HistoSegNet: Semantic Segmentation of Histological Tissue Type in Whole Slide Images

Lyndon Chan [1], Mahdi S. Hosseini[1,4], Corwyn Rowsell [2,3], Konstantinos N. Plataniotis [1] and Savvas Damaskinos [4]
[1]The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto
[2]Division of Pathology, St. Michaels Hospital, Toronto, ON, M4N 1X3, Canada
[3]Department of Laboratory Medicine and Pathobiology, University of Toronto
[4]Huron Digital Pathology, St. Jacobs, ON, N0B 2N0, Canada
{lyndon.chan, mahdi.hosseini}@mail.utoronto.ca

Abstract

In digital pathology, tissue slides are scanned into Whole Slide Images (WSI) and pathologists first screen for diagnostically-relevant Regions of Interest (ROIs) before reviewing them. Screening for ROIs is a tedious and time-consuming visual recognition task which can be exhausting. The cognitive workload could be reduced by developing a visual aid to narrow down the visual search area by highlighting (or segmenting) regions of diagnostic relevance, enabling pathologists to spend more time diagnosing relevant ROIs. In this paper, we propose HistoSegNet, a method for semantic segmentation of histological tissue type (HTT). Using the HTT-annotated Atlas of Digital Pathology (ADP) database, we train a Convolutional Neural Network on the patch annotations, infer Gradient-Weighted Class Activation Maps, average overlapping predictions, and post-process the segmentation with a fully-connected Conditional Random Field. Our method out-performs more complicated weakly-supervised semantic segmentation methods and can generalize to other datasets without retraining.

1 Introduction

Digital pathology with Whole Slide Imaging (WSI) platforms allows pathologists to conveniently navigate tissue slides for diagnosis. Typically, pathologists quickly screen slides for tissue regions relevant to the disease being diagnosed, known as regions-of-interest (ROI) and then review these regions for tissues with abnormal appearance [31]. For instance, a pathologist diagnosing a slide for adenocarcinoma will first screen for glandular tissue regions, and review those that appear abnormally disordered before assigning a diagnosis. However, histology slides are very large images and the visual search can be exhausting. This is further complicated when pathology departments typically diagnose hundreds to thousands of slides each day [8], and diagnostic accuracy suffers when pathologists are fatigued [19, 26]. Computer-Aided Diagnosis (CAD) tools to highlight (segment) regions of diagnostic relevance as a visual aid are already widespread in radiology [40] and similar tools can be applied to pathology, where screening slides for relevant tissues occupies 36%-57% of productive time [19]. With an increasing number of histopathology cases and chronic shortage of pathologists in coming years, developing a CAD tool could decrease time to diagnosis and improve diagnostic accuracy for patients [46].

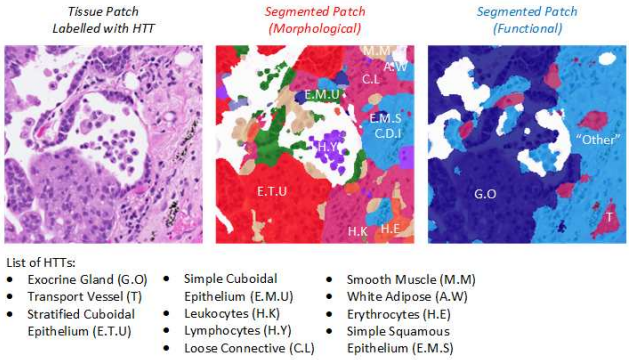


Figure 1. *Our approach trains on patch-level histological tissue type annotations and predicts morphological and functional tissue types at the pixel level.*

Semantic segmentation methods have been developed for histopathological images, but are trained on only specific tissues from specific organs for diagnosing specific diseases. If these methods were used as diagnostic aids, they would need to be retrained for every new diagnostic case. In this paper, we propose a semantic segmentation algorithm that is trained on annotated patch-level of histological tissue type (HTT) drawn from healthy tissues from different organs and predicts pixel-level labels (see Figure 1 for visu-

alization of HTTs in both patch and pixel levels). Our aim is to develop a feasible diagnostic aid to highlight regions of relevant tissues within WSI scans.

1.1 Problem and Contributions

The problem we aim to solve is to assign a semantic label to each image pixel of a WSI that is populated by diverse HTTs. The HTT of a pixel can be either non-tissue related (e.g. background, dust speck) or tissue related, in which case it can be further classified with a morphological type (the type of constituent cell) and sometimes also a functional type (whether it belongs to a glandular or vascular structure). This is similar to the stuff-thing distinction in computer vision [9], whereby all objects have "stuff" but might not be "things". In this paper we (1) propose the first publicly-released semantic segmentation tool on brightfield histopathology images for a wide variety of HTTs (> 10) trained on healthy tissues from different organs, (2) demonstrate its quantitative performance on a hand-segmented subset of the Atlas of Digital Pathology (ADP) database [11] and compare against two recent semantic segmentation methods, (3) evaluate its qualitative performance on unseen slides with a pathologist's feedback, and (4) analyze its generalizability to other pathology datasets. The novel stagewise system design is more sophisticated than previous methods for histological semantic segmentation [16, 3] and is shown to out-perform more complicated state-of-the-art weakly-supervised semantic segmentation (WSSS) methods applied to histopathology images.

1.2 Related Works

Weakly-Supervised Semantic Segmentation. Fully-supervised semantic segmentation approaches are highly accurate due to training at the pixel-level [24]. However, these annotations are time-consuming and expensive which need weak (or inexact [52]) supervision to infer pixel-level labels from image-level annotations. These methods fall under four categories: (a) graphical model-based methods which extract regions of homogeneous appearance and predict latent variable labels from the image level for each region e.g. superpixels [48] and graphlets [50], (b) multi-instance learning (MIL) based methods which constrain their optimization to assign at least one pixel per image to each image label including STF [42], MIL-ILP [29] and SPN [20], (c) self-supervised based methods which generate interim segmentation masks from image-level annotations and learn pixel-level segmentations. Some methods iterate between fine and coarse predictions, such as EM-Adapt [28] and AF-SS [30]. Other methods produce CAMs [51] or saliency maps [36] as initial seeds and refine them to train a FCN, such as [10], DCSM [34], Guided Segmentation [27], and AffinityNet [1], and (d) discriminative localization-based methods which use image-level annotations to generate discriminative object localizations as initial seeds (usually using a CNN and CAM) and then improve these iteratively, such as SEC [17], SRG [12], AE-PSL [45], and MCOF [44].

