Learning Whole-Slide Segmentation from Inexact and Incomplete Labels Using Tissue Graphs

Valentin Anklin[1,2], Pushpak Pati[1,2(✉)], Guillaume Jaume[1,3],
Behzad Bozorgtabar[3], Antonio Foncubierta-Rodriguez[1],
Jean-Philippe Thiran[3], Mathilde Sibony[4,5], Maria Gabrani[1],
and Orcun Goksel[1,6]

[1] IBM Research-Europe, Zurich, Switzerland
[2] ETH Zurich, Zurich, Switzerland
pus@zurich.ibm.com
[3] EPFL, Lausanne, Switzerland
[4] Cochin Hospital, Paris, France
[5] University of Paris, Paris, France
[6] Uppsala University, Uppsala, Sweden

Abstract. Segmenting histology images into diagnostically relevant regions is imperative to support timely and reliable decisions by pathologists. To this end, computer-aided techniques have been proposed to delineate relevant regions in scanned histology slides. However, the techniques necessitate task-specific large datasets of annotated pixels, which is tedious, time-consuming, expensive, and infeasible to acquire for many histology tasks. Thus, weakly-supervised semantic segmentation techniques are proposed to leverage weak supervision which is cheaper and quicker to acquire. In this paper, we propose SegGini, a weakly-supervised segmentation method using graphs, that can utilize weak *multiplex* annotations, *i.e.*, *inexact* and *incomplete* annotations, to segment arbitrary and large images, *scaling* from tissue microarray (TMA) to whole slide image (WSI). Formally, SegGini constructs a tissue-graph representation for an input image, where the graph nodes depict tissue regions. Then, it performs weakly-supervised segmentation via node classification by using *inexact* image-level labels, *incomplete* scribbles, or both. We evaluated SegGini on two public prostate cancer datasets containing TMAs and WSIs. Our method achieved state-of-the-art segmentation performance on both datasets for various annotation settings while being comparable to a pathologist baseline. Code and models are available at: https://github.com/histocartography/seg-gini.

Keywords: Weakly-supervised semantic segmentation · Scalable digital pathology · Multiplex annotations · Graphs in digital pathology

---

V. Anklin, P. Pati and G. Jaume—Contributed equally to this work.

---

1   Introduction

Automated delineation of diagnostically relevant regions in histology images is pivotal in developing automated computer-aided diagnosis systems in computational pathology. Accurate delineation assists the focus of the pathologists to improve diagnosis [33]. In particular, this attains high value in analyzing gigapixel histology images. To this end, several supervised methods have been proposed to efficiently segment glands [8,29], tumor regions [3,7], and tissue types [5]. Though these methods achieve high-quality semantic segmentation, they demand tissue, organ and task-specific dense pixel-annotated training datasets. However, acquiring such annotations for each diagnostic scenario is laborious, time-consuming, and often not feasible. Thus, weakly supervised semantic segmentation (WSS) methods [10,43] are proposed to learn from weak supervision, such as *inexact* coarse image labels, *incomplete* supervision with partial annotations, and *inaccurate* supervision where annotations may not always be ground-truth.

WSS methods using various learning approaches, such as graphical model, multi-instance learning, self-supervised learning, are reviewed in [10]. WSS methods using various types of weak annotations are presented in [2,43]. Despite the success in delivering excellent segmentation performance, mostly with natural images, WSS methods encounter challenges in histology images [10], since histology images contain, (i) finer-grained objects (*i.e.*, large intra- and inter-class variations) [34], and (ii) often ambiguous boundaries among tissue components [39]. Nevertheless, some WSS methods were proposed for histology. Among those, the methods in [13,14,16,18,35,38] perform patch-wise image segmentation and cannot incorporate global tissue microenvironment context. While [9,28] propose to operate on larger image-tiles, they remain constrained to working with fixed and limited-size images. Thus, a WSS method operating on arbitrary and large histology images by utilizing both local and global context is needed. Further, most methods focus on binary classification tasks. Though HistoSegNet [9] manages multiple classes, it requires training images with *exact* fine-grained image-level annotations. *Exact* annotations demand pathologists to annotate images beyond standard clinical needs and norms. Thus, a WSS method should ideally be able to learn from *inexact*, coarse, image-level annotations. Additionally, to generalize to other WSS tasks in histology, methods should avoid complex, task-specific post-processing steps, as in HistoSegNet [9]. Notably, WSS methods in literature only utilize a single type of annotation. Indeed, complementary information from easily or readily available multiplex annotations can boost WSS performance.

