# Detection of Glands and Villi by Collaboration of Domain Knowledge and Deep Learning*

Jiazhuo Wang[1], John D. MacKenzie[2],
Rageshree Ramachandran[3], and Danny Z. Chen[1]

[1] Department of Computer Science & Engineering, University of Notre Dame, USA
[2] Department of Radiology & Biomedical Imaging, UCSF, USA
[3] Department of Pathology, UCSF, USA

**Abstract.** Architecture distortions of glands and villi are indication of chronic inflammation. However, the "duality" nature of these two structures causes lots of ambiguity for their detection in H&E histology tissue images, especially when multiple instances are clustered together. Based on the observation that once such an object is detected for certain, the ambiguity in the neighborhood of the detected object can be reduced considerably, we propose to combine deep learning and domain knowledge in a unified framework, to simultaneously detect (the closely related) glands and villi in H&E histology tissue images. Our method iterates between exploring domain knowledge and performing deep learning classification, and the two components benefit from each other. (1) By exploring domain knowledge, the generated object proposals (to be fed to deep learning) form a more complete coverage of the true objects and the segmentation of object proposals can be more accurate, thus improving deep learning's performance on classification. (2) Deep learning can help verify the class of each object proposal, and provide feedback to repeatedly "refresh" and enhance domain knowledge so that more reliable object proposals can be generated later on. Experiments on clinical data validate our ideas and show that our method improves the state-of-the-art for gland detection in H&E histology tissue images (to our best knowledge, we are not aware of any method for villi detection).

## 1 Introduction

Architecture distortions of glands and villi are strong signs of chronic inflammation [9]. Also, a quantitative measurement of the degree of such distortions may help determine the severity of the chronic inflammation. A crucial step towards these goals is the ability to detect accurately these two biological structures.

As shown in Fig. 1(a)-(b), both glands and villi are actually composed of the same structure: epithelium. A gland encloses lumen and is surrounded by extracellular material, while a villus encloses extracellular material but is surrounded

* This research was supported in part by NSF Grant CCF-1217906, a grant of the National Academies Keck Futures Initiative (NAKFI), and NIH grant K08-AR061412-02 Molecular Imaging for Detection and Treatment Monitoring of Arthritis.

N. Navab et al. (Eds.): MICCAI 2015, Part II, LNCS 9350, pp. 20–27, 2015.
DOI: 10.1007/978-3-319-24571-3_3

by lumen. In H&E histology tissue images, the detection challenges of glands and villi are mainly due to such "duality" of the two structures (especially when multiple instances are clustered together), as well as the complex tissue background (containing different biological structures, e.g., different types of cells, connective tissue, etc), and the variable appearances of glands and villi due to morphology, staining, and scale.

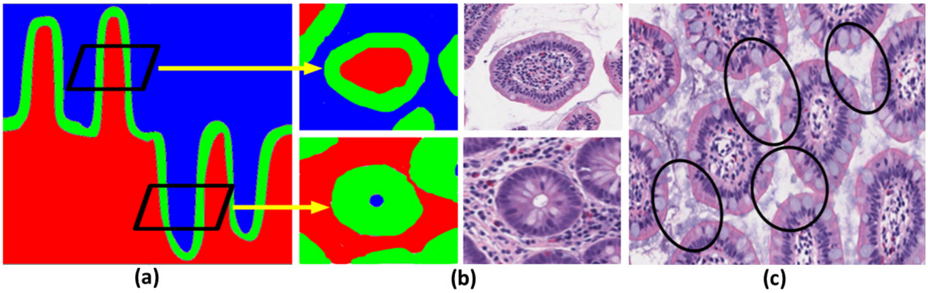

**Fig. 1.** (a) A 3-D illustration of the dual glands and villi: Villi (top) are evagination of epithelium (green) into lumen (blue), and glands (bottom) are invagination of epithelium into extracellular material (red); (b) histology tissue images are 2-D slices of the 3-D structures; (c) some areas (black circles) that may cause false positives of glands.

Some methods [4,8,10] were proposed for glands detection in H&E histology tissue images, which used a similar framework: (1) Find lumen regions; (2) for each lumen region, perform a region-growing like process to find the epithelium enclosing the lumen which is considered as the boundary of a gland. Applying such a method in the presence of villi clusters could generate many false positives for glands, because a lumen region among different villi may be found in the first step, and then the epithelium regions bounding these villi may be taken incorrectly as the boundaries of a gland enclosing the lumen (see Fig. 1(c)). Also, due to certain slicing angles for the images, the lumen regions inside some glands may not be very obvious; thus, this methodology may tend to produce false negatives for such glands. A recent glands detection method [2] was proposed for H-DAB images, and it also did not consider the influence of the dual villi. To our best knowledge, we are not aware of any previous work on detecting villi.

