign to each  
241  $G$  an  $S = S_*(G)$ , and to each  $S$  a  $G = G_*(S)$ .

### 242 5.1. Detection Accuracy

243 The F1 score is employed to measure the detection accuracy of individual  
244 glandular objects. A segmented glandular object that intersects with at  
245 least 50% of its ground truth object is counted as true positive, otherwise it  
246 is counted as false positive. The number of false negatives is calculated as  
247 the difference between the number of ground truth objects and the number

248 of true positives. Given these definitions, the F1 score is defined by

$$\text{F1score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \quad (1)$$

249 where

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2)$$

250 and TP, FP, and FN denote respectively the number of true positives, false  
251 positives, and false negatives from all images in the dataset.

## 252 5.2. Volume-Based Segmentation Accuracy

### 253 5.2.1. Object-Level Dice Index

254 The Dice index (Dice, 1945) is a measure of agreement or similarity be-  
255 tween two sets of samples. Given  $G$ , a set of pixels belonging to a ground  
256 truth object, and  $S$ , a set of pixels belonging to a segmented object, the Dice  
257 index is defined as follows:

$$\text{Dice}(G, S) = \frac{2|G \cap S|}{|G| + |S|}, \quad (3)$$

258 where  $|\cdot|$  denotes set cardinality. The index ranges over the interval  $[0, 1]$ ,  
259 where the higher the value, the more concordant the segmentation result  
260 and the ground truth. A Dice index of 1 implies a perfect agreement. It  
261 is conventional that the segmentation accuracy on an image is calculated by  
262  $\text{Dice}(G_{\text{all}}, S_{\text{all}})$ , where  $G_{\text{all}}$  denotes the set of pixels of all ground truth objects  
263 and  $S_{\text{all}}$  denotes the set of pixels of all segmented objects. The calculation  
264 made in this way measures the segmentation accuracy only at the pixel level,  
265 not at the gland level, which was the main focus of the competition.

266 To take the notion of an individual gland into account, we employ the  
267 object-level Dice index (Sirinukunwattana et al., 2015). Let  $n_{\mathcal{G}}$  be the number  
268 of non-empty ground truth glands, as annotated by the expert pathologist.  
269 Similarly let  $n_{\mathcal{S}}$  be the number of glands segmented by the algorithm, that  
270 is the number of non-empty segmented objects. Let  $G_i \in \mathcal{G}$  denote the  $i^{\text{th}}$   
271 ground truth object, and let  $S_j \in \mathcal{S}$  denote the  $j^{\text{th}}$  segmented object. The  
272 object-level Dice index is defined as

$$\text{Dice}_{\text{obj}}(\mathcal{G}, \mathcal{S}) = \frac{1}{2} \left[ \sum_{i=1}^{n_{\mathcal{G}}} \gamma_i \text{Dice}(G_i, S_*(G_i)) + \sum_{j=1}^{n_{\mathcal{S}}} \sigma_j \text{Dice}(G_*(S_j), S_j) \right], \quad (4)$$

273 where

$$\gamma_i = |G_i| / \sum_{p=1}^{n_G} |G_p|, \quad \sigma_j = |S_j| / \sum_{q=1}^{n_S} |S_q| \quad (5)$$

274 On the right hand side of (4), the first summation term reflects how well each  
 275 ground truth object overlaps its segmented object, and the second summa-  
 276 tion term reflects how well each segmented object overlaps its ground truth  
 277 objects. Each term is weighted by the relative area of the object, giving less  
 278 emphasis to small segmented and small ground truth objects.

279 In the competition, the object-level Dice index of the whole test dataset  
 280 was calculated by including all the ground truth objects from all images in  
 281  $\mathcal{G}$  and all the segmented objects from all images in  $\mathcal{S}$ .

### 282 5.2.2. Adjusted Rand Index

283 We also included the adjusted Rand index (Hubert and Arabie, 1985) as  
 284 another evaluation measure of segmentation accuracy. This index was used  
 285 for additional assessment of the algorithm performance in Section 8.3.

286 The adjusted Rand index measures similarity between the set of all ground  
 287 truth objects  $\mathcal{G}$  and the set of all segmented objects  $\mathcal{S}$ , based on how pixels  
 288 in a pair are labeled. Two possible scenarios for the pair to be concordant  
 289 are that (i) they are placed in the same ground truth object in  $\mathcal{G}$  and the  
 290 same segmented object in  $\mathcal{S}$ , and (ii) they are placed in different ground  
 291 truth objects in  $\mathcal{G}$  and in different segmented objects in  $\mathcal{S}$ . Define  $n_{ij}$  as the  
 292 number of pixels that are common to both the  $i^{\text{th}}$  ground truth object and  
 293 the  $j^{\text{th}}$  segmented object,  $n_{i,\cdot}$  as the total number of pixels in the  $i^{\text{th}}$  ground  
 294 truth object,  $n_{\cdot,j}$  as the total number of pixels in the  $j^{\text{th}}$  segmented object,  
 295 and  $n$  as the total number of pixels. Following a simple manipulation, it can  
 296 be shown that the probability of agreement is equal to

$$P_{\text{agreement}} = \left[ \binom{n}{2} + 2 \sum_{i=1}^{n_G} \sum_{j=1}^{n_S} \binom{n_{ij}}{2} - \sum_{i=1}^{n_G} \binom{n_{i,\cdot}}{2} - \sum_{j=1}^{n_S} \binom{n_{\cdot,j}}{2} \right] / \binom{n}{2}. \quad (6)$$

297 Here, the numerator term corresponds to the total number of agreements,  
 298 while the denominator term corresponds to the total number of all possi-  
 299 ble pairs of pixels. Under the assumption that the partition of pixels into  
 300 ground truth objects in  $\mathcal{G}$  and segmented objects in  $\mathcal{S}$  follows a generalized

301 hypergeometric distribution, the adjusted Rand index can be formulated as

$$\text{ARI}(\mathcal{G}, \mathcal{S}) = \frac{\sum_{i=1}^{n_{\mathcal{G}}} \sum_{j=1}^{n_{\mathcal{S}}} \binom{n_{i,j}}{2} - \sum_{i=1}^{n_{\mathcal{G}}} \binom{n_i}{2} \sum_{j=1}^{n_{\mathcal{S}}} \binom{n_{\cdot,j}}{2} / \binom{n}{2}}{\frac{1}{2} \left[ \sum_{i=1}^{n_{\mathcal{G}}} \binom{n_{i,\cdot}}{2} + \sum_{j=1}^{n_{\mathcal{S}}} \binom{n_{\cdot,j}}{2} \right] - \sum_{i=1}^{n_{\mathcal{G}}} \binom{n_{i,\cdot}}{2} \sum_{j=1}^{n_{\mathcal{S}}} \binom{n_{\cdot,j}}{2} / \binom{n}{2}}}. \quad (7)$$

302 The adjusted Rand index is bounded above by 1, and it can be negative.

### 303 5.3. Boundary-Based Segmentation Accuracy

304 We measure the boundary-based segmentation accuracy between the seg-  
 305 mented objects in  $\mathcal{S}$  and the ground truth objects in  $\mathcal{G}$  using the object-level  
 306 Hausdorff distance. The usual definition of a Hausdorff distance between  
 307 ground truth object  $G$  and segmented object  $S$  is

$$H(G, S) = \max\left\{\sup_{x \in G} \inf_{y \in S} d(x, y), \sup_{y \in S} \inf_{x \in G} d(x, y)\right\} \quad (8)$$

308 where  $d(x, y)$  denotes the distance between pixels  $x \in G$  and  $y \in S$ . In this  
 309 work, we use the Euclidean distance. According to (8), Hausdorff distance is  
 310 the most extreme value from all distances between the pairs of nearest pixels  
 311 on the boundaries of  $S$  and  $G$ . Thus, the smaller the value of the Hausdorff  
 312 distance, the higher the similarity between the boundaries of  $S$  and  $G$ , and  
 313  $S = G$  if their Hausdorff distance is zero.