Histopathological Semantic Segmentation. In histopathological images, semantic segmentation aims to label each pixel with either diagnoses (e.g. cancer/non-cancer [2]) or tissue/cell types (e.g. gland [38], nuclei [14], mitotic/non-mitotic [32, 41]). These methods include: (a) sliding patch-based methods which train and predict at the center pixel of a sliding patch to obtain finer predictions e.g. CNNs are commonly used for mitosis [6, 25], cellular [35], neuronal [5] and gland [21, 15] segmentation, (b) superpixel-based methods which train and predict at the superpixel level e.g. CNNs applied to scaled superpixels for tissue type [47] and nuclear [39] segmentation, (c) pixel-based methods which train and predict at the pixel level and typically apply a FCN with contour separation processing [4, 22] and (d) weakly-supervised methods which train at the image and predict at the pixel levels e.g. using patch-based MIL in [49, 43].

Color	HTT (code)	% GT Images	% ADP Images*	% GT Pixels
	Background	100%	-	17.91%
	Simple Squamous Epithelial (E.M.S)	68%	18.91%	0.38%
	Simple Cuboidal Epithelial (E.M.U)	34%	29.66%	6.62%
	Simple Columnar Epithelial (E.M.O)	18%	14.34%	3.84%
	Stratified Squamous Epithelial (E.T.S)	6%	2.01%	2.51%
	Stratified Cuboidal Epithelial (E.T.U)	32%	20.73%	4.58%
	Stratified Columnar Epithelial (E.T.O)	8%	4.43%	1.02%
	Pseudostratified Epithelial (E.P)	6%	0.28%	1.18%
	Dense Irregular Connective (C.D.I)	50%	25.36%	17.07%
	Dense Regular Connective (C.D.R)	6%	0.38%	1.61%
	Loose Connective (C.L)	54%	49.63%	11.71%
	Erythrocytes (H.E)	72%	42.47%	2.28%
	Leukocytes (H.K)	32%	9.84%	0.33%
	Lymphocytes (H.Y)	60%	29.61%	1.33%
	Compact Bone (S.M.C)	2%	1.69%	1.96%
	Spongy Bone (S.M.S)	2%	1.32%	0.63%
	Endochondral Bone (S.E)	4%	0.22%	0.14%
	Hyaline Cartilage (S.C.H)	2%	0.06%	1.21%
	Marrow (S.R)	4%	0.89%	1.63%
	White Adipose (A.W)	18%	3.03%	2.69%
	Brown Adipose (A.B)	2%	0.01%	0.31%
	Marrow Adipose (A.M)	4%	0.78%	0.21%
	Smooth Muscle (M.M)	50%	23.85%	4.80%
	Skeletal Muscle (M.K)	8%	4.43%	0.68%
	Neuropil (N.P)	14%	12.44%	10.28%
	Nerve Cell Bodies (N.R.B)	12%	10.41%	1.73%
	Nerve Axons (N.R.A)	6%	0.33%	1.32%
	Microglial Cells (N.G.M)	6%	3.36%	0.05%
	Schwann Cells (N.G.W)	2%	0.12%	0.01%
-	Total	100%	100%	100.00%
	Background	72%	-	11.03%
	Other	100%	-	63.82%
	Exocrine Gland (G.O)	36%	39.48%	15.44%
	Endocrine Gland (G.N)	8%	6.31%	4.84%
	Transport Vessel (T)	82%	34.21%	4.88%
-	Total	100%	100%	100.00%

Table 1. *ADP 3rdlevel morphological (top block) and functional (bottom block) HTTs: occurrence frequency at image-level in tuning set, at image-level in entire ADP, at pixel level in tuning set.*

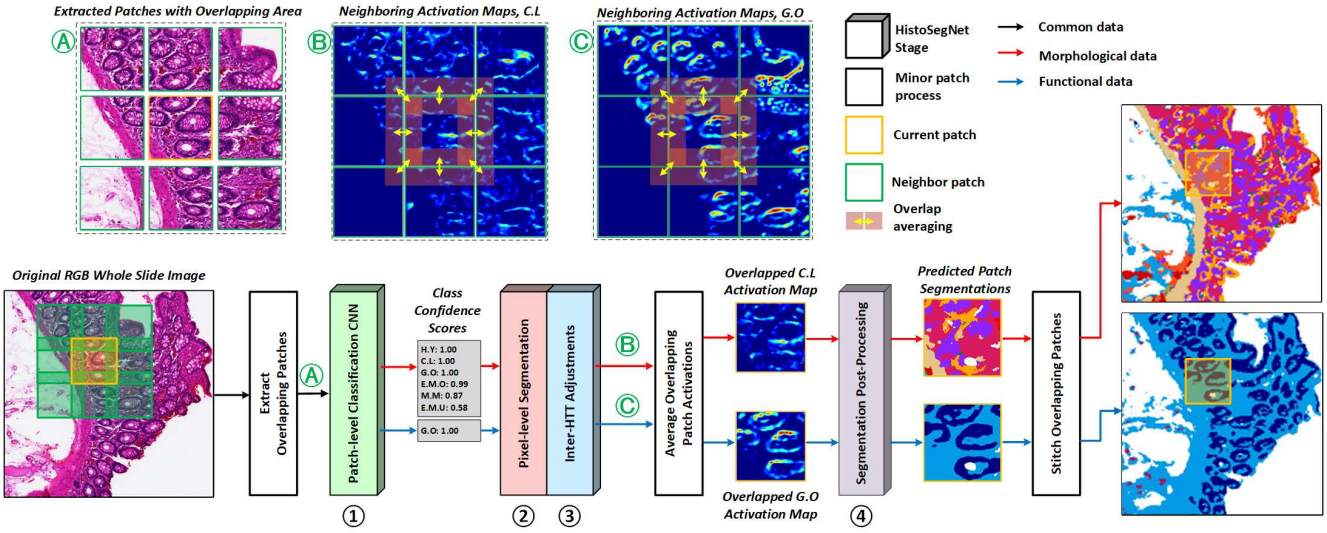


Figure 2. *The proposed HistoSegNet algorithm consists of four major stages. Initially, the whole slide image is divided into overlapping patches, then for each patch, (1) HTT confidence scores are generated with the patch-level classification CNN, (2) pixel-level class activation maps are generated, and (3) adjustments are made to the activation maps to account for relations between HTTs. Then, the activation maps of overlapping maps are averaged and (4) post-processed before being stitched together to the whole slide image level.*