To this end, we propose SEGGINI, "SEGmentation using Graphs with Inexact aNd Incomplete labels". SEGGINI represents a histology image using a superpixel-based tissue-graph and follows a classification approach to segment it. Our major contributions are, (i) SEGGINI is the first WSS method scalable to arbitrary image sizes, unlike pixel-based WSS or fully-connected graph-based WSS [26,41], (ii) to the best of our knowledge, SEGGINI is the first WSS method to simultaneously learn from weak multiplex supervision, *i.e.*, *inexact* image-level labels and *incomplete* scribbles. (iii) SEGGINI incorporates both local and global
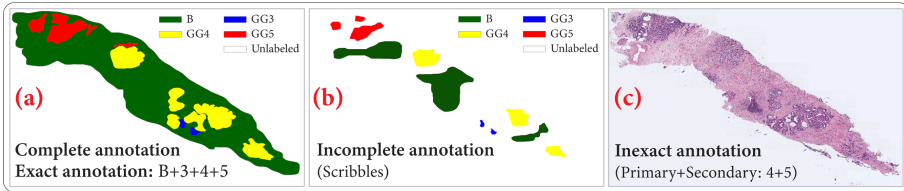
Fig. 1. Overview of various annotation types for a sample prostate cancer WSI, (a) *complete* pixel-level and *exact* image-level annotation, (b) *incomplete* scribbles of Gleason patterns, and (c) *inexact* image-level Gleason grade (P+S).

inter-tissue-region relations to build contextualized segmentation, principally in agreement with inter-pixel relation based state-of-the-art WSS method [2].

We evaluate our method on two H&E stained prostate cancer datasets [27,42] and segment Gleason patterns, *i.e.*, Benign (B), Grade3 (GG3), Grade4 (GG4) and Grade5 (GG5), by using *incomplete* scribbles of Gleason patterns and *inexact* image-level Gleason grades. Image-level grades are defined as the combination of the most common (*primary*, P) and the second most common (*secondary*, S) cancer growth patterns in the image. Figure 1 exemplifies *incomplete* and *inexact* annotations, along with *complete* pixel-level and *exact* image-level annotation.

2   Methods

This section presents the proposed SEGGINI methodology (Fig. 2) for scalable WSS of histology images. First, an input image is preprocessed and transformed into a tissue graph representation, where the graph nodes denote tissue superpixels. Then, a Graph Neural Network (GNN) learns contextualized features for the graph nodes. The resulting node features are processed by a *Graph-head*, a *Node-head*, or both based on the type of weak supervision. The outcomes of the heads are used to segment Gleason patterns. Additionally, a classification is performed to identify image-level Gleason grades from the segmentation map.

Preprocessing and Tissue Graph Construction.   An input H&E stained image X is stain-normalized using the algorithm in [31] to reduce any appearance variability due to tissue preparation. Then, the normalized image is transformed into a Tissue-Graph (TG) (Fig. 2(a)), as proposed in [20]. Formally, we define a TG as $G := (V, E, H)$, where the nodes V encode meaningful tissue regions in form of *superpixels*, and the edges E represent inter-tissue interactions. Each node $v \in V$ is encoded by a feature vector $h(v) \in \mathbb{R}^d$. We denote the node features set, $h(v), \forall v \in V$ as $H \in \mathbb{R}^{|V| \times d}$. Motivated by [6], we use superpixels as visual primitives, since rectangular patches may span multiple distinct structures.