314 To calculate the overall segmentation accuracy between a pair of corre-  
 315 sponding segmented and ground truth objects, we now introduce object-level  
 316 Hausdorff distance by imitating the definition of object-level Dice index (4).  
 317 The object-level Hausdorff distance is defined as

$$H_{\text{obj}}(\mathcal{G}, \mathcal{S}) = \frac{1}{2} \left[ \sum_{i=1}^{n_{\mathcal{G}}} \gamma_i H(G_i, S_*(G_i)) + \sum_{j=1}^{n_{\mathcal{S}}} \sigma_j H(G_*(S_j), S_j) \right], \quad (9)$$

318 where the meaning of the mathematical notation is similar to that given in  
 319 Section 5.2.1. In case a ground truth object  $G$  does not have a corresponding  
 320 segmented object (i.e.  $S_*(G) = \emptyset$ ), the Hausdorff distance is calculated  
 321 between  $G$  and the nearest segmented object  $S \in \mathcal{S}$  to  $G$  (in the Hausdorff  
 322 distance) in that image instead. The same applies for a segmented object  
 323 that does not have a corresponding ground truth object.

## 324 6. Ranking Scheme

325 Each submitted entry was assigned one ranking score per evaluation met-  
326 ric and set of test data. Since there were 3 evaluation metrics (F1 score  
327 for gland detection, object-level Dice index for volume-based segmentation  
328 accuracy, and object-level Hausdorff index for boundary-based segmentation  
329 accuracy) and 2 sets of test data, the total number of ranking scores was  
330 6. The best performing entry was assigned ranking score 1, the second best  
331 was assigned ranking score 2, and so on. In case of a tie, the standard com-  
332 petition ranking was applied. For instance, F1 score 0.8, 0.7, 0.7, and 0.6  
333 would result in the ranking scores 1, 2, 2, and 4. The final ranking was then  
334 obtained by adding all 6 ranking scores (rank sum). The entry with smallest  
335 sum was placed top in the final ranking.

## 336 7. Methods

337 The top ranking methods are described in this section. They are selected  
338 from the total of 13 methods that participated in all stages of the challenge.  
339 The cut-off for the inclusion in this section was made where there was a  
340 substantial gap in the rank sums (see Appendix A, Figure A.5). Of the 7  
341 selected methods, only 6 preferred to have their methods described here.

### 342 7.1. *CUMedVision*<sup>4</sup>

343 A novel deep contour-aware network (Chen et al., 2016) was presented.  
344 This method explored the multi-level feature representations with fully con-  
345 volutional networks (FCN) (Long et al., 2015). The network outputted seg-  
346 mentation probability maps and depicted the contours of gland objects simul-  
347 taneously. The network architecture consisted of two parts: a down-sampling  
348 path and an up-sampling path. The down-sampling path contained convo-  
349 lutional and max-pooling layers while the up-sampling path contained con-  
350 volutional and up-sampling layers, which increased the resolutions of feature  
351 maps and outputted the prediction masks. In total, there were 5 max-pooling  
352 layers and 3 up-sampling layers. Each layer with learned parameters was fol-  
353 lowed by a non-linear mapping layer (element-wise rectified linear activation).

---

<sup>4</sup>Department of Computer Science and Engineering, The Chinese University of Hong Kong.

354 In order to separate touching glands, the feature maps from hierarchical  
355 layers were up-sampled with two different branches to output the segmented  
356 object and contour masks respectively. The parameters of the down-sampling  
357 path were shared and updated for these two kinds of masks. This could  
358 be viewed as a multi-task learning framework with feature representations,  
359 simultaneously encoding the information of segmented objects and contours.  
360 To alleviate the problem of insufficient training data (Chen et al., 2015),  
361 an off-the-shelf model from DeepLab (Chen et al., 2014), trained on the  
362 2012 PASCAL VOC dataset<sup>5</sup>, was used to initialize the weights for layers in  
363 the down-sampling path. The parameters of the network were obtained by  
364 minimizing the loss function with standard back-propagation<sup>6</sup>.

365 The team submitted two entries for evaluation. **CUMedVision1** was  
366 produced by FCN with multi-level feature representations relying only on  
367 gland object masks, while **CUMedVision2** was the results of the deep  
368 contour-aware network, which considers gland object and contour masks si-  
369 multaneously.

## 370 7.2. CVML<sup>7</sup>

371 In the first, preprocessing, stage the images were corrected to compensate  
372 for variations in the appearance due to a variability of the tissue staining pro-  
373 cess. This was implemented through histogram matching, where the target  
374 histogram was calculated from the whole training data, and the individual  
375 image histograms were used as inputs. The main processing stage was based  
376 on two methods: a convolutional neural network (CNN) (Krizhevsky et al.,  
377 2012) for a supervised pixel classification, and a level set segmentation for  
378 grouping pixels into spatially coherent structures. The employed CNN used  
379 an architecture with two convolutional, pooling and fully connected layers.  
380 The network was trained with three target classes. The classes were designed  
381 to represent (1) the tubular interior of the glandular structure (inner class),  
382 (2) epithelial cells forming boundary of the glandular structure (boundary  
383 class) and (3) inter-gland tissue (outer class). The inputs to the CNN were  
384  $19 \times 19$  pixel patches sliding across the adjusted RGB input image. The two  
385 convolutional layers used  $6 \times 6$  and  $4 \times 4$  kernels with 16 and 36 feature maps

---

<sup>5</sup><http://host.robots.ox.ac.uk:8080/pascal/VOC/voc2012/index.html>

<sup>6</sup>More details will be available at: [http://www.cse.cuhk.edu.hk/~hchen/research/2015miccai\\_gland.html](http://www.cse.cuhk.edu.hk/~hchen/research/2015miccai_gland.html)

<sup>7</sup>School of Engineering, University of Central Lancashire, Preston, UK.

386 respectively. The pooling layers, implementing the mean function, used  $2 \times 2$   
387 receptive fields and  $2 \times 2$  stride. The first and second fully connected layers  
388 used the rectified linear unit and softmax functions respectively. The outputs  
389 from the CNN were two probability maps representing the probability of each  
390 image pixel belonging to the inner and boundary classes. These two prob-  
391 ability maps were normalized between -1 and 1 and used as a propagation  
392 term, along with an advection term and a curvature flow term. These terms  
393 were part of the hybrid level set model described in Zhang et al. (2008). In  
394 the post-processing stage, a sequence of morphological operations was per-  
395 formed to removed small objects, fill holes and disconnect weakly connected  
396 objects. Additionally, if an image boundary intersecting an object forms a  
397 hole, the corresponding pixels was labeled as part of that object. The team  
398 submitted a single entry for evaluation, henceforth referred to as **CVML**.

### 399 7.3. *ExB*<sup>8</sup>

400 This method first preprocessed the data by performing per channel zero  
401 mean and unit variance normalization, where the mean and variance were  
402 computed from the training data. The method then exploited the local invari-  
403 ance properties of the task by applying a set of transformations to the data.  
404 At training time, the dataset was augmented by applying affine transforma-  
405 tions, Gaussian blur and warping. During testing, both image mirroring and  
406 rotation were applied.

407 The main segmentation algorithm consisted of a multi-path convolutional  
408 neural network. Each path was equipped with a different set of convolutional  
409 layers and configured to capture features from different views in a local-global  
410 fashion. All the different paths were connected to a set of two fully connected  
411 layers. A leaky rectified linear unit was used as a default activation function  
412 between layers, and a softmax layer was used after the last fully connected  
413 layer. Every network was trained via stochastic gradient descent with mo-  
414 mentum, using a step-wise learning rate schedule (Krizhevsky et al., 2012).  
415 The network was randomly initialized such that unit variance was preserved  
416 across layers. It was found that using more than three paths led to heavy  
417 over-fitting – this was due to insufficient training data.

418 Simple-path networks were trained to detect borders of glands. The  
419 ground truth for these networks was constructed using a band of width

---

<sup>8</sup>ExB Research and Development.

420  $K \in [5, 10]$  pixels along a real gland border. These values of  $K$  were found  
421 to produce optimal and equivalent quantitative results, measured by the F1  
422 score and the object-Dice index. The output of these networks was used to  
423 better calibrate the final prediction.

424 In the post-processing step, a simple method was applied to clean noise  
425 and fill holes in the structures. Thresholding was applied to remove spurious  
426 structures with diameter smaller than a certain epsilon. Filling-hole criteria  
427 based on diameter size was also used.

428 Using the initial class discrimination (benign and malignant), a simple  
429 binary classifier constructed from a convolutional neural network with 2 con-  
430 volutional and 1 fully connected layers was trained. This binary classifier  
431 used the raw image pixels as input. The output of the classifier was used  
432 together with the border networks and the post-processing method to apply  
433 a different set of parameters/thresholds depending on the predicted class.  
434 The hyperparameters for the entire pipeline, including post-processing and  
435 border networks, were obtained through cross-validation.