2 Dataset

The ADP database was introduced in [11] and contains digital pathology patches of different healthy histological tissues with different stains from the same medical institute, labeled from a hierarchical HTT taxonomy. The images are sized 1088×1088 and scanned at 0.25μm/pixel with Huron TissueScope LE1.2 scanner. HTTs are assigned if the labeler could find at least one occurrence of a non-cellular tissue or at least five occurrences of a cellular tissue, see [11] for details. Out of the third-level tissues, we have ignored undifferentiated classes and those without any examples. We also added two non-ADP classes of "Background" for non-tissue regions and "Other" for non-functional tissue regions i.e. neither glandular nor vascular. We further separate these 31 tissues into morphological (28 in total, plus "Background") and functional (3 in total, plus "Background" and "Other") types. See Table 1 for the color-coded morphological and functional HTTs; provided in the fourth column are their occurrences in ADP. As the original ADP database was annotated at the patch level, we additionally hand-segmented a 50-patch subset to quantitatively tune our method. To ensure this tuning set would contain tissues representative of the larger ADP database, we ensured that the frequency of image occurrence for each HTT (except G.O) would be no lower than its proportion in ADP (see the third and fourth columns in Table 1). We found that hand-annotating each patch took about 18.7 minutes for the morphological types and 2.6 minutes for the functional types.

3 Methodology

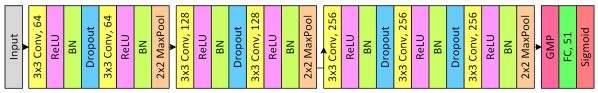
In this section, we explain our proposed four-stage HistoSegNet algorithm. Once a patch is extracted from the slide (with 25% overlap), it is passed to (1) the patch-level HTT classification Convolutional Neural Network (CNN) stage to predict possible tissue classes, (2) the pixel-level HTT segmentation stage to predict pixel-level activation maps, and (3) the inter-HTT adjustment stage to adjust the activation maps with additional information. Then, the activation maps of neighboring patches are averaged at the overlapping areas and passed to (4) the HTT segmentation post-processing stage before stitching back to the slide level. Note that HistoSegNet accepts 224×224-pixel patches that are resized from a scan resolution of $0.25 \times \frac{224}{1088} = 1.2143\mu$m/pixel. Processing is conducted mostly independently for the morphological and functional segmentation modes. We decided to overlap the patch predictions between stages (3) and (4) to minimize boundary artifacts. A summary illustration of the HistoSegNet pipeline is shown in Figure 2 and a detailed description of the constituent operations in mathematical notation can be found in the Supplementary Materials. Code and further documentation for HistoSegNet can be found online [1].

3.1 Patch-level classification CNN

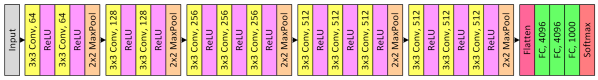
We use a HTT classification CNN to predict multiple HTT labels for a given patch. The CNN is pre-trained on predicting the 31 HTTs in the third level of the ADP database, excluding undifferentiated and absent types. Our network architecture (see Figure 3(a)) is identical to VGG-16 [37] (see Figure 3(b)) but for several important differences: (1) the softmax layer is replaced by a sigmoid layer, (2) batch normalization is added after each convolutional layer activation, and (3) the flattening layer is replaced by

[1] https://github.com/lyndonchan/hsn_v1

a global max pooling layer. Furthermore, dropout is used between normalization and convolutional layers, and we removed the last two convolutional blocks and two fully-connected layers. We decided to add batch normalization and dropout to regularize our network [13] and we used the softmax layer to implement multi-label prediction. We were inspired by the global average pooling layer [23], which reduces overfitting, to use the related global max pooling layer, since tissues are labeled regardless of their spatial extent. After some experimentation, we found that removing two convolutional blocks and fully-connected layers was optimal for improving classification performance, reducing training time, and increasing segmentation resolution (for Gradient-Weighted Class Activation Map / Grad-CAM). As a result, our network consists of three convolutional blocks, followed by a global max pooling layer, a single fully-connected layer, and a sigmoid layer. Each convolutional block consists of a single convolutional layer, a ReLU activation layer, and a batch normalization layer. Furthermore, no color normalization was applied since the same WSI scanner and staining protocol were used for all images. We provide additional validation of the CNN's performance in the Supplementary Materials.



(a) Our Modified Architecture



(b) VGG16 Architecture

Figure 3. *Adoption of the modified CNN from VGG16.*

3.2 Pixel-level Segmentation

To infer pixel-level HTT predictions from the patch-level predictions of the CNN, we use Gradient-Weighted Class Activation Maps (Grad-CAM), a weakly-supervised semantic segmentation (WSSS) method [33] which generalizes the Class Activation Map (CAM) method [51]. We decided to use Grad-CAM over similar WSSS methods because of its simplicity (no re-training is required) and versatility (it is applicable to any CNN architecture). See Figure 4 for an overview of the constituent operations of the pixel-level segmentation stage.

Grad-CAM. The Grad-CAM method first conducts a partial backpropagation from the class confidence score y_c to the final convolutional activation output $\hat{A}^L_{d_L}$ and then performs a 2D average to obtain the "importance" of the d_L-th feature map to the c-th output class, α_{c,d_L}:

$$\alpha_{c,d_L} \leftarrow \frac{1}{N_L^2} \sum_{i=1}^{N_L} \sum_{j=1}^{N_L} \frac{\partial y_c}{\partial \hat{A}^L_{d_L}(i,j)} \tag{1}$$

Then, the incoming feature maps are weighted and summed before passing them through a ReLU activation:

$$\tilde{U}_c \leftarrow \text{ReLU}\Big(\sum_{d_L=1}^{D_L} \alpha_{c,d_L} \hat{A}^L_{d_L}\Big) \tag{2}$$

Finally, the Grad-CAM is upsampled back to the original image size using bilinear interpolation.

Scaling by HTT Confidence Scores. Afterwards, we scale the Grad-CAMs by their patch-level HTT confidence scores y_c wherever they pass the confidence threshold θ_c, i.e. $\hat{U}_c \leftarrow y_c \check{U}_c$. This is necessary because Grad-CAM values invariably range from 0 to 1, so scaling them by their patch-level confidence scores ensures that confident activations get boosted relative to non-confident ones, as can be seen in the activation for E.M.U in Figure 4.