The TG construction follows three steps: (i) superpixel construction to define V, (ii) superpixel feature extraction to define H, and (iii) graph topology construction to define E. For superpixels, we first use the unsupervised
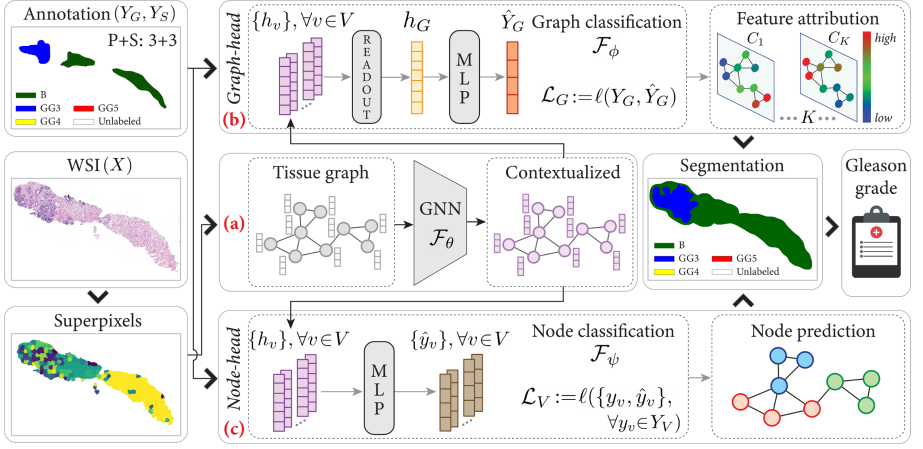
Fig. 2. Overview of the proposed SᴇɢGɪɴɪ methodology. Following superpixel extraction, (a) Tissue-graph construction and contextualization, (b) *Graph-head*: WSS via graph classification, (c) *Node-head*: WSS via node classification.

SLIC algorithm [1] emphasizing on space proximity. Over-segmented superpixels are produced at a lower magnification to capture homogeneity, offering a good compromise between granularity and noise smoothing. The superpixels are hierarchically merged based on channel-wise color similarity of superpixels at higher magnification, *i.e.*, channel-wise 8-bin color histograms, mean, standard-deviation, median, energy, and skewness. These then form the TG nodes. The merging reduces node complexity in the TG, thereby enabling a scaling to large images and contextualization to distant nodes, as explained in next section. To characterize the TG nodes, we extract morphological and spatial features. Patches of 224×224 are extracted from the original image and encoded into 1280-dimensional features with MobileNetV2 [23] pre-trained on ImageNet [11]. For a node $v \in V$, morphological features are computed as the mean of individual patch-level representations that belong to v. Spatial features are computed by normalizing superpixel centroids by the image size. We define the TG topology by constructing a region adjacency graph (RAG) [22] from the spatial connectivity of superpixels.

Contextualized Node Embeddings. Given a TG, we aim to learn discriminative node embeddings (Fig. 2(a)) that benefit from the nodes' context, *i.e.*, the tissue microenvironment and inter-tissue interactions. The contextualized node embeddings are further used for semantic segmentation. To this end, we use a GNN, that operates on graph-structured data [12,17,37]. In particular, we use Graph Isomorphism Network (GIN) [37] layers, a powerful and fast GNN architecture that functions as follows. For each node $v \in V$, GIN uses a *sum*-operator to *aggregate* the features of the node's neighbors $\mathcal{N}(v)$. Then, it *updates* the node features $h(v)$ by combining the *aggregated* features with the current node fea-

tures $h(v)$ via a multi-layer perceptron (MLP). After T GIN layers, *i.e.*, acquiring context up to T-hops, the intermediate node features $h^{(t)}(v)$, $t = 1, \ldots, T$ are concatenated to define the contextualized node embeddings [36]. Formally, a GNN $\mathcal{F}_\theta$ with batch normalization (BN) is described for $v, u \in V$ as,

$$h^{(t+1)}(v) = \text{MLP}\Big(\text{BN}\big(h^{(t)}(v) + \sum_{u \in \mathcal{N}(v)} h^{(t)}(u)\big)\Big), \, t = \{0, .., T-1\} \qquad (1)$$

$$h(v) = \text{Concat}\left(\left\{\, h^{(t)}(v) \,\Big|\, t = 1, .., T \,\right\}\right) \qquad (2)$$