436 For this method, the team submitted 3 entries. **ExB 1** was a two-path  
437 network including both the border network for detecting borders of glands  
438 and the binary classification to differentiate between the post-processing pa-  
439 rameters. **ExB 2** was similar to ExB 1 without the use of the border network.  
440 **ExB 3** used a two-path network without any post-processing.

#### 441 7.4. *Image Analysis Lab Uni Freiburg*<sup>9</sup>

442 The authors applied a u-shaped deep convolutional network “u-net”<sup>10</sup>  
443 (Ronneberger et al., 2015) for the segmentation. The input was the raw  
444 RGB image and the output was a binary segmentation map (glands and  
445 background). The network consisted of an analysis-path constructed from a  
446 sequence of convolutional layers and max-pooling layers, followed by a synthe-  
447 sis path with a sequence of up-convolutional layers and convolutional layers,  
448 resulting in 23 layers in total. Additional shortcut-connections propagated  
449 the feature maps at all detail levels from the analysis to the synthesis path.  
450 The network was trained from scratch in an end-to-end fashion with only the

---

<sup>9</sup>Computer Science Department and BIOS Centre for Biological Signalling Studies, University of Freiburg, Germany.

<sup>10</sup>The implementation of the u-net is freely available at <http://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>.

451 images and ground truth segmentation maps provided by the challenge orga-  
452 nizers. To teach the network the desired invariances and to avoid overfitting,  
453 the training data were augmented with randomly transformed images and  
454 the correspondingly transformed segmentation maps. The applied transfor-  
455 mations were random elastic deformations, rotations, shifts, flips, and blurs.  
456 The color transformations were random multiplications applied in the HSV  
457 color space. To avoid accidentally joining touching objects, a high pixel-wise  
458 loss weight was introduced for pixels in thin gaps between objects in the  
459 training dataset (see Ronneberger et al. (2015)). The exact same u-net lay-  
460 out with the same hyperparameters as in Ronneberger et al. (2015) was used  
461 for the challenge. The only difference were more training iterations and a  
462 slower decay of the learning rate.

463 The team submitted two entries. The first entry **Freiburg1** was a con-  
464 nected component labelling applied to the raw network output. The second  
465 entry **Freiburg2** post-processed the segmentation maps with morphological  
466 hole-filling and deletion of segments smaller than 1000 pixels.

#### 467 7.5. LIB<sup>11</sup>

468 Intestinal glands were divided according to their appearance into three  
469 categories: hollow, bounded, and crowded. A hollow gland was composed of  
470 lumen and goblet cells and it could be a hole in the tissue surface. A bounded  
471 gland had the same composition, but in addition, it was surrounded by a thick  
472 epithelial layer. A crowded gland was composed of bunches of epithelial cells  
473 clustered together and it might have shown necrotic debris.

474 The tissue was first classified into one of the above classes before beginning  
475 the segmentation. The classification relied on the characterization of the  
476 spatial distribution of cells and the topology of the tissue. Therefore, a  
477 closing map was generated with a cumulative sum of morphological closing  
478 by a disk of increasing radius (1 to 40 pixels) on the binary image of nuclear  
479 objects, which were segmented by the  $k$ -means algorithm in the RGB colour  
480 space. The topological features were calculated from a normalized closing  
481 map in MSER fashion (Maximally Stable Extremal Region, Matas et al.  
482 (2004)) as the number of regions below three different thresholds (25%, 50%  
483 and 62.5%) and above one threshold (90%), their sizes and the mean of

---

<sup>11</sup>Sorbonne Universités, UPMC Univ Paris 06, CNRS, INSERM, Biomedical Imaging Laboratory (LIB), Paris, France.

484 their corresponding values in the closing map. The first three thresholds  
485 characterized the holes and the fourth one characterized the thickness of  
486 nuclear objects. After classifying the tissue with a Naive Bayes classifier  
487 trained on these features, a specific segmentation algorithm was applied.

488 Three segmentation algorithms were presented, one for each category.  
489 Hollow glands were delineated by morphological dilation on regions below  
490 50%. Bounded gland candidates were first detected as hollow glands, then  
491 the thickness of nuclear objects surrounding the region was evaluated by gener-  
492 ating a girth map and a solidity map (Ben Cheikh et al., 2016), then after  
493 classifying nuclear objects, the epithelial layer was added or the candidate  
494 was removed. Crowded glands were identified as populous regions (regions  
495 above 90%), and then morphological filtering was applied for refinement. The  
496 team submitted a single entry labeled as **LIB** for evaluation.

#### 497 *7.6. vision4GlaS*<sup>12</sup>

498 Given an H&E-stained RGB histopathological section, the gland segmen-  
499 tation method was based on a pixel-wise classification and an active contour  
500 model, and it proceeded in three steps (Kainz et al., 2015). In a first prepro-  
501 cessing step the image was rescaled to half the spatial resolution, and color  
502 deconvolution separated the stained tissue components. The red channel of  
503 the deconvolved RGB image represented the tissue structure best and was  
504 therefore considered for further processing. Next, two convolutional neu-  
505 ral networks (CNNs) (LeCun et al., 2010) of seven layers each were trained  
506 for pixel-wise classification on a set of image patches. Each network was  
507 trained with ReLU nonlinearities, and stochastic gradient descent with mo-  
508 mentum, weight decay, and dropout regularization to minimize a negative  
509 log-likelihood loss function. The first CNN, called Object-Net, was trained  
510 to distinguish four classes: (i) benign background, (ii) benign gland, (iii)  
511 malignant background, and (iv) malignant gland. For each image patch the  
512 probability distribution over the class labels was predicted, using a softmax  
513 function. The Object-Net consisted of three convolutional layers followed  
514 by max-pooling, a final convolutional layer and three fully connected layers.

---

<sup>12</sup>Institute of Biophysics, Center for Physiological Medicine, Medical University of Graz,  
Graz, Austria; Institute of Neuroinformatics, University of Zurich and ETH Zurich, Zurich,  
Switzerland; Institute for Computer Graphics and Vision, BioTechMed, Graz University  
of Technology, Graz, Austria; Ludwig Boltzmann Institute for Clinical Forensic Imaging,  
Graz, Austria.

515 The second – architecturally similar – CNN called Separator-Net, learned to  
516 predict pixels of gland-separating structures in a binary classification task.  
517 Ground truth was generated by manually labeling image locations, close  
518 to two or more gland borders, as gland-separating structures. In the final  
519 step the segmentation result was obtained by combining the outputs of the  
520 two CNNs. Predictions for benign and malignant glands were merged, and  
521 predictions of gland-separating structures were subtracted to emphasize the  
522 foreground probabilities. Background classes were handled similarly. Using  
523 these refined foreground and background maps, a figure-ground segmentation  
524 based on weighted total variation was employed to find a globally optimal  
525 solution. This approach optimized a geodesic active contour energy, which  
526 minimized contour length while adhering to the refined CNN predictions  
527 (Bresson et al., 2007). The team submitted a single entry, referred to as  
528 **vision4GlaS**.

## 529 **8. Results and Discussion**

### 530 *8.1. Summary of the Methods*

531 The methods described above take one of the following two approaches  
532 to segmentation: (a) they start by identifying pixels corresponding to glands  
533 which are then grouped together to form separated, spatially coherent ob-  
534 jects; (b) they begin with candidate objects that are then classified as glands  
535 or non-glands. All methods that are based on CNNs (CUMedVision, CVML,  
536 ExB, Freiburg, and vision4GlaS) follow the former approach. CVML, ExB,  
537 and vision4GlaS built CNN classifiers that assign a gland-related or non-  
538 gland-related label to every pixel in an image, by taking patch(es) centered  
539 at the pixel as input. ExB, in particular, use multi-path networks into which  
540 patches at different sizes are fed, in order to capture contextual informa-  
541 tion at multiple scales. CUMedVision and Freiburg, on the other hand, base  
542 their pixel classifier on a fully convolutional network architecture (Long et al.,  
543 2015), allowing simultaneous pixel-wise label assignment at multiple pixel lo-  
544 cations. To separate gland-related pixels into individual objects, CVML and  
545 vision4GlaS deploy contour based approaches. ExB trains additional net-  
546 works for glandular boundary, while CUMedVision and Freiburg explicitly  
547 include terms for boundary in the training loss function of their networks.  
548 The only method that follows the latter approach for object segmentation  
549 is LIB. In this method, candidate objects forming part of a gland (i.e., lu-

550 men, epithelial boundary) are first identified, and then classified into different  
551 types, followed by the final step of segmentation.