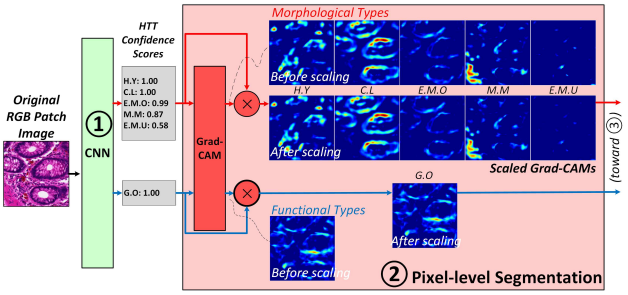


Figure 4. *The constituent operations of the pixel-level segmentation stage: activation maps are obtained with Grad-CAM and scaled with their HTT confidence scores.*

3.3 Inter-HTT Adjustments

The original ADP database has no non-tissue and non-functional labels where we must artificially produce the added "Background" for both morphological and functional modes and "Other" activations for functional mode. These activation maps must be produced to avoid predictions where no valid pixel class from ADP exists.

"Background" Activation. Background pixels in digital pathology images are known to have high white-illumination values, except for tissues which stain transparent (i.e. white adipose and brown adipose tissues for the morphological mode; and exocrine glandular, endocrine glandular, and transport vessels for the functional mode). First, the smoothed white-illumination image is obtained by applying a scaled-and-shifted sigmoid to the mean-RGB image $\overline{\mathrm{X}}$; then, we subtract the appropriate transparent-staining class activations; and finally we filter with a 2D Gaussian blur $H_{\mu,\sigma}$ to reduce the prediction resolution

$$\hat{U}_B \leftarrow \frac{0.75}{1+\exp\left[-4(\overline{\mathrm{X}}-240)\right]}$$
$$\hat{U}_B^{\text{morph}} \leftarrow (\hat{U}_B - \max(\hat{U}^{\text{A.W}}, \hat{U}^{\text{A.B}})) * H_{0,2}$$
$$\hat{U}_B^{\text{func}} \leftarrow (\hat{U}_B - \max(\hat{U}^{\text{G.O}}, \hat{U}^{\text{G.N}}, \hat{U}^{\text{T}})) * H_{0,2}.$$

"Other" Activation. For the functional mode, non-background pixels belonging to non-functional tissues intuitively must have low activations for both background and

all other functional tissues. First, we take the 2D maximum of: (1) all other functional type activations, (2) white and brown adipose activations (from the morphological mode), and (3) the background activation. Then, we subtract this probability map from one and scale it

$$\hat{U}_O^{\text{func}} \leftarrow 0.05 \left[1 - \max\left(\{\hat{U}_c^{\text{func}}\}_{c=1}^{C}, \hat{U}_B^{\text{func}}, \hat{U}_A\right)\right]. \quad (3)$$

Class-Specific Grad-CAM. A final adjustment is made to differentiate Grad-CAMs overlapping in the same patch; inspired by Shimoda *et al.* [34], we subtracted each activation map from the 2D maximum of the other Grad-CAMs, producing an activation map that we call a "Class-Specific Grad-CAM" (CSGC). Note how, in Figure 5, the functional background activation at the top of the patch is suppressed by the exocrine gland activation at the same location.

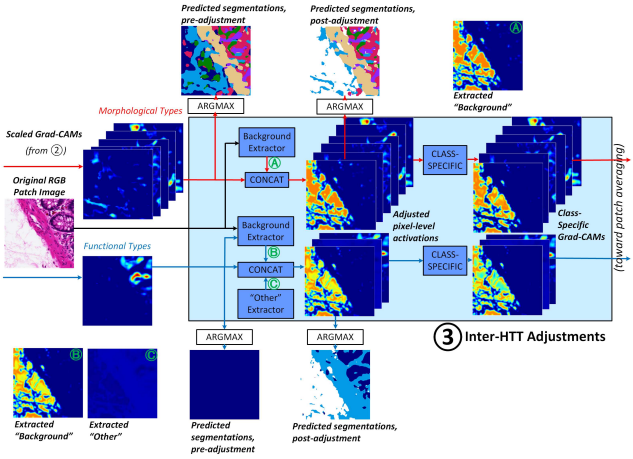


Figure 5. *The constituent operations of the inter-HTT adjustment stage: the "Background" is concatenated to the activation maps for both morphological and functional types and the "Other" activation for functional types only, then each activation map is subtracted from the maximum of the others.*

3.4 Segmentation Post-Processing

The resultant CSGCs produce blobby predictions which poorly conform to object contours - this is a well-known problem of CNN-based segmentation algorithms. Hence, we post-process the segments to maximize the visual homogeneity of their constituent pixels using a fully-connected Conditional Random Field (CRF) proposed by Krähenbühl *et al.* [18]. For multi-class segmentation, the CRF uses an appearance kernel and a smoothness kernel to compute the pairwise distance between two pixels' features $\mathbf{f} = [p, I]$ (position $p = (p_x, p_y)$ and RGB values $I = (I_R, I_G, I_B)$):

$$k(\mathbf{f}_i, \mathbf{f}_j) = w^{(1)} e^{-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2}} + w^{(2)} e^{-\frac{|p_i - p_j|^2}{2\theta_\gamma^2}}$$

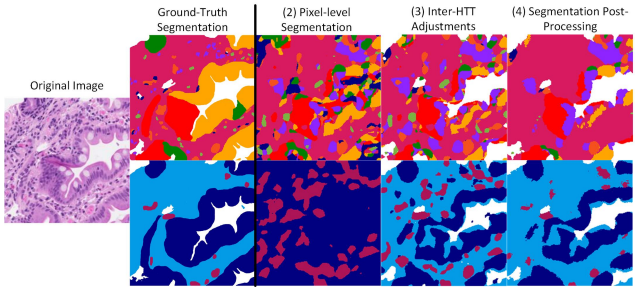
We applied the CRF for 5 iterations and used different settings for the two modes: for the morphological mode, we used $w^{(1)} = 50$, $\theta_\alpha = 10$, $\theta_\beta = 40$, $w^{(2)} = 20$, and $\theta_\gamma = 1$; for the functional mode, we used $w^{(1)} = 25$, $\theta_\alpha = 10$, $\theta_\beta = 4$, $w^{(2)} = 40$, and $\theta_\gamma = 3$.

3.5 Stage-by-Stage Ablative Analysis

To assess the contributions of each stage in our proposed method, we analyze each stage's segmentation performance in Figure 6 and runtime cost in Table 2. Our method was applied on the 50-image tuning set for measuring performance and on a typical slide of 3343 patches for measuring runtime (with an NVIDIA RTX 2070 GPU).