In this paper, we propose to combine domain knowledge and deep learning to simultaneously detect glands and villi (since they are closely related) in H&E histology tissue images. Our method is based on the observation that once we detect an object (of some class, i.e., glands or villi) for certain, we can propagate this information to the neighborhood of the detected object, so that the detection ambiguity nearby is reduced. The main steps of our method are as follows.

(1) We extract (pseudo-)probability maps (PPMs) for possible candidates of the target objects, by using domain knowledge on the appearances of glands and villi. (2) Using PPMs, we generate object proposals and feed them to deep

convolutional neural networks (CNN) [6], to verify whether each object is really of the class claimed by PPMs (reflecting domain knowledge). (3) If the object proposals pass the verification, then we update PPMs (essentially, propagating the information that we have detected some objects for certain), so that new object proposals can be generated using the updated domain knowledge. We repeat the last two steps until no more object can be detected for certain.

Our work shows that the close collaboration between domain knowledge and deep learning allows multiple instances of glands and villi to be detected effectively. Experimental results (summarized in Table 1) on clinical data validate our ideas and show that we improve the state-of-the-art for glands detection.

## 2 Methodology

### 2.1 Extraction of (Pseudo-)Probability Maps

Each histology tissue slide may contain multiple instances of glands and villi, possibly in different scales. Thus, the first step of our method aims to initially extract (pseudo-)probability maps (PPMs) that contain information of both the locations and scales for all objects of the two target classes (glands and villi). We will generate object proposals based on the PPMs in the next step. Our main idea for this step is to conduct a generalized Hough transform voting process.

This idea is based on two considerations after exploring domain knowledge of the appearances of glands and villi. (I) Each epithelium region suggests that a target object is nearby, but its class (i.e., gland or villus), location, and scale are not yet clear, at least from the perspective of this single epithelium region. (II) A more clear and complete picture of all objects could be obtained after each epithelium region votes (based on its own view). This is because, collectively, true positives of objects are more likely to receive more votes from such epithelium regions. Our idea and steps are discussed in more detail below.

(1) We first obtain a superpixel segmentation [1] (Fig. 2(b)) of the image. We then classify each superpixel as epithelium, lumen, or extracellular material (since they are all related to the appearances of glands and villi), using Random Forest and hand-crafted features (e.g., colors, texture based on Gabor filters).

(2) Since each epithelium superpixel suggests that a target object is nearby, it would vote for some points in a 4-D object voting space (voting process is illustrated below), where each dimension corresponds to a factor of each single object, i.e., its class (glands or villi), $x$ and $y$ coordinates in the image, and scale (we empirically use 8 scales corresponding to roughly $S=\{0.011, 0.014, 0.025, 0.05, 0.067, 0.111, 0.167, 0.25\}$ times $LS$, the length of the shorter side of the image).

(2.a) We first find the lumen (LM) and extracellular material (EM) within a distance of $d = S(i) \times LS$ from this epithelium superpixel (ES), if any. (2.b) For each class and each scale, we map the ES to some locations, based on the following observation for choosing the mapping/voting direction (which narrows down the voting space to be covered): If the ES is part of a gland, then the gland is likely to lie on the same side as LM, but on the opposite side from EM (found in (2.a) near this ES); if it is actually part of a villus, then due to the "duality"