552 A variety of data transformation and augmentation were employed to deal  
553 with variation within the data. In order to counter the effect of stain vari-  
554 ation, CVML and ExB performed transformations of the RGB color chan-  
555 nels, vision4GlaS used a stain deconvolution technique to obtain only the  
556 basophilic channel in their preprocessing step. By contrast, Freiburg tackled  
557 the issue of stain variability through data augmentation, which implicitly  
558 forces the networks to be robust to stain variation to some extent. As is  
559 common among methods using CNNs, spatial transformations, such as affine  
560 transformations (e.g. translation, rotation, flip), elastic deformations (e.g.  
561 pincushion and barrel distortions), and blurring, were also used in the data  
562 augmentation to teach the network to learn features that are spatially invari-  
563 ant. The other benefit of data augmentation is it provides, to some extent,  
564 avoidance of over-fitting.

565 ExB, LIB, and vision4GlaS incorporated histologic grades of glands in  
566 their segmentation approach. In ExB, procedures and/or parameter values  
567 used in boundary detection and post-processing steps were different, subject  
568 to the predicted histologic grade of an image. vision4GlaS classified pixels  
569 based on histological information. Although not explicit, LIB categorized  
570 candidate objects forming glands according to their appearance, related to  
571 histologic grades, before treating them in different ways.

572 As a post-processing step, many segmentation algorithms employed sim-  
573 ple criteria and/or a sequence of morphological operations to improve their  
574 segmentation results. A common treatment was to eliminate small spurious  
575 segmented objects. Imperfections in pixel labelling can result in the appear-  
576 ance of one or more holes in the middle of an object. Filling such holes is often  
577 necessary. In addition to these operations, CVML performed morphological  
578 operations to separate accidentally joined objects.

## 579 *8.2. Evaluation Results*

580 Table 2 summarizes the overall evaluation scores and ranks achieved by  
581 each entry from each test part. We list the entries according to the order  
582 of their rank sum, which indicates the overall performance across evaluation  
583 measures and tasks of the entries. The lower the rank sum, the more favor-  
584 able the performance. The top three entries according to the overall rank  
585 sum in descending order are CUMedVision2, ExB1, and ExB3. However,

Table 2: Summary results. The evaluation is carried out according to the challenge criteria described in Section 6. A ranking score is assigned to each algorithm according to its performance in each evaluation measure, obtained from each test part. The entries are listed in a descending order based on their rank sum

Method	F1score				Dice <sub>obj</sub>				H <sub>obj</sub>				Rank Sum
	Part A		Part B		Part A		Part B		Part A		Part B		
	Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank	
CUMedVision2	0.912	1	0.716	3	0.897	1	0.781	5	45.418	1	160.347	6	17
ExB1	0.891	4	0.703	4	0.882	4	0.786	2	57.413	6	145.575	1	21
ExB3	0.896	2	0.719	2	0.886	2	0.765	6	57.350	5	159.873	5	22
Freiburg2	0.870	5	0.695	5	0.876	5	0.786	3	57.093	3	148.463	3	24
CUMedVision1	0.868	6	0.769	1	0.867	7	0.800	1	74.596	7	153.646	4	26
ExB2	0.892	3	0.686	6	0.884	3	0.754	7	54.785	2	187.442	8	29
Freiburg1	0.834	7	0.605	7	0.875	6	0.783	4	57.194	4	146.607	2	30
CVML	0.652	9	0.541	8	0.644	10	0.654	8	155.433	10	176.244	7	52
LIB	0.777	8	0.306	10	0.781	8	0.617	9	112.706	9	190.447	9	53
vision4GlaS	0.635	10	0.527	9	0.737	9	0.610	10	107.491	8	210.105	10	56

586 if rank sum is considered with respect to the test part, the three best entries are CUMedVision2, ExB2, and ExB3 for part A; whereas in part B, 587 CUMedVision1, ExB1, and Freiburg2 come at the top. A summary of the 588 ranking results from the competition can be found in Appendix A. Some segmentation results and their corresponding evaluation scores are illustrated in 589 Figure 2 to give a better idea of how the evaluation scores correlate with the 590 quality of the segmentation. 591 592

### 593 8.3. Additional Experiments

594 In the challenge, the split of the test data into two parts – Part A (60 595 images) for off-site test and Part B (20 images) for on-site test – to some 596 extent introduces bias into the performance evaluation of the segmentation 597 algorithms due to equal weight given to performance on the two test parts. 598 The algorithms that perform particularly well on Test Part B would therefore 599 get a better evaluation score even though they may not have performed as 600 well on Test Part A, where the majority of the test dataset is to be found. 601 In addition, the imbalance between the benign and malignant classes in Test 602 Part B, only 4 benign (20%) and 16 malignant (80%) images, would also favor 603 algorithms that perform well on the malignant class. In order to alleviate 604 these issues, we merged the two test parts and re-evaluated the performance 605 of all the entries. In addition, as suggested by one of the participating teams,

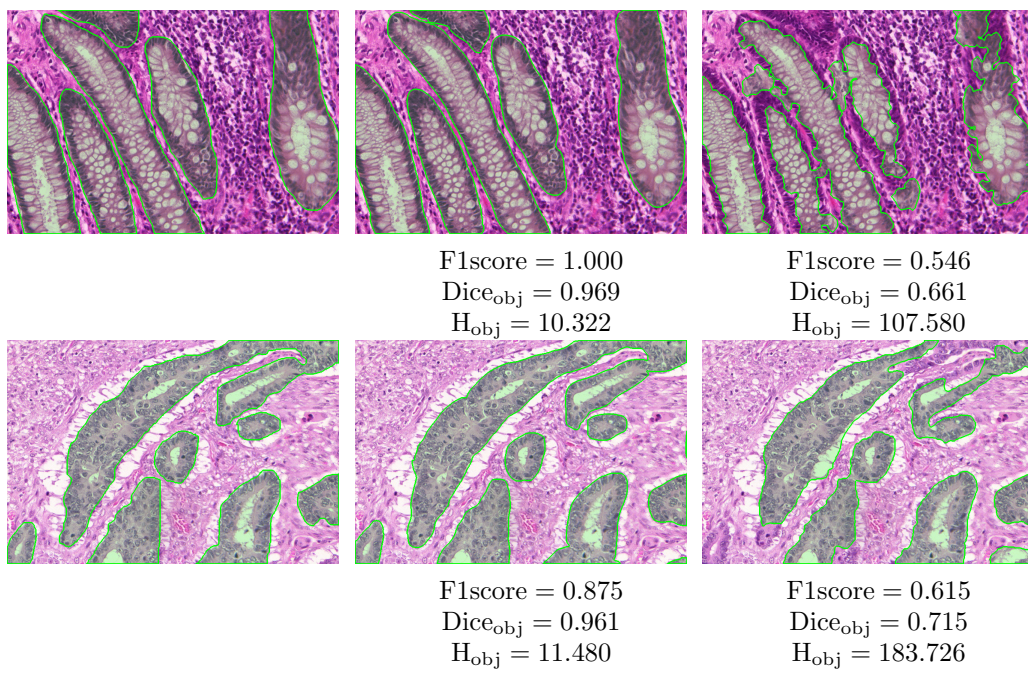


Figure 2: Example images showing segmentation results from some submitted entries. In each row, (left) ground truth, (middle) the best segmentation result, and (right) the worst segmentation result. For each image, the corresponding set of evaluation scores for the segmentation result is reported underneath the image.