Stage	morph	func
(2) Pixel-level Segmentation	0.2549	0.2059
(3) Inter-HTT Adjustments	0.2067	0.5174
(4) Segmentation Post-Processing	0.2206	0.5505
Overall	0.2206	0.5505

(a) Quantitative results (mIoU)



(b) Qualitative results on sample patch

Figure 6. *Ablative analysis of HistoSegNet performance*

Table 2. *Run time of HistoSegNet (sec/img) stages: segmenting a slide of 3343 patches*

Stage	morph	func
(1) Patch-level Classification CNN	0.0050	0.0050
(2) Pixel-level Segmentation	0.3033	0.0956
(3) Inter-HTT Adjustments	0.2569	0.0212
Average Overlapping Patch Activations	0.7584	0.1855
(4) Segmentation Post-Processing	0.2644	0.2898
Stitch Overlapping Patches	0.0006	0.0006
Overall	1.5879	0.5970

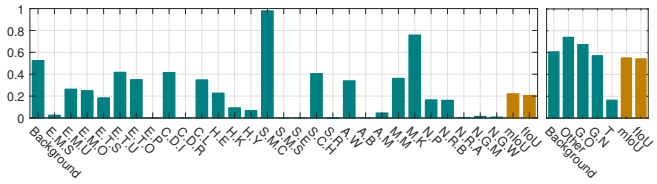
4 Results

In this section, we present our results for evaluating HistoSegNet at both patch and slide levels. For patch-level evaluation, we employ (1) the tuning set introduced in Section 2, and (2) the Gland Segmentation (GlaS) Challenge database. For slide-level evaluation, we obtained a pathologist's expert opinion on our segmentation of several unseen slides. For the following experiments, both training and testing were conducted in Keras (TensorFlow backend) with an NVIDIA GTX 1080 Ti GPU.

4.1 Quantitative Evaluation

Since we have hand-segmented 50 images from the ADP database at the pixel level, we assess the pixel-level quantitative performance of HistoSegNet and present the results below. Each patch is processed independently and for the c-th HTT, the resultant pixel-level predictions P_c

are compared with the ground truth segmentations T_c using the intersection-over-union metric (or Jaccard index) i.e. $\text{IoU}_c = |P_c \cap T_c|/|P_c \cup T_c|$. To obtain the overall performance over all HTTs, we utilize the mean IoU $\text{mIoU} = \frac{1}{C}\sum_{c=1}^{C}\text{IoU}_c$ which weighs all HTTs equally, and our custom "inverse log frequency-weighted IoU" $\text{fIoU} = \sum_{c=1}^{C}\frac{1}{\log|T_c|}\text{IoU}_c$ which weighs HTTs with fewer ground-truth pixels more. From Figures 7(a) (morphological types) and 7(b) (functional types), it can be seen that HistoSegNet performs better on the tuning set overall for functional types ($\text{mIoU} = 0.5505$, $\text{fIoU} = 0.5421$) than for morphological types ($\text{mIoU} = 0.2206$, $\text{fIoU} = 0.2057$). For the morphological mode, the best performing HTTs are Compact Bone (S.M.C) and Skeletal Muscle (M.K) and the worst performing HTTs are those with few ground-truth examples (e.g. E.P), whereas performance is more consistent for the functional mode and is worst for Transport Vessel (T).



(a) Morphological (b) Func.

Figure 7. *Intersection over Union between predicted and ground-truth segmentations in the tuning set.*

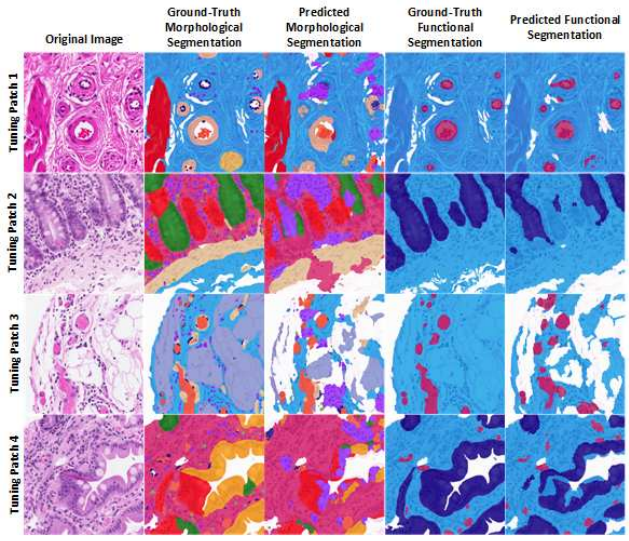


Figure 8. *Segmentation of selected patches from the tuning set, compared with the ground-truth segmentations. For the segmentation color-keys, refer to the Table 1.*

In Figure 8, the ground truth segmentations are closely approximated by predicted segmentations which occasionally recognize small regions of tissues omitted in the ground truth. Some tissues are labeled at the cellular level in the ground truth and our proposed method either segments these individual cells or the general areas occupied by them.

4.2 Pathologist Validation on unseen WSIs

The quality of HistoSegNet on unseen WSI scans are evaluated in this section by an experienced gastrointestinal pathologist. In Figure 9, we visually demonstrate the merit of producing a pixel-level diagnostic aid rather than at the patch level.

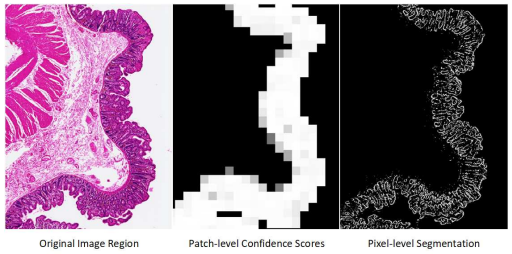


Figure 9. *The pixel-level segmentation captures fine details and shapes in the simple columnar epithelium that are lost in the patch-level prediction.*

In Figure 10, we show a WSI of H&E stained colonic tissue that was evaluated in parallel to segmented functional and morphological images. In particular, five and three different ROIs are annotated on segmented morphological and functional images. The details of evaluation on each ROI are listed in caption of the same figure. Overall, the functional segmented images were found to be highly concordant with the H&E WSI with respect to the location of exocrine glands and transport vessels, and reliably distinguished these functional tissue types from surrounding tissues. The morphological segmented image was found to correctly identify and distinguish mucosal elements including columnar epithelium, lymphocytes, and loose connective tissue, and very precisely delineated the smooth muscle of the muscularis mucosae. Erythrocytes showed a high degree of concordance. The segmented image also separated other muscular structures from other kinds of soft tissue, but was less reliable in distinguishing the specific type of muscle (smooth vs. skeletal), particularly in thicker structures such as large vessels or the muscularis propria. Nerve tissues were also not reliably separated from other soft tissue types. Further visual evaluations on more WSIs are provided in Supplementary Materials.