Weakly Supervised Semantic Segmentation. The contextualized node embeddings $h(v)$, $\forall v \in V$ for a graph G, corresponding to an image X, are processed by SegGini to assign a class label $\in \{1, .., K\}$ to each node v, where K is the number of semantic classes. SegGini can incorporate multiplex annotations, *i.e.*, *inexact* image label Y_X and *incomplete* scribbles Y_S. Then, the weak supervisions for G are, the graph label Y_G, *i.e.*, the image label Y_X, and node labels $y_v \in Y_V$ that are extracted from Y_S by assigning the most prevalent class within each node. This is a reasonable assumption, as the tissue regions are built to be semantically homogeneous. The *Graph-head* (Fig. 2(b)) and the *Node-head* (Fig. 2(c)) are executed for using Y_G and Y_V, respectively. Noticeably, unlike [9], SegGini does not involve any post-processing, thus being a generic method that can be applied to various organs, tissue types, segmentation tasks, etc.

The *Graph-head* consists of a graph classification and a feature attribution module. First, a graph classifier $\mathcal{F}_\phi$ predicts $\hat{Y}_G$ for G. $\mathcal{F}_\phi$ includes, (i) a global average pooling *readout* operation to produce a fixed-size graph embedding h_G from the node embeddings $h(v), \forall v \in V$, and (ii) a MLP to map h_G to Y_G. As G directly encodes X, the need for patch-based processing is nullified. $\mathcal{F}_\theta$ and $\mathcal{F}_\phi$ are trained on a graph-set $\mathcal{G}$, extracted from the image-set $\mathcal{X}$, by optimizing a multi-label weighted binary cross-entropy loss $\mathcal{L}_G := l(Y_G, \hat{Y}_G)$. The class-weights are defined by $w_i = \log(N/n_i), i = 1, ..., K$, where $N = |\mathcal{X}|$, and n_i is the class example count; such that higher weight is assigned to smaller classes to mitigate class imbalance during training. Second, in an offline step, we employ a discriminative *feature attribution* technique to measure importance scores $\forall v \in V$ towards the classification of each class. Specifically, we use GraphGrad-CAM [15,21], a version of Grad-CAM [24] that can operate with GNNs. *Argmax* across class-wise node attribution maps from GraphGrad-CAM determines the node labels.

The *Node-head* simplifies image segmentation into classifying nodes $v \in V$. It inputs $h(v)$, $\forall v \in V$ to a MLP classifier $\mathcal{F}_\psi$ to predict node-labels y_v, $\forall v \in V$. $\mathcal{F}_\theta$ and $\mathcal{F}_\psi$ are trained using the multi-class weighted cross-entropy loss $\mathcal{L}_V := l(y_v, \hat{y}_v)$. The class-weights are defined by $w_i = \log(N/n_i), i = 1, ..., K$, where N is the number of annotated nodes, and n_i is the class node count. The node-wise predicted classes produce the final segmentation.

Multiplexed Supervision: For multiplex annotations, both heads are executed to perform WSS. $\mathcal{F}_\theta$, $\mathcal{F}_\phi$, and $\mathcal{F}_\psi$ are jointly trained to optimize a weighted loss

$\mathcal{L} = \lambda\mathcal{L}_G + (1 - \lambda)\mathcal{L}_V$, with which complementary information from multiplex annotations helps improve the individual classification tasks and thus improving WSS. Subsequently, we employ the classification approach in [4] to determine the Gleason grades from the generated segmentation maps.

3   Experiments

We evaluate our method on 2 prostate cancer datasets for Gleason pattern segmentation and Gleason grade classification.

UZH dataset [42] comprises of five TMAs with 886 spots, digitized at $40\times$ resolution (0.23 μm/pixel). Spots (3100×3100 pixels) contain *complete* pixel-level annotations and *inexact* image-level grades. We follow a 4-fold cross-validation at TMA-level with testing on TMA-80 as in [4]. The second pathologist annotations on the test TMAs are used as a pathologist-baseline.