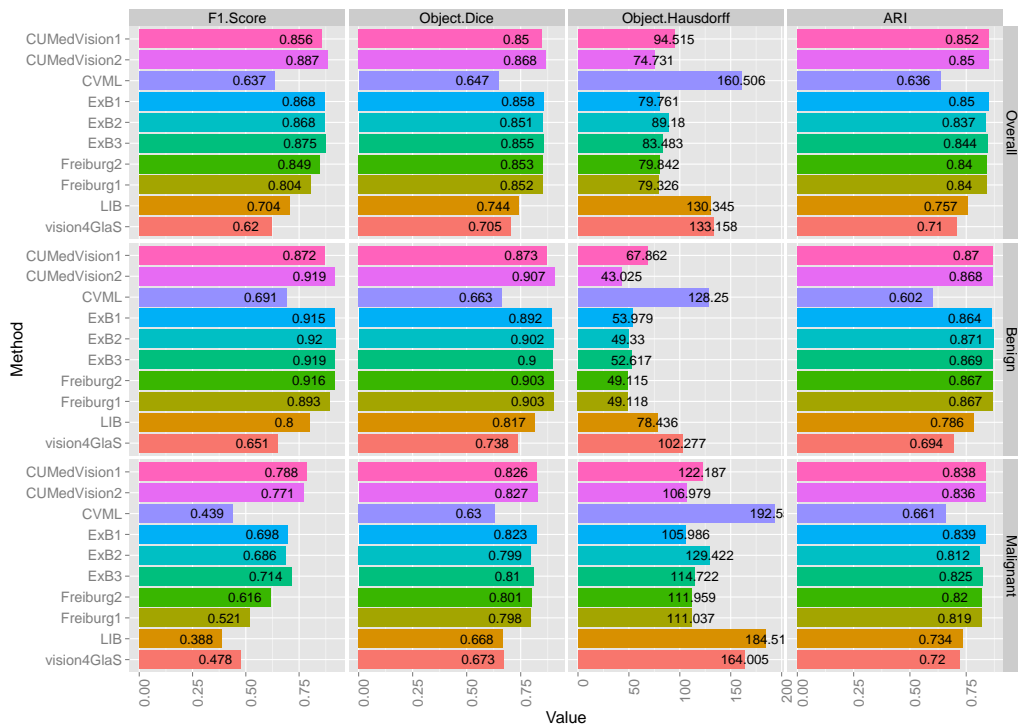


Figure 3: Performance scores achieved by different entries on the combined test data. Evaluation is conducted on three subsets of the data: (1st row) the whole test data, (2nd row) benign, and (3rd row) malignant.

Table 3: Ranking results of the entries when the two parts of test data are combined. Two set of ranking scheme are considered: a) F1score + Dice<sub>obj</sub>+H<sub>obj</sub> and b) F1score + ARI + H<sub>obj</sub>. In addition to the evaluation on the whole test data (overall), the entries are evaluated on a subset of the data according to the histologic labels, i.e. benign and malignant.

Entry	Final Ranking					
	F1score + Dice <sub>obj</sub> +H <sub>obj</sub>			F1score + ARI + H <sub>obj</sub>		
	Overall	Benign	Malignant	Overall	Benign	Malignant
CUMedVision1	7	7	3	4	6	3
CUMedVision2	1	1	1	1	2	2
CVML	10	10	10	10	10	10
ExB1	2	6	2	2	7	1
ExB2	6	3	7	7	1	7
ExB3	3	5	4	3	3	4
Freiburg1	4	4	6	6	5	6
Freiburg2	5	2	5	5	4	5
LIB	8	8	9	8	8	9
vision4GlaS	9	9	8	9	9	8

606 the adjusted Rand index is included as another performance measurement  
607 for segmentation.

608 The evaluation scores calculated from the combined two test parts are  
609 presented as bar chart in Figure 3. The final rankings based on the rank  
610 sums of evaluation scores calculated from the combined two test parts are  
611 reported in Table 3. Here, two set of rank sums are considered: one calculated  
612 according to the criteria of the competition (i.e., F1score+Dice<sub>obj</sub>+H<sub>obj</sub>), and  
613 the other where the adjusted Rand index is used instead of the object-level  
614 Dice index to evaluate segmentation accuracy (i.e., F1score + ARI + H<sub>obj</sub>).  
615 For both sets of rank sums, the new ranking orders are largely similar to  
616 those reported in Section 8.2, with a few swaps in the order, while the top  
617 three entries remaining the same, namely CUMedVision2, ExB1, ExB3.

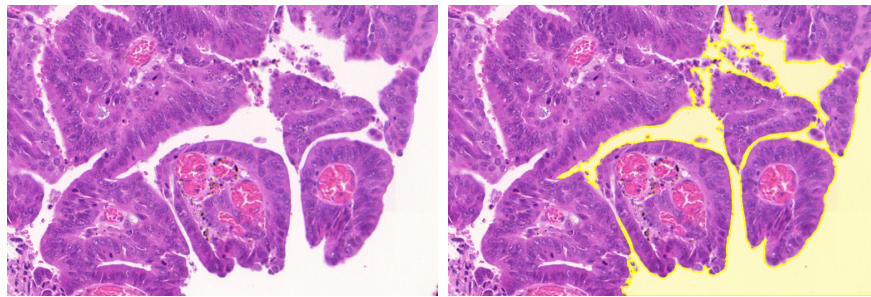
618 The main factors that negatively affect the performance of the methods  
619 are a number of challenges presented by the dataset. Firstly, large white  
620 empty areas corresponding to the lumen of the gastrointestinal tract which  
621 are not in the interior of intestinal glands can easily confuse the segmentation  
622 algorithms (Figure 4a). Secondly, characteristics of non-glandular tissue can

623 sometimes resemble that of the glandular tissue. For instance, connective tis-  
624 sue in muscularis mucosa or sub-mucosa layers of the colon is stained white  
625 and pinkish and has less dense nuclei, thus resembling the inner part of glands  
626 (Figure 4b). In the case where there is less stain contrast between nuclei and  
627 cytoplasm due to elevated levels of Hematoxylin stain, non-glandular tissue  
628 with dense nuclei can look similar to malignant epithelial tissue (Figure 4c).  
629 Thirdly, small glandular objects are blended into the surrounding tissue and  
630 can be easily mis-detected (Figure 4d). A careful inspection of the segmenta-  
631 tion results generated by each entry showed that methods by CUMedVision,  
632 ExB, and Freiburg better avoid over-segmentation or under-segmentation  
633 when facing the above-mentioned pitfalls.

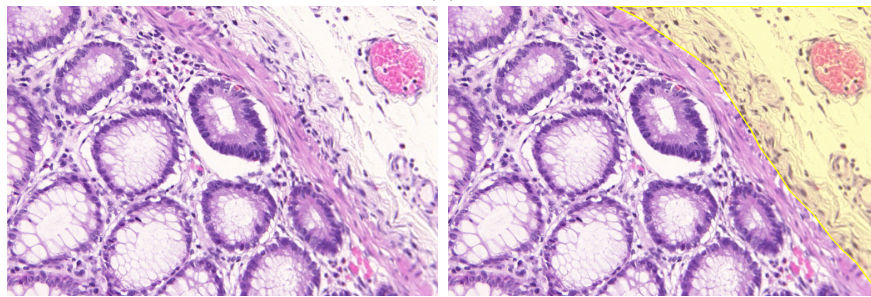
634 The performance of each entry with respect to the histologic grade of  
635 cancer was also examined. Their evaluation scores based on benign and  
636 malignant samples are reported in the second and the third rows of Figure  
637 3 respectively, and the ranking orders derived from the rank sums of the  
638 scores are shown in Table 3. Based on these results, one can get a better  
639 contrast between the performance of the entries that enforce border separa-  
640 tion and those that do not. By applying a predicted border mask to separate  
641 clumped segmented objects, CUMedVision2 performs better than CUMedVi-  
642 sion1, which tends to produce segmentation results that merge neighboring  
643 glands together, in both benign and malignant cases. Similarly, ExB1 is  
644 able to segment malignant glands better than ExB2 and ExB3 that do not  
645 utilize border separation. However, this can have an adverse effect if the al-  
646 gorithm already yields segmentation results that separate individual objects  
647 well, such as in the case of ExB1 which under-segments benign glandular  
648 objects as compared to its counterparts ExB2 and ExB3.