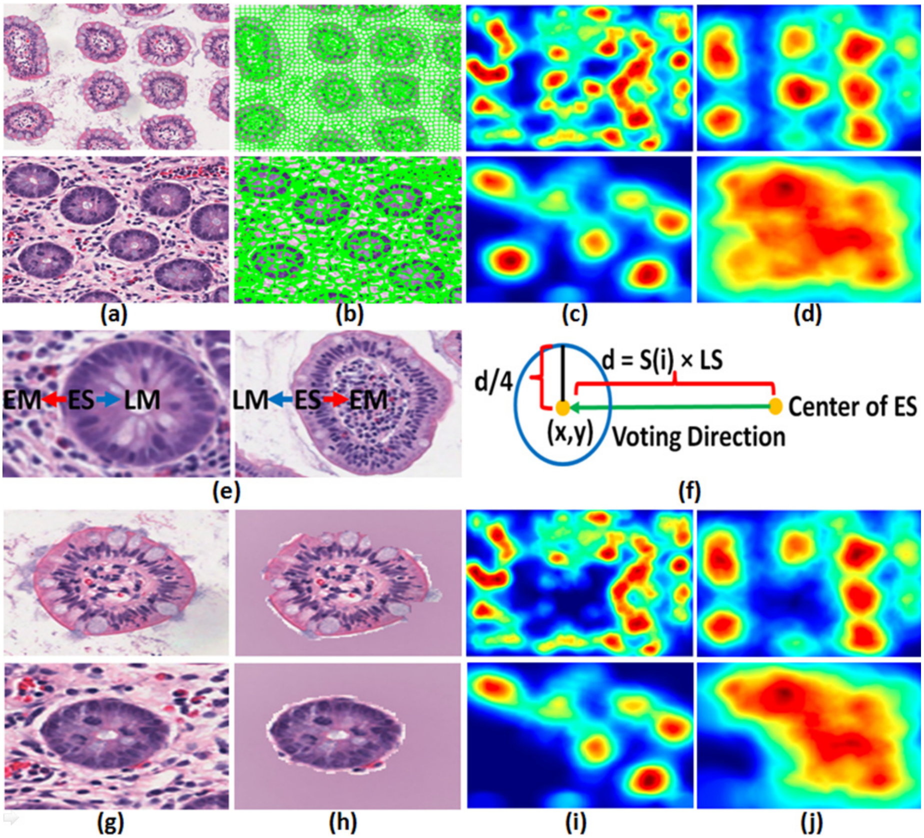

**Fig. 2.** (a) Image samples; (b) superpixel segments; initial PPMs (high values in red; low values in blue) for respectively glands (c) and villi (d) at a single scale; (e) blue and red arrows are voting directions for glands and villi, respectively; (f) the voted points in a circle (blue) by an ES at a single scale for one class; (g) image patches containing detected objects (the presented villus (top) is around the center of the top image in (a), and the presented gland (bottom) is at the bottom left of the bottom image in (a)); (h) image patches with background masked out; updated PPMs for respectively glands (i) and villi (j) at a single scale, after the information of the objects in (g) detected is propagated to the neighborhood (note the changes of the (pseudo-)probability values around the detected objects).

of glands and villi, the villus is likely to lie on the same side as EM, but on the opposite side from LM (see Fig. 2(e)). More specifically, the ES would vote for the points in a circle (of a radius $d/4$) centered at the $(x, y)$ coordinates, which are of a distance $d$ away from the center of the ES, towards the chosen directions accordingly, for respectively glands and villi, and the 8 scales (see Fig. 2(f)).

(3) We view the corresponding 3-D hyper-plane w.r.t. to the class dimension in the built 4-D voting space as the initial PPMs of each class, containing the location and scale information for the candidate objects (see Fig. 2(c)-(d)).

### 2.2 Object Proposals Generation and Class Verification

**Overview.** This step aims to make sure that detected objects are indeed true positive objects, so that in the next step, we would not propagate wrong information to their neighborhoods for resolving ambiguity. We first apply graph search [5] (which uses high level priors) to conduct segmentation for object proposals (generated based on PPMs), and then feed two small image patches containing the same object proposal (with or without the background masked out) to respectively **two** convolutional neural networks (CNN) with the same architecture, to verify whether the class of that object proposal is the one claimed by PPMs.

The observation that object proposals generated using domain knowledge can help CNN is critical. (1) If being blind to domain knowledge (e.g., in [3,11,6]), many true objects might be missed by the generated object proposals. Thus, during training, CNN cannot well model the target objects; during testing, false negatives can be produced. (2) Although the segmented form of object proposals at the pixel level can help improve CNN's performance on object classification [6], segmentation in [6] is done in a bottom-up fashion, using only low level image cues. Based on which class of the PPMs (reflecting domain knowledge) that an object proposal is generated from, we can obtain more accurate segmentation in a top-down fashion, by utilizing class-specific high level semantic priors.

**CNN Training.** For every class, we find, in the corresponding **initial** PPM, each local maximum point above a certain threshold as one object proposal, and perform graph search based segmentation for it. If the segmented foreground region $R_{Seg}$ and a manually marked ground truth object region $R_{GT}$ of that class satisfy $\frac{|R_{Seg} \cap R_{GT}|}{|R_{Seg}|} > 0.6$ and $\frac{|R_{Seg} \cap R_{GT}|}{|R_{GT}|} > 0.6$, then we take the object proposal as a positive training example of that class; otherwise, a negative one. Note the relatively high overlap threshold 0.6 is to make trained CNN be conservative so that false positives are less likely to pass CNN verification during testing.