SICAPv2 dataset [27] contains 18 783 patches of size 512×512 with *complete* pixel annotations and WSI-level grades from 155 WSIs at $10\times$ resolution. We reconstruct the original WSIs and annotation masks from the patches, containing up to 11000^2 pixels. We follow a 4-fold cross-validation at patient-level as in [27]. An independent pathologist's annotations are included as a pathologist-baseline.

We evaluate the methods for four annotation settings, *complete* ($\mathcal{C}$) and *incomplete* ($\mathcal{IC}$) pixel annotations, *inexact* image labels ($\mathcal{IE}$) as well as $\mathcal{IE} + \mathcal{IC}$. $\mathcal{IC}$ annotations with various pixel percentages are created by randomly selecting regions from $\mathcal{C}$ (see supplementary material for more details). We report per-class and average Dice scores as segmentation metrics, and weighted F1-score as a classification metric. We present means and standard-deviations on the test set for 4-fold cross-validation for all experiments.

Baselines: We compare SEGGINI with several state-of-the-art methods:

– UZH-CNN [4] and FSConv [27], for segmentation and classification using $\mathcal{C}$
– Neural Image Compression (NIC) [30], Context-Aware CNN (CACNN) [25], and CLAM [18], for weakly-supervised classification using $\mathcal{IE}$
– HistoSegNet [9], for weakly supervised segmentation using $\mathcal{IE}$.

These baselines are implemented based on code and algorithms in the corresponding publications. Baselines [18,25,30] directly classify WSI Gleason grades, and do not provide segmentation of Gleason patterns. Also, HistoSegNet [9] was trained herein with $\mathcal{IE}$, instead of *exact* image labels, since accessing the *exact* annotations would require using $\mathcal{C}$, that violates weak supervision constraints.

Implementations were conducted using PyTorch [19] and DGL [32] on an NVIDIA Tesla P100. SEGGINI model consists of 6-GIN layers, where the MLP in GIN, the *graph-head*, and the *node-head* contain 2-layers each with PReLU activation and 32-dimensional node embeddings, inspired by [40]. For graph augmentation, the nodes were augmented randomly with rotation and mirroring. A hyper-parameter search was conducted to find the optimal batch size $\in \{4, 8, 16\}$,

learning rate $\in \{10^{-3}, 5 \times 10^{-4}, 10^{-4}\}$, dropout $\in \{.25, .5\}$, and $\lambda \in \{.25, .5, .75\}$ for each setting. The methods were trained with Adam optimizer to select the model with the best validation loss. For fair comparison, we evaluated all the baselines with similar patch-level augmentations and hyper-parameter searches.

Table 1. Results on UZH dataset as *Mean±std* using complete ($\mathcal{C}$), inexact ($\mathcal{IE}$), incomplete ($\mathcal{IC}$), and $\mathcal{IE} + \mathcal{IC}$ settings. Setting-wise best scores are in **bold**.

Annot.		per-class Dice				avg. Dice	weight-F1
	Method	Benign	Grade3	Grade4	Grade5		
$\mathcal{C}$	UZH-CNN[4]	**69.5±6.0**	54.7±3.9	63.6±3.2	34.6±4.6	55.6±1.8	49.2±4.3
	SEGGINI	64.2±8.0	**71.3±1.9**	**72.9±2.8**	**55.6±3.3**	**66.0±3.1**	**56.8±1.7**
$\mathcal{IE}$	CLAM[18]	-	-	-	-	-	45.7±4.6
	NIC[30]	-	-	-	-	-	33.5±5.5
	CACNN[25]	-	-	-	-	-	26.1±5.1
	HistoSegNet[9]	**89.0±3.8**	42.4±10.9	56.8±10.4	34.8±12.9	55.7±3.2	41.6±9.3
	SEGGINI	63.0±9.3	**69.6±5.6**	**67.6±5.4**	**55.7±7.0**	**64.0±1.8**	**52.4±3.2**
	Pathologist	83.33	44.53	69.29	57.28	63.60	48.98