#### 649 *8.4. General Discussion*

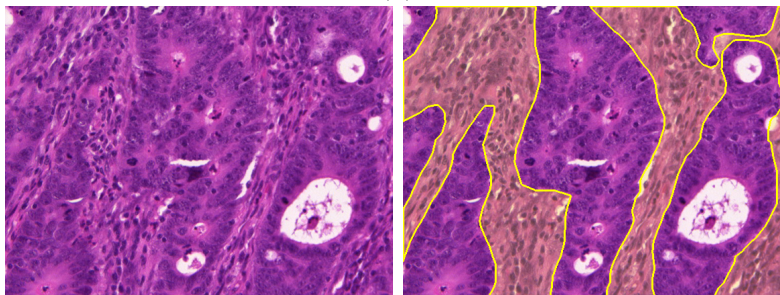
650 The objectives of this challenge were to raise the research community’s  
651 awareness of the existence of the intestinal gland segmentation problem in  
652 routine stained histology images, and at the same time to provide a plat-  
653 form for a standardized comparison of the performance of automatic and  
654 semi-automatic algorithms. The challenge attracted a lot of attention from  
655 researchers, as can be seen from the number of registered teams/individuals  
656 and the number of submissions at each stage of the competition. Interest-  
657 ingly, some of the teams had no experience in working with histology images  
658 before. We would like to emphasize that finding the best performing ap-  
659 proach is not the main objective of the competition, but rather pushing the



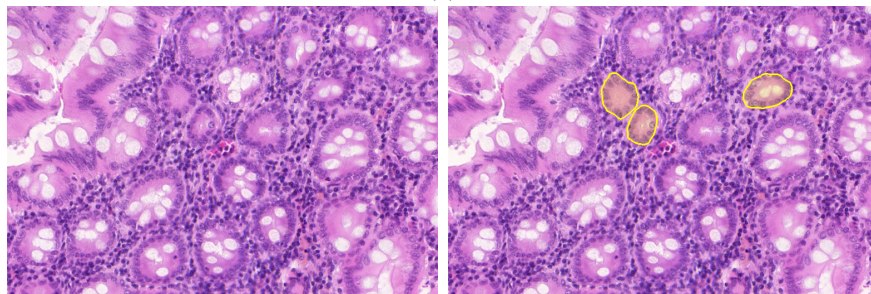
(a)



(b)



(c)



(d)

Figure 4: Example images showing some challenging features in the dataset: (a) lumen of the gastrointestinal tract, (b) sub-mucosa layer, (c) area with dense nuclei in mucosa layer, and (d) small glands. Each example is shown with (left) the original image and (right) the overlaid image highlighting the area with challenging characteristic.

660 boundaries of the-state-of-the-art approaches. Already, we have seen quite  
661 interesting developments from many participating teams and the leading al-  
662 gorithms have produced excellent results, both qualitatively and quantita-  
663 tively.

664 As noted in the Introduction, morphometric analysis of the appearance  
665 of cells and tissues, especially those forming glands from which tumors origi-  
666 nate, is one of the key components towards precision medicine, and segmen-  
667 tation is the first step to attain morphological information. Some may have  
668 argued that there is no need to perform segmentation, but instead, to fol-  
669 low conventional pattern recognition approaches by extracting mathematical  
670 features which normally capture local and/or global tissue architecture and  
671 then identifying features that are most suited to the objective of the study.  
672 It is true that there are a number of successful works that follow such an  
673 approach (Jafari-Khouzani and Soltanian-Zadeh, 2003; Tabesh et al., 2007;  
674 Altunbay et al., 2010; Basavanhally et al., 2010; Ozdemir and Gunduz-Demir,  
675 2013; Gultekin et al., 2015). However, because these extracted features are  
676 often physically less interpretable in the eyes of practitioners, it is difficult to  
677 adopt such an approach in clinical settings. On the other hand, the appear-  
678 ance of glands such as size and shape obtained through segmentation is easy  
679 to interpret. Segmentation also helps to localize other type of information  
680 (e.g., texture, spatial arrangement of cells) that is specific to the glandular  
681 areas.

682 Even though the dataset used in the challenge included images of different  
683 histologic grades taken from several patients, it lacked other aspects. First of  
684 all, inter-observer variability was not taken into account as the ground truth  
685 was generated by a single expert. This is because the intricate and arduous  
686 nature of the problem makes it difficult to find several volunteer experts to  
687 perform manual segmentation. Considerable experience is required in order  
688 to delineate boundaries of malignant glands, which are not so well-defined  
689 as those of the benign ones. Moreover, a single image can contain a large  
690 number of glands to be segmented, making the task very laborious. Sec-  
691 ondly, digitization variability was also not considered in this dataset. It is, in  
692 fact, very important to evaluate the robustness of algorithms when the data  
693 are scanned by different instruments. As whole-slide scanners are becoming  
694 increasingly available, this type of real-world problem should be expected.

695 The choice of evaluation measures would also affect the comparative re-  
696 sults. In this challenge, we emphasized object segmentation and accordingly  
697 defined the object-level Dice index and the object-level Hausdorff distance to

698 measure segmentation accuracy at the object level rather than at the pixel  
699 level. Nonetheless, it has been suggested that these measures are too strict,  
700 as they put a severe penalty on mismatch of the objects. One could replace  
701 these measures by less conservative ones, for example, adjusted Rand in-  
702 dex (Hubert and Arabie, 1985) or a topology preserving warping error (Jain  
703 et al., 2010) for a volume-based metric and elastic distance (Younes, 1998;  
704 Joshi et al., 2007) for a boundary-based metric. For this reason, we included  
705 adjusted rand index as an alternative to object-level Dice index in Section  
706 8.3. As we have already pointed out, this results in only a minor change in the  
707 ranking order of the entries. Another aspect that was not explicitly included  
708 in the evaluation was execution times. Nevertheless, all the algorithms were  
709 capable of completing the segmentation task on the on-site test data (Part  
710 B) in the given amount of time with or without limitation of resources. Time  
711 efficiency is required to process large scale data, such as whole-slide images,  
712 whose volume is growing by the day as slides are routinely scanned. Still, in  
713 medical practice, accuracy is far more important than speed.

714 It is worth noting that the used evaluation metrics used here are clini-  
715 cally relevant. As mentioned in the Introduction, morphology of intestinal  
716 glands is the key criterion for colorectal cancer grading. This includes shape,  
717 size, and formation of the glands. Thus, in terms of clinical relevance, the  
718 object-Hausdorff distance is used in accessing the shape similarity between  
719 the segmentation results and the ground truth. The object-Dice index is used  
720 in assessing the closeness between the volume of the segmentation results and  
721 that of the ground truth, which is important in estimating the size of individ-  
722 ual glands. Although not directly clinically relevant, F1 score is important  
723 in assessing the accuracy of the identified glands. Since the morphological  
724 assessment is done on the basis of tissue slide including several thousands of  
725 glands, an algorithm with high value of F1 score is more preferable as it can  
726 detect a larger number of glands.

727 Gland segmentation algorithms presented here are not ready for deploy-  
728 ment into clinic in their present form. Although some of the top algorithms  
729 produce good segmentation results for the contest dataset and will probably  
730 fare well in the real world, there needs to be a large-scale validation involving  
731 data from multiple centers annotated by multiple pathologists before any of  
732 these algorithms can be deployed in a diagnostic application.

733 The challenge is now completed, but the dataset will remain available  
734 for research purposes so as to continually attract newcomers to the problem  
735 and to encourage development of state-of-the-art methods. Extension of the

736 dataset to address inter-observer and inter-scanner variability seems to be the  
737 most achievable aim in the near future. Beyond the scope of segmentation,  
738 there lie various extremely interesting future research directions. Previous  
739 studies have shown the strong association between the survival of colorectal  
740 cancer patients and tumor-related characteristics, including lymphocytic in-  
741 filtration (Galon et al., 2006; Fridman et al., 2012), desmoplasia (Tommelein  
742 et al., 2015), tumor budding (Mitrovic et al., 2012), and necrosis (Richards  
743 et al., 2012). A systematic analysis of these characteristics with the help  
744 of gland segmentation as part of automatic image analysis framework could  
745 lead to a better understanding of the relevant cancer biology as well as bring  
746 precision and accuracy into assessment and prediction of the outcome of the  
747 cancer.

## 748 **9. Conclusions**

749 This paper presented a summary of the Gland Segmentation in Colon  
750 Histology Images (GlaS) Challenge Contest which was held in conjunction  
751 with the 18th International Conference on Medical Image Computing and  
752 Computer Assisted Interventions (MICCAI'2015). The goal of the challenge  
753 was to bring together researchers interested in the gland segmentation prob-  
754 lem, to validate the performance of their existing or newly invented algo-  
755 rithms on the same standard dataset. In the final round, the total num-  
756 ber of submitted entries for evaluation was 19, and we presented here in  
757 this paper 10 of the leading entries. The dataset used in the challenge has  
758 been made publicly available and can be accessed at the challenge website  
759 (<http://www.warwick.ac.uk/bialab/GlasContest/>). Those who are in-  
760 terested in developing or improving their own approaches are encouraged to  
761 use this dataset for quantitative evaluation.