4.3 Comparison with State-of-the-Art WSSS

To determine whether our proposed method outperforms mainstream WSSS methods retrained on ADP, we implemented two state-of-the-art WSSS methods originally developed for the PASCAL VOC 2012 segmentation dataset [7] with code available online: SEC [17] (test mIoU of 51.7) and DSRG [12] (test mIoU of 63.4). We reconfigured these methods for the ADP database by generating foreground cues (sized 41×41 pixels) with the CNN portion of our proposed method and transferring our background/other activation maps as additional cues, and retraining the FCN portions of SEC and DSRG on the ADP database (resized

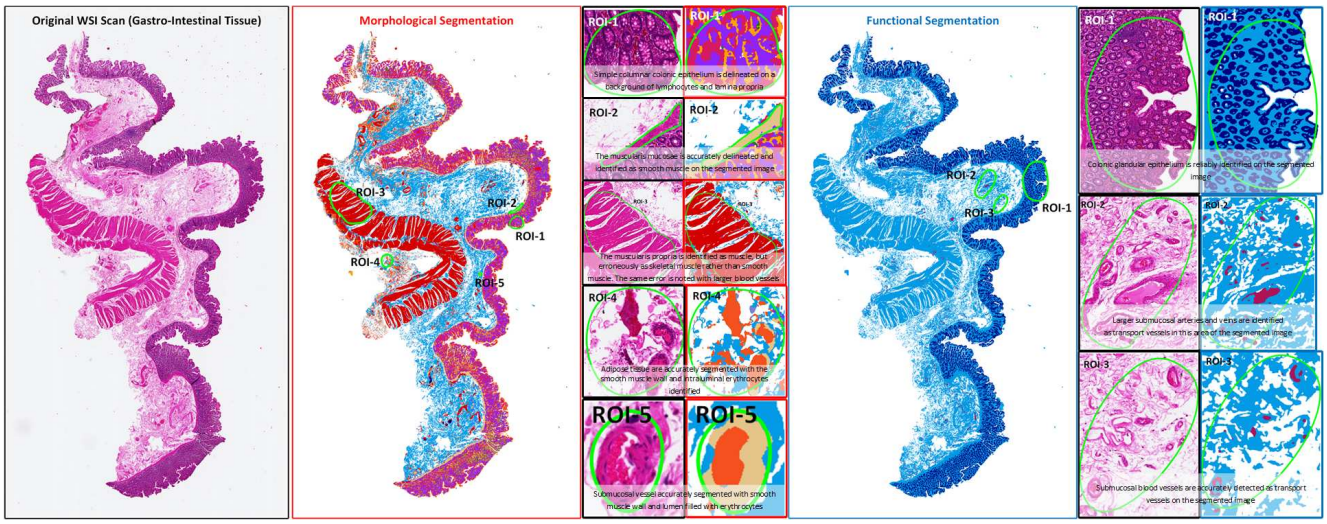
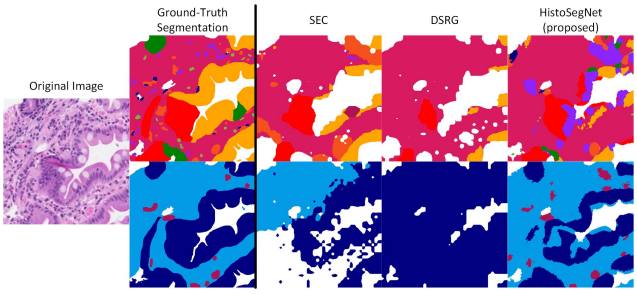


Figure 10. *Semantic segmentation results (for both morphological and functional types) on colonic tissue slide evaluated by experienced gastrointestinal pathologist in different ROIs, with displayed with comments.*

Method	morph	func
SEC [17]	0.1628	0.3225
DSRG [12]	0.1375	0.4732
HistoSegNet (proposed)	0.2206	0.5505

(a) Quantitative results (mIoU)



(b) Qualitative results on sample patch

Figure 11. *Comparison of different WSSS methods*

to 321 × 321 pixels). A step-wise learning rate decay policy was used (decay of 0.5 every 4 epochs starting from 10^{-4}) over 16 epochs with a momentum of 0.9 and batch size of 12. Quantitative results (Table 4.3) show that our proposed method outperforms both SEC and DSRG; visually, it is clear that SEC and DSRG are intended to segment larger objects but our proposed method is capable of producing finer segments (Figure 4.3).

4.4 GlaS Challenge Dataset Evaluation

The GlaS@MICCAI'2015 Gland Segmentation Challenge Contest evaluated different methods for instance segmentation of colon glands in different cancer grades from H&E-stained histology slides [38]. The images were scanned at a resolution of 0.620μm/pixel with a Zeiss MIRAX MIDI and then split into (mostly 775 × 522-pixel) patches annotated at the pixel level. In this section, we segment these GlaS images with HistoSegNet to assess its predictive performance on increasingly diseased tissues, as the ADP dataset mainly consists of healthy tissues.

As HistoSegNet accepts 224 × 224-pixel patches with a scan resolution of 1.2143μm/pixel, we first down-sampled the GlaS images by 1.9585 times, fed 224 × 224-pixel crops from the GlaS image into HistoSegNet and then overlapped the predictions before upsampling by 1.9585 times again. Also, as only two classes are available in GlaS (i.e. glandular or non-glandular), we applied HistoSegNet to predict only G.O and "Other" in functional mode.

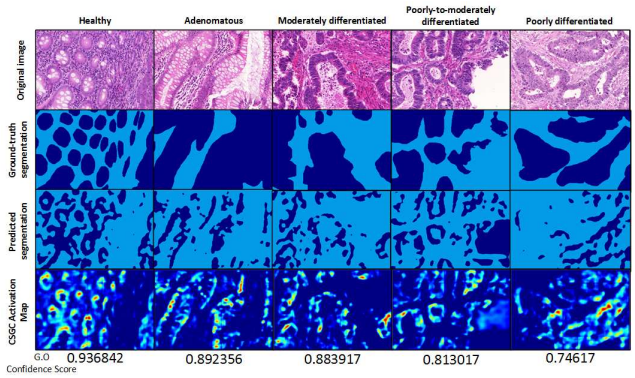


Figure 12. *Qualitative performance of HistoSegNet on select images of GlaS dataset, demonstrating that HistoSegNet consistently produces less confident predictions when given diseased tissues.*

In Figure 12, the qualitative performance of HistoSegNet can be seen on select images of the GlaS dataset. Note how HistoSegNet generally detects the outlines of the glands well, but tends not to form predictions within those outlines, which shows the generalizability of our proposed method to digital pathology images scanned with different setups. Also note how HistoSegNet's patch-level and pixel-level predictions become progressively less confident and accurate as the tumor grade worsens (from left to right) and even the predicted outline eventually misses entire glands.