We crop a small image patch containing the object proposal, warp it to $256 \times 256$ pixels, and use in training one CNN. We further mask out the background region in it by the mean values of all training image patches (see Fig. 2(g)-(h)), and use it in training the other CNN. Our two CNNs have the same architecture and are trained using the same learning algorithm as [7]; we also apply data augmentation and drop out to reduce overfitting as [7].

Note CNN once trained using **initial** PPMs, will be used in all rest iterations.

**CNN Testing.** During CNN testing, we also perform graph search based segmentation for each object proposal (generated using the **current** PPMs, as described in detail below), and feed the two image patches (with or without background masked out) to respectively the two CNNs. We use the average probabilities output by the two CNNs to predict the class. Once an object proposal is verified by CNNs as a true positive for either glands or villi (i.e., the prediction result is not a non-object), we will propagate this information to the neighborhood regions by updating the PPMs (to be described in Section 2.3). We stop the algorithm if CNN cannot verify any true object.

Since during the testing, the PPMs change dynamically as more and more object proposals pass the verification by CNNs, and object proposals are generated based on the new versions of PPMs after update, we need to determine what would be an appropriate order to generate and process object proposals. One possible order is of a greedy and one-by-one fashion: Each time, we find a point with the **largest** (pseudo-)probability value in current PPMs, verify the corresponding object proposal by CNNs, and update PPMs if necessary.

Another possible way is of a batch fashion: Each time, we generate a batch of object proposals, verify all of them by CNNs, and update PPMs if necessary. Of course, object proposals generated within the same batch should not be closely related to one another (otherwise, we may have conflicting information to propagate subsequently). We formulate the problem of finding a non-conflicting batch as computing a *maximal weighted independent set* (MWIS) in a vertex weighted graph $G$, constructed as follows: Each vertex of $G$ is for a local maximal point above a certain threshold in the PPM of either class; connect two vertices by an edge if the $(x, y)$ coordinates of one vertex in the image are within 2 times the scale of those of the other vertex; the weight of each vertex is the (pseudo-)probability value of the corresponding point. (Hence, we may view MWIS as being partially greedy.)

**Graph Search.** We apply graph search [5] to segment each object proposal. We utilize various category-specific high level semantic priors to construct the needed input for graph search, as follows. We use the scale information of each object proposal to set the lengths of the re-sampling rays and the geometric smoothness constraint. We simply apply distance transform to the border pixels between detected epithelium and extracellular material (resp., lumen) to set up on-boundary cost of glands (resp., villi). We simply set in-region cost for glands (resp., villi) to be low for pixels inside epithelium or lumen (resp., extracellular material), and high for pixels inside extracellular material (resp., lumen). Note better segmentation may be obtained with more sophisticated cost functions.

### 2.3 Information Propagation

This step aims to propagate the information that an object, $Obj$, is detected, so that the detection ambiguity in its neighborhood is reduced. This is why we generate object proposals dynamically (to take advantage of the reduced ambiguity) instead of all in one-shot as in [3,11,6]. Our idea is to update PPMs (see Fig. 2(i)-(j)) which are used to generate new object proposals. Specifically, for the segmented foreground region of $Obj$, $R_{Obj}$, we find each epithelium superpixel (ES) such that its region $R_{ES}$ satisfies $\frac{|R_{ES} \cap R_{Obj}|}{|R_{ES}|} > 0.8$ and $\frac{|R_{ES} \cap R_{Obj}|}{|R_{Obj}|} > 0.8$. Since these ESs are quite unlikely to be part of other target objects, we remove their votes for all points that they have voted for in the 4-D voting space, thus resulting in new PPMs, with the information of $Obj$'s detection being incorporated and propagated.

## 3 Experiments and Discussions

We collected clinical H&E histology tissue slides (scanned at 40X magnification) from patients suspected of having inflammatory bowl disease (the study was performed under an IRB and in compliance with the privacy provisions of HIPPA of 1996). We manually marked 1376 glands and 313 villi and their boundaries as ground truth. We used 2-fold cross validation (CNN training for each fold takes a week on a single CPU) to evaluate the performance using three metrics, $precision = \frac{TP}{TP+FP}$, $recall = \frac{TP}{TP+FN}$, and $Fscore = \frac{2\times precision \times recall}{precision+recall}$. A detected object is counted as TP if its segmented foreground region $R_{Seg}$ and a ground truth region $R_{GT}$ (not corresponding to any other detected object) satisfy $\frac{|R_{Seg} \cap R_{GT}|}{|R_{Seg}|} > 0.6$ and $\frac{|R_{Seg} \cap R_{GT}|}{|R_{GT}|} > 0.6$; otherwise, it is counted as FP. If a ground truth region corresponds to no detected object, it is counted as FN.