## 762 **10. Acknowledgements**

763 This paper was made possible by NPRP grant number NPRP5-1345-1-228  
764 from the Qatar National Research Fund (a member of Qatar Foundation).  
765 The statements made herein are solely the responsibility of the authors. Kor-  
766 suk Sirinukunwattana acknowledges the partial financial support provided by  
767 the Department of Computer Science, University of Warwick, UK. CUMed-  
768 Vision team acknowledges Hong Kong RGC General Research Fund (Project  
769 No. CUHK 412513). The work of Olaf Ronneberger was supported by the

770 Excellence Initiative of the German Federal and State Governments (EXC  
771 294). Philipp Kainz was supported by the Excellence Grant 2014 of the  
772 Federation of Austrian Industries (IV), and Martin Urschler acknowledges  
773 funding by the Austrian Science Fund (FWF): P 28078-N33.

774 The authors thank Professor David Epstein for his extensive help with  
775 the wording and mathematical notation in this paper, and Nicholas Trahearn  
776 for his help with the wording.

## 777 **Appendix A. The Complete Contest Results**

778 A summary of the ranking results from the contest is given in Figure A.5.  
779

## 780 **References**

781 Altunbay, D., Cigir, C., Sokmensuer, C., Gunduz-Demir, C., 2010. Color  
782 graphs for automated cancer diagnosis and grading. *IEEE Transactions*  
783 *on Biomedical Engineering* 57, 665–674.

784 Barker, N., 2014. Adult intestinal stem cells: critical drivers of epithelial  
785 homeostasis and regeneration. *Nature Reviews Molecular Cell Biology* 15,  
786 19–33.

787 Basavanhally, A.N., Ganesan, S., Agner, S., Monaco, J.P., Feldman, M.D.,  
788 Tomaszewski, J.E., Bhanot, G., Madabhushi, A., 2010. Computerized  
789 image-based detection and grading of lymphocytic infiltration in HER2+  
790 breast cancer histopathology. *IEEE Transactions on Biomedical Engineer-*  
791 *ing* 57, 642–653.

792 Ben Cheikh, B., Bertheau, P., Racoceanu, D., 2016. A structure-based ap-  
793 proach for colon gland segmentation in digital pathology, in: *SPIE Medical*  
794 *Imaging*, International Society for Optics and Photonics.

795 Bosman, F.T., Carneiro, F., Hruban, R.H., Theise, N.D., et al., 2010. WHO  
796 classification of tumours of the digestive system. Ed. 4, World Health  
797 Organization.

798 Bresson, X., Esedolu, S., Vandergheynst, P., Thiran, J.P., Osher, S., 2007.  
799 Fast global minimization of the active contour/snake model. *Journal of*  
800 *Mathematical Imaging and Vision* 28, 151–167.

Method	Rank						Sum
	F1 Score		Object Dice		Object Hausdorff		
	Part A	Part B	Part A	Part B	Part A	Part B	
CUMedVision 2	1	3	1	5	1	6	17
ExB 1	4	4	4	2	6	1	21
ExB 3	2	2	2	6	5	5	22
Freiburg 2 <sup>a</sup>	5	5	5	3	3	3	24
CUMedVision 1	6	1	8	1	8	4	28
ExB 2	3	6	3	7	2	8	29
Freiburg 1 <sup>a</sup>	8	8	6	4	4	2	32
CVIP Dundee <sup>b</sup>	7	7	7	8	7	10	46
CVML	10	9	11	9	11	7	57
LIB	9	17	9	12	9	9	65
vision4GlaS	11	10	10	14	10	11	66
LIST	13	11	14	11	14	14	77
Ching-Wei Wang 1 <sup>c</sup>	12	12	15	13	16	16	84
Bioimage Informatics	16	15	17	10	18	12	88
Ching-Wei Wang 2 <sup>c</sup>	14	13	16	15	17	17	92
SUTECH	15	18	13	18	13	15	92
ISI Kolkatta	18	16	18	19	12	13	96
FIMM	19	19	12	17	15	19	101
Ching-Wei Wang 3 <sup>c</sup>	17	14	19	16	19	18	103

<sup>a</sup> Image Analysis Lab Uni Freiburg: Freiburg 2 = post-processing, Freiburg 1 = raw

<sup>b</sup> CVIP Dundee: feature level fusion

<sup>c</sup> Ching-Wei Wang: Ching-Wei Wang 1 = no preprocess fill hole, Ching-Wei Wang 2 = no preprocess hole, Ching-Wei Wang 3 = preprocess fill hole

Figure A.5: The ranking results from the GlaS Challenge Contest.

- 801 Chen, H., Dou, Q., Ni, D., Cheng, J.Z., Qin, J., Li, S., Heng, P.A., 2015.  
802 Automatic fetal ultrasound standard plane detection using knowledge  
803 transferred recurrent neural networks, in: Medical Image Computing and  
804 Computer-Assisted Intervention–MICCAI 2015. Springer, pp. 507–514.
- 805 Chen, H., Qi, X., Yu, L., Heng, P.A., 2016. DCAN: Deep contour-aware net-  
806 works for accurate gland segmentation. arXiv preprint arXiv:1604.02677  
807 .
- 808 Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2014.  
809 Semantic image segmentation with deep convolutional nets and fully con-  
810 nected CRFs. arXiv preprint arXiv:1412.7062 .
- 811 Cohen, A., Rivlin, E., Shimshoni, I., Sabo, E., 2015. Memory based ac-  
812 tive contour algorithm using pixel-level classified images for colon crypt  
813 segmentation. *Computerized Medical Imaging and Graphics* 43, 150–164.
- 814 Compton, C.C., 2000. Updated protocol for the examination of specimens  
815 from patients with carcinomas of the colon and rectum, excluding carcinoid  
816 tumors, lymphomas, sarcomas, and tumors of the vermiform appendix: a  
817 basis for checklists. *Archives of Pathology & Laboratory Medicine* 124,  
818 1016–1025.
- 819 Crosnier, C., Stamataki, D., Lewis, J., 2006. Organizing cell renewal in the  
820 intestine: stem cells, signals and combinatorial control. *Nature Reviews*  
821 *Genetics* 7, 349–359.
- 822 Cross, S.S., Betmouni, S., Burton, J.L., Dubé, A.K., Feeley, K.M., Holbrook,  
823 M.R., Landers, R.J., Lumb, P.B., Stephenson, T.J., 2000. What levels of  
824 agreement can be expected between histopathologists assigning cases to  
825 discrete nominal categories? a study of the diagnosis of hyperplastic and  
826 adenomatous colorectal polyps. *Modern Pathology* 13, 941–944.
- 827 Dice, L.R., 1945. Measures of the amount of ecologic association between  
828 species. *Ecology* 26, 297–302.
- 829 Fanshawe, T.R., Lynch, A.G., Ellis, I.O., Green, A.R., Hanka, R., 2008. As-  
830 ssuming agreement between multiple raters with missing rating information,  
831 applied to breast cancer tumour grading. *PLoS One* 3, e2925.

- 832 Fridman, W.H., Pagès, F., Sautès-Fridman, C., Galon, J., 2012. The immune  
833 contexture in human tumours: impact on clinical outcome. *Nature Reviews*  
834 *Cancer* 12, 298–306.
- 835 Fu, H., Qiu, G., Shu, J., Ilyas, M., 2014. A novel polar space random field  
836 model for the detection of glandular structures. *IEEE Transactions on*  
837 *Medical Imaging* 33, 764–776. doi:10.1109/TMI.2013.2296572.
- 838 Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-  
839 Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al., 2006.  
840 Type, density, and location of immune cells within human colorectal tu-  
841 mors predict clinical outcome. *Science* 313, 1960–1964.
- 842 Gibson, P.R., Anderson, R.P., Mariadason, J.M., Wilson, A.J., 1996. Protec-  
843 tive role of the epithelium of the small intestine and colon. *Inflammatory*  
844 *Bowel Diseases* 2, 279–302.
- 845 Gultekin, T., Koyuncu, C.F., Sokmensuer, C., Gunduz-Demir, C., 2015.  
846 Two-tier tissue decomposition for histopathological image representation  
847 and classification. *IEEE Transactions on Medical Imaging* 34, 275–283.
- 848 Gunduz-Demir, C., Kandemir, M., Tosun, A.B., Sokmensuer, C., 2010. Au-  
849 tomatic segmentation of colon glands using object-graphs. *Medical Image*  
850 *Analysis* 14, 1–12.
- 851 Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M.,  
852 Yener, B., 2009. Histopathological image analysis: A review. *IEEE Re-*  
853 *views in Biomedical Engineering* 2, 147–171.
- 854 Gurcan, M.N., Madabhushi, A., Rajpoot, N., 2010. Pattern recognition in  
855 histopathological images: An ICPR 2010 contest, in: *Recognizing Patterns*  
856 *in Signals, Speech, Images and Videos*. Springer, pp. 226–234.
- 857 Hubert, L., Arabie, P., 1985. Comparing partitions. *Journal of Classification*  
858 2, 193–218.
- 859 Humphries, A., Wright, N.A., 2008. Colonic crypt organization and tumori-  
860 genesis. *Nature Reviews Cancer* 8, 415–424.
- 861 Jafari-Khouzani, K., Soltanian-Zadeh, H., 2003. Multiwavelet grading of  
862 pathological images of prostate. *IEEE Transactions on Biomedical Engi-*  
863 *neering* 50, 697–704.