In Table 3, we present a more thorough quantitative evaluation of the performance of HistoSegNet on segmenting the GlaS images at each tumor grade: the G.O confidence score and the Dice index and Hausdorff distance for both "Single Gland" and "Multiple Glands" modes. The original challenge metric evaluated the segmented glands as separate instances ("Multiple Glands" mode) but HistoSegNet tends to produce many disconnected gland predictions, so we also evaluated the segmentations as single gland objects ("Single Gland" mode). We assessed the performance at the patch level with the G.O confidence score and at the pixel level with the Dice index (measuring mask overlap) and the Hausdorff distance (measuring shape dissimilarity).

Overall, HistoSegNet is still able to segment relevant tissues in slides scanned by a different setup. Furthermore, as the tumor grade worsens, its segmentations become increasingly less confident, less overlapping, and more misshapened, which suggests that they can be used to as a predictive indicator of the level of disease in tissue.

		Single Gland		Multiple Glands	
Grade	Avg. G.O Score	Mean Dice	Mean Hausdorff	Mean Dice	Mean Hausdorff
healthy	0.9750	0.5970	130.48	0.2359	466.22
adenomatous	0.7542	0.3705	236.79	0.1468	504.25
moderately differentiated	0.7506	0.4862	138.02	0.1604	515.46
moderately-to-poorly differentated	0.7629	0.4585	153.12	0.1460	499.78
poorly differentiated	0.6628	0.4102	170.20	0.1137	584.62

Table 3. *Segmentation performance on the GlaS Challenge Dataset, over different tumour Grades: Avg. G.O Score is the mean of confidence scores for G.O across all images in that grade, semantic segmentation is evaluated for Single Gland mode, and instance segmentation is evaluated for Multiple Glands mode.*

5 Conclusion

In this paper, we presented a new semantic segmentation method for computational pathology to annotate WSIs at the pixel level with respect to different HTTs. Our method is more sophisticated than previous methods for semantic segmentation of histological tissue, but out-performs more complicated state-of-the-art WSSS methods applied to histopathology images. We achieved this by training a weakly-supervised semantic segmentation method on patch-level annotations from the Atlas of Digital Pathology database. The proposed segmentation method, which we call HistoSegNet, consists of several sequential stages. First, we trained a CNN using the ADP database, applied the Grad-CAM to construct tissue activation maps, and then performed proper HTT adjustments followed by fully-connected CRF to enhance the visual homogeneity of segmentation. We tuned our method on a hand-segmented subset of 50 images from ADP. We evaluated the quantitative performance of HistoSegNet the ground-truth tuning set and its qualitative performance on unseen WSI scans by consulting an experienced pathologist to provide a medical diagnostic opinion. We further studied the generalizability of the HistoSegNet on the GlaS gland segmentation challenge dataset to segment exocrine glands without retraining and observe how the segmentation deteriorates as the tumor grade worsens.

References

[1] Jiwoon Ahn and Suha Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. *CoRR*, abs/1803.10464, 2018. 4322

[2] Guilherme Aresta and et.al. BACH: grand challenge on breast cancer histology images. *CoRR*, abs/1808.04277, 2018. 4322

[3] Claus Bahlmann, Amar Patel, Jeffrey Johnson, Jie Ni, Andrei Chekkoury, Parmeshwar Khurd, Ali Kamen, Leo Grady, Elizabeth Krupinski, Anna Graham, et al. Automated detection of diagnostically relevant regions in h&e stained digital pathology slides. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, page 831504. International Society for Optics and Photonics, 2012. 4322

[4] Hao Chen, Xiaojuan Qi, Lequan Yu, and Pheng-Ann Heng. Dcan: Deep contour-aware networks for accurate gland segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 4322

[5] Dan Ciresan, Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems 25*, pages 2843–2851. Curran Associates, Inc., 2012. 4322

[6] Dan C. Cireşan, Alessandro Giusti, Luca M. Gambardella, and Jürgen Schmidhuber. Mitosis detection in breast cancer histology images with deep neural networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 411–418, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg. 4322

[7] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136, 2015. 4326

[8] Navid Farahani, Anil V Parwani, and Liron Pantanowitz. Whole slide imaging in pathology: advantages, limitations, and emerging perspectives. *Pathol Lab Med Int*, 7:23–33, 2015. 4321

[9] David A Forsyth, Jitendra Malik, Margaret M Fleck, Hayit Greenspan, Thomas Leung, Serge Belongie, Chad Carson, and Chris Bregler. Finding pictures of objects in large collections of images. In *International workshop on object representation in computer vision*, pages 335–360. Springer, 1996. 4322

[10] Seunghoon Hong, Donghun Yeo, Suha Kwak, Honglak Lee, and Bohyung Han. Weakly supervised semantic segmentation using web-crawled videos. *CoRR*, abs/1701.00352, 2017. 4322

[11] Mahdi S Hosseini, Lyndon Chan, Gabriel Tse, Michael Tang, Jun Deng, Sajad Norouzi, Corwyn Rowsell, Konstantinos N

Plataniotis, and Savvas Damaskinos. Atlas of digital pathology: A generalized hierarchical histological tissue type-annotated database for deep learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11747–11756, 2019. 4322, 4323

[12] Zilong Huang, Xinggang Wang, Jiasi Wang, Wenyu Liu, and Jingdong Wang. Weakly-supervised semantic segmentation network with deep seeded region growing. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7014–7023, 2018. 4322, 4326, 4327

[13] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. 4324

[14] Andrew Janowczyk and Anant Madabhushi. Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7, 2016. 4322

[15] Philipp Kainz, Michael Pfeiffer, and Martin Urschler. Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. *CoRR*, abs/1511.06919, 2015. 4322

[16] Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel, Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A Valous, Dyke Ferber, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study. *PLoS medicine*, 16(1):e1002730, 2019. 4322

[17] Alexander Kolesnikov and Christoph H. Lampert. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. *CoRR*, abs/1603.06098, 2016. 4322, 4326, 4327

[18] Philipp Krähenbühl and Vladlen Koltun. Efficient inference in fully connected crfs with gaussian edge potentials. In *Advances in neural information processing systems*, pages 109–117, 2011. 4325