**Table 1.** Quantitative performance of different methods.

| | Precision | | Recall | | F score | |
|---|---|---|---|---|---|---|
| | Glands | Villi | Glands | Villi | Glands | Villi |
| DK-Only | 0.79 | 0.73 | 0.75 | 0.86 | 0.77 | 0.79 |
| CNN-Only | 0.82 | 0.71 | 0.69 | 0.81 | 0.75 | 0.76 |
| DK-CNN-Greedy | 0.95 | 0.94 | 0.80 | 0.85 | 0.87 | 0.89 |
| DK-CNN-MWIS | 0.96 | 0.95 | 0.78 | 0.79 | 0.86 | 0.86 |
| Glands-Detect [10] | 0.42 | - | 0.54 | - | 0.47 | - |

To validate the key ideas we proposed, we construct several versions of our method (DK-Only, CNN-Only, DK-CNN-Greedy, and DK-CNN-MWIS), by either simply taking away or replacing by alternatives some important components of our method, as follows. DK-Only depends only on exploring domain knowledge (i.e., we take away the CNN verification) to detect objects. In CNN-Only, all object proposals and their segmentation are generated in one-shot, in a bottom-up fashion, and being blind to domain knowledge as in [6]; instead of dynamically and in a top-down fashion by exploring domain knowledge. We use the greedy fashion to address the object proposals during testing in these two versions. DK-CNN-Greedy uses both domain knowledge and CNN, and is based on the greedy fashion. DK-CNN-MWIS also uses both domain knowledge and CNN, but is based on the maximal weighted independent set approach.

Table 1 shows the quantitative performance of these four versions and the state-of-the-art glands detection method [10] (referred to as Glands-Detect). One can see the following points. (1) The two complete versions DK-CNN-Greedy and DK-CNN-MWIS perform much better than DK-Only, indicating that CNN is suitable for the verification task on object detection, which reduces the amount of incorrect information propagated to the neighborhoods to resolve ambiguity. (2) These two complete versions also outperform CNN-Only, indicating that it is better to dynamically generate object proposals and their segmentation, based on the gradually "refreshed" domain knowledge, than generating them all at once and being blind to suggestions of domain knowledge. (3) Combining the above two points, it shows that the close collaboration between domain knowledge and deep learning can effectively detect multiple instances of glands and villi. (4) DK-CNN-Greedy and DK-CNN-MWIS perform similarly well. This means that generating object proposals in a greedy or batch fashion (partially greedy) is not quite critical, as long as domain knowledge and the power of deep learning are fully explored. (5) Glands-Detect does not use an appropriate verification scheme to address the situation when villi, with a dual structure (causing lots of ambiguity), are also present in the images; it requires lumen regions inside glands to be obvious, and thus does not perform well on our images. (6) All four versions of our method have a higher precision than recall, this is because we use relatively high overlap thresholds. We plan to do complete ROC analysis by varying overlap thresholds.

## References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. 34(11), 2274–2282 (2012)
2. Fu, H., Qiu, G., Shu, J., Ilyas, M.: A novel polar space random field model for the detection of glandular structures. IEEE Trans. Med. Imaging 33, 764–776 (2014)
3. Girshick, R.B., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR, pp. 580–587 (2013)
4. Gunduz-Demir, C., Kandemir, M., Tosun, A.B., Sokmensuer, C.: Automatic segmentation of colon glands using object-graph. Medical Image Analysis 14(1) (2010)
5. Haeker, M., Wu, X., Abràmoff, M.D., Kardon, R., Sonka, M.: Incorporation of regional information in optimal 3-D graph search with application for intraretinal layer segmentation of optical coherence tomography images. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 607–618. Springer, Heidelberg (2007)
6. Hariharan, B., Arbeláez, P., Girshick, R., Malik, J.: Simultaneous detection and segmentation. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part VII. LNCS, vol. 8695, pp. 297–312. Springer, Heidelberg (2014)
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1106–1114 (2012)
8. Naik, S., Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A.: Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. In: MIAAB (2007)
9. Naini, B.V., Cortina, G.: A histopathologic scoring system as a tool for standardized reporting of chronic (ileo) colitis and independent risk assessment for inflammatory bowel disease. Human Pathology 43, 2187–2196 (2012)
10. Nguyen, K., Sarkar, A., Jain, A.K.: Structure and context in prostatic gland segmentation and classification. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 115–123. Springer, Heidelberg (2012)
11. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. In: ICLR (2014)