- 864 Jain, V., Bollmann, B., Richardson, M., Berger, D.R., Helmstaedter, M.N.,  
865 Briggman, K.L., Denk, W., Bowden, J.B., Mendenhall, J.M., Abraham,  
866 W.C., et al., 2010. Boundary learning by optimization with topological  
867 constraints, in: 2010 IEEE Conference on Computer Vision and Pattern  
868 Recognition (CVPR), IEEE. pp. 2488–2495.
- 869 Joshi, S.H., Klassen, E., Srivastava, A., Jermyn, I., 2007. Removing shape-  
870 preserving transformations in square-root elastic (SRE) framework for  
871 shape analysis of curves, in: Energy Minimization Methods in Computer  
872 Vision and Pattern Recognition, Springer. pp. 387–398.
- 873 Kainz, P., Pfeiffer, M., Urschler, M., 2015. Semantic segmentation of colon  
874 glands with deep convolutional neural networks and total variation seg-  
875 mentation. arXiv preprint arXiv:1511.06919 .
- 876 Komuta, K., Batts, K., Jessurun, J., Snover, D., Garcia-Aguilar, J., Rothen-  
877 berger, D., Madoff, R., 2004. Interobserver variability in the pathological  
878 assessment of malignant colorectal polyps. *British Journal of Surgery* 91,  
879 1479–1484.
- 880 Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification  
881 with deep convolutional neural networks, in: *Advances in Neural Informa-*  
882 *tion Processing Systems*, pp. 1097–1105.
- 883 LeCun, Y., Kavukcuoglu, K., Farabet, C., 2010. Convolutional networks  
884 and applications in vision, in: *Proceedings of 2010 IEEE International*  
885 *Symposium on Circuits and Systems (ISCAS)*, IEEE. pp. 253–256.
- 886 Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks  
887 for semantic segmentation, in: *Proceedings of the IEEE Conference on*  
888 *Computer Vision and Pattern Recognition*, pp. 3431–3440.
- 889 Matas, J., Chum, O., Urban, M., Pajdla, T., 2004. Robust wide-baseline  
890 stereo from maximally stable extremal regions. *Image and Vision Com-*  
891 *puting* 22, 761–767.
- 892 May, M., 2010. A better lens on disease. *Scientific American* 302, 74–77.
- 893 Mitrovic, B., Schaeffer, D.F., Riddell, R.H., Kirsch, R., 2012. Tumor bud-  
894 ding in colorectal carcinoma: time to take notice. *Modern Pathology* 25,  
895 1315–1325.

- 896 Murphy, K., Van Ginneken, B., Reinhardt, J.M., Kabus, S., Ding, K., Deng,  
897 X., Cao, K., Du, K., Christensen, G.E., Garcia, V., et al., 2011. Evaluation  
898 of registration methods on thoracic CT: the EMPIRE10 challenge. *IEEE*  
899 *Transactions on Medical Imaging* 30, 1901–1920.
- 900 Ozdemir, E., Gunduz-Demir, C., 2013. A hybrid classification model for  
901 digital pathology using structural and statistical pattern recognition. *IEEE*  
902 *Transactions on Medical Imaging* 32, 474–483.
- 903 Richards, C., Roxburgh, C., Anderson, J., McKee, R., Foulis, A., Horgan,  
904 P., McMillan, D., 2012. Prognostic value of tumour necrosis and host  
905 inflammatory responses in colorectal cancer. *British Journal of Surgery*  
906 99, 287–294.
- 907 Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks  
908 for biomedical image segmentation, in: Navab, N., Hornegger, J., Wells,  
909 W., Frangi, A. (Eds.), *Medical Image Computing and Computer-Assisted*  
910 *Intervention MICCAI 2015*. Springer International Publishing. volume  
911 9351 of *Lecture Notes in Computer Science*, pp. 234–241.
- 912 Roux, L., Racoceanu, D., Lomnie, N., Kulikova, M., Irshad, H., Klossa, J.,  
913 Capron, F., Genestie, C., Naour, G., Gurcan, M., 2013. Mitosis detection  
914 in breast cancer histological images An ICPR 2012 contest. *Journal of*  
915 *Pathology Informatics* 4, 8. doi:10.4103/2153-3539.112693.
- 916 Rubin, R., Strayer, D.S., Rubin, E., et al., 2008. *Rubin’s pathology: clinico-*  
917 *pathologic foundations of medicine*. Lippincott Williams & Wilkins.
- 918 Shanmugathasan, M., Jothy, S., 2000. Apoptosis, anoikis and their relevance  
919 to the pathobiology of colon cancer. *Pathology International* 50, 273–279.
- 920 Sirinukunwattana, K., Snead, D., Rajpoot, N., 2015. A stochastic polygons  
921 model for glandular structures in colon histology images. *IEEE Transac-*  
922 *tions on Medical Imaging* 34, 2366–2378. doi:10.1109/TMI.2015.2433900.
- 923 Tabesh, A., Teverovskiy, M., Pang, H.Y., Kumar, V.P., Verbel, D., Kotsianti,  
924 A., Saidi, O., 2007. Multifeature prostate cancer diagnosis and Gleason  
925 grading of histological images. *IEEE Transactions on Medical Imaging* 26,  
926 1366–1378.

- 927 Tommelein, J., Verset, L., Boterberg, T., Demetter, P., Bracke, M.,  
928 De Wever, O., 2015. Cancer-associated fibroblasts connect metastasis-  
929 promoting communication in colorectal cancer. *Frontiers in Oncology* 5.
- 930 Veta, M., Van Diest, P.J., Willems, S.M., Wang, H., Madabhushi, A., Cruz-  
931 Roa, A., Gonzalez, F., Larsen, A.B., Vestergaard, J.S., Dahl, A.B., et al.,  
932 2015. Assessment of algorithms for mitosis detection in breast cancer  
933 histopathology images. *Medical Image Analysis* 20, 237–248.
- 934 Washington, M.K., Berlin, J., Branton, P., Burgart, L.J., Carter, D.K.,  
935 Fitzgibbons, P.L., Halling, K., Frankel, W., Jessup, J., Kakar, S., et al.,  
936 2009. Protocol for the examination of specimens from patients with pri-  
937 mary carcinoma of the colon and rectum. *Archives of Pathology & Labo-  
938 ratory Medicine* 133, 1539.
- 939 Wu, H.S., Xu, R., Harpaz, N., Burstein, D., Gil, J., 2005a. Segmentation  
940 of intestinal gland images with iterative region growing. *Journal of Mi-  
941 croscopy* 220, 190–204.
- 942 Wu, H.S., Xu, R., Harpaz, N., Burstein, D., Gil, J., 2005b. Segmenta-  
943 tion of microscopic images of small intestinal glands with directional 2-d  
944 filters. *Analytical and Quantitative Cytology and Histology/the Inter-  
945 national Academy of Cytology and American Society of Cytology* 27,  
946 291–300.
- 947 Younes, L., 1998. Computable elastic distances between shapes. *SIAM Jour-  
948 nal on Applied Mathematics* 58, 565–586.
- 949 Zhang, Y., Matuszewski, B.J., Shark, L.K., Moore, C.J., 2008. Medical  
950 image segmentation using new hybrid level-set method, in: *BioMedical  
951 Visualization, 2008. MEDIVIS'08. Fifth International Conference, IEEE.*  
952 pp. 71–76.