[19] Elizabeth A Krupinski, Allison A Tillack, Lynne Richter, Jeffrey T Henderson, Achyut K Bhattacharyya, Katherine M Scott, Anna R Graham, Michael R Descour, John R Davis, and Ronald S Weinstein. Eye-movement study and human performance using telepathology virtual slides. implications for medical education and differences with experience. *Human pathology*, 37(12):1543–1556, 2006. 4321

[20] Suha Kwak, Seunghoon Hong, and Bohyung Han. Weakly supervised semantic segmentation using superpixel pooling network. 2017. 4322

[21] W. Li, S. Manivannan, S. Akbar, J. Zhang, E. Trucco, and S. J. McKenna. Gland segmentation in colon histology images using hand-crafted features and convolutional neural networks. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 1405–1408, April 2016. 4322

[22] Huangjing Lin, Hao Chen, Qi Dou, Liansheng Wang, Jing Qin, and Pheng-Ann Heng. Scannet: A fast and dense scanning framework for metastatic breast cancer detection from whole-slide images. *CoRR*, abs/1707.09597, 2017. 4322

[23] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013. 4324

[24] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 4322

[25] Christopher Malon and Eric Cosatto. Classification of mitotic figures with convolutional neural networks and seeded blob features. In *Journal of pathology informatics*, 2013. 4322

[26] Sanjay Mukhopadhyay, Michael D Feldman, Esther Abels, Raheela Ashfaq, Senda Beltaifa, Nicolas G Cacciabeve, Helen P Cathro, Liang Cheng, Kumarasen Cooper, Glenn E Dickey, et al. Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: A multicenter blinded randomized noninferiority study of 1992 cases (pivotal study). *The American journal of surgical pathology*, 42(1):39, 2018. 4321

[27] Seong Joon Oh, Rodrigo Benenson, Anna Khoreva, Zeynep Akata, Mario Fritz, and Bernt Schiele. Exploiting saliency for object segmentation from image level labels. *CoRR*, abs/1701.08261, 2017. 4322

[28] George Papandreou, Liang-Chieh Chen, Kevin Murphy, and Alan L. Yuille. Weakly- and semi-supervised learning of a DCNN for semantic image segmentation. *CoRR*, abs/1502.02734, 2015. 4322

[29] Pedro O. Pinheiro and Ronan Collobert. From image-level to pixel-level labeling with convolutional networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 4322

[30] Xiaojuan Qi, Zhengzhe Liu, Jianping Shi, Hengshuang Zhao, and Jiaya Jia. Augmented feedback in semantic segmentation under image level supervision. In *ECCV*, 2016. 4322

[31] Rebecca Randell, Roy A. Ruddle, Rhys Thomas, and Darren Treanor. Diagnosis at the microscope: a workplace study of histopathology. *Cognition, Technology & Work*, 14(4):319–335, Nov 2012. 4321

[32] Ludovic Roux, Daniel Racoceanu, Nicolas Loménie, Maria Kulikova, Humayun Irshad, Jacques Klossa, Frédérique Capron, Catherine Genestie, Gilles le Naour, and Metin Nafi Gurcan. Mitosis detection in breast cancer histological images an icpr 2012 contest. In *Journal of pathology informatics*, 2013. 4322

[33] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 618–626, 2017. 4324

[34] Wataru Shimoda and Keiji Yanai. Weakly supervised semantic segmentation using distinct class specific saliency maps. *Computer Vision and Image Understanding*, 2018. 4322, 4325

[35] Anat Shkolyar, Amit Gefen, Dafna Benayahu, and Hayit Greenspan. Automatic detection of cell divisions (mitosis) in live-imaging microscopy images using convolutional neural networks. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 743–746, 2015. 4322

[36] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *arXiv preprint arXiv:1312.6034*, 2013. 4322

[37] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 4323

[38] Korsuk Sirinukunwattana, Josien PW Pluim, Hao Chen, Xiaojuan Qi, Pheng-Ann Heng, Yun Bo Guo, Li Yang Wang, Bogdan J Matuszewski, Elia Bruni, Urko Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35:489–502, 2017. 4322, 4327

[39] Riku. Turkki, Nina. Linder, Panu. Kovanen, Teijo. Pellinen, and Johan. Lundin. Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples. *Journal of Pathology Informatics*, 7(1):38, 2016. 4322

[40] Bram van Ginneken, Cornelia M. Schaefer-Prokop, and Mathias Prokop. Computer-aided diagnosis: How to move from the laboratory to the clinic. *Radiology*, 261(3):719–732, 2011. PMID: 22095995. 4321

[41] Mitko Veta and et.al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. *Medical Image Analysis*, 11 2014. 4322

[42] A. Vezhnevets and J. M. Buhmann. Towards weakly supervised semantic segmentation by means of multiple instance and multitask learning. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3249–3256, June 2010. 4322

[43] Xi Wang, Hao Chen, Christopher Han-Kie Gan, Huangjing Lin, Qi Dou, Qitao Huang, Muyan Cai, and Pheng-Ann Heng. Weakly supervised learning for whole slide lung cancer image classification. 2018. 4322

[44] Xiang Wang, Shaodi You, Xi Li, and Huimin Ma. Weakly-supervised semantic segmentation by iteratively mining common object features. *CoRR*, abs/1806.04659, 2018. 4322

[45] Yunchao Wei, Jiashi Feng, Xiaodan Liang, Ming-Ming Cheng, Yao Zhao, and Shuicheng Yan. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. *CoRR*, abs/1703.08448, 2017. 4322

[46] Michael L Wilson, Kenneth A Fleming, Modupe A Kuti, Lai Meng Looi, Nestor Lago, and Kun Ru. Access to pathology and laboratory medicine services: a crucial gap. *The Lancet*, 391(10133):1927–1938, 2018. 4321

[47] Jun Xu, Xiaofei Luo, Guanhao Wang, Hannah Gilmore, and Anant Madabhushi. A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. *Neurocomputing*, 191:214–223, 2016. 4322

[48] Jia Xu, Alexander G Schwing, and Raquel Urtasun. Tell me what you see and i will show you where it is. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3190–3197, 2014. 4322

[49] Yan Xu, Jun-Yan Zhu, Eric I-Chao Chang, Maode Lai, and Zhuowen Tu. Weakly supervised histopathology cancer image segmentation and classification. *Medical Image Analysis*, 18(3):591 – 604, 2014. 4322

[50] Luming Zhang, Yue Gao, Yingjie Xia, Ke Lu, Jialie Shen, and Rongrong Ji. Representative discovery of structure cues for weakly-supervised image segmentation. *IEEE Transactions on Multimedia*, 16(2):470–479, 2014. 4322

[51] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2921–2929, 2016. 4322, 4324

[52] Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2017. 4322