Annot.	Method	avg. Dice			
		5% pixel	10% pixel	25% pixel	50% pixel
$\mathcal{IC}$	SEGGINI	58.2±3.1	62.9±2.2	63.3±2.3	**65.3±3.1**
$\mathcal{IC} + \mathcal{IE}$	SEGGINI	**63.7±2.9**	**65.6±3.0**	**63.6±2.0**	64.2±2.6

Results and Discussion: Table 1 and 2 present the segmentation and classification results of SEGGINI and the baselines, divided in groups for their use of different annotations. For the $\mathcal{C}$ setting, SEGGINI significantly outperforms UZH-CNN [4] on per-class and average segmentation as well as classification metrics, while reaching segmentation performance comparable with pathologists. For the $\mathcal{IE}$ setting, SEGGINI outperforms HistoSegNet on segmentation and classification tasks. Interestingly, SEGGINI also outperforms the classification-tailored baselines [18,25,30]. SEGGINI delivers comparable segmentation performance for *inexact* and *complete* supervision, *i.e.*, 64% and 66% average Dice, respectively. Comparing $\mathcal{IC}$ and $\mathcal{IE} + \mathcal{IC}$, we observe that $\mathcal{IE} + \mathcal{IC}$ produces better segmentation, especially in the low pixel-annotation regime. Such improvement, however, lessens with increased pixel annotations, which is likely due to the homogeneous Gleason patterns in the test set with only one or two patterns per TMA. Notably, SEGGINI with $\mathcal{IE}$ setting outperforms UZH-CNN with $\mathcal{C}$ setting.

On SICAPv2 dataset in $\mathcal{C}$ setting, SEGGINI outperforms FSConv for both segmentation and classification, and performs comparable to the pathologist-baseline for classification. SICAPv2 is highly imbalanced with a large fraction

of benign regions. Thus, SegGini yields better results for benign class, while relatively poor performance for Grade5, which is rare in the dataset. For the $\mathcal{IE}$ setting, SegGini significantly outperforms HistoSegNet that trains using tile-labels, set the same as WSI-labels. This indicates HistoSegNet's inapplicability to WSIs with WSI-level supervision. For $\mathcal{IE}$, SegGini performs superior to [25,30] and comparable to [18]. Combining $\mathcal{IE}$ and $\mathcal{IC}$ for segmentation, the complementarity of annotations substantially boosts the performance. SegGini with $\mathcal{IE} + \mathcal{IC}$ consistently outperforms $\mathcal{IC}$ for various % of pixel annotations. Notably, $\mathcal{IE} + \mathcal{IC}$ outperforms $\mathcal{C}$ while using only 50% pixels. This confirms the benefit of learning from *multiplex* annotations. SegGini's inference time to process a WSI (11K × 11K pixels at 10×) is 14 ms, comparable to CLAM (11 ms). TG building takes 276.7 s including 183.5 s for superpixel detection and merging, 20.5 s for patch feature extraction and 85.7 s for RAG building. Figure 3 presents qualitative results on both datasets for various annotation settings. $\mathcal{IE} + \mathcal{IC}$ produces satisfactory segmentation while correcting any errors in $\mathcal{IE}$ by incorporating scribbles. The results indicate that SegGini provides competitive segmentation even with *inexact* supervision. Thus, we can leverage readily available slide-level Gleason grades from clinical reports, to substantially boost the segmentation, potentially together with a few *incomplete* scribbles from pathologists.

Table 2. Results on SICAPv2 as $Mean \pm std$ using complete ($\mathcal{C}$), inexact ($\mathcal{IE}$), incomplete ($\mathcal{IC}$), and $\mathcal{IE} + \mathcal{IC}$ settings. Setting-wise best scores are in **bold**.

Annot.	Method	per-class Dice				avg. Dice	weight-F1
		Benign	Grade3	Grade4	Grade5		
$\mathcal{C}$	FSConv [27]	$59.4_{\pm 3.0}$	$23.7_{\pm 2.6}$	$30.7_{\pm 2.7}$	$\mathbf{9.1_{\pm 2.9}}$	$31.3_{\pm 2.5}$	$59.9_{\pm 5.0}$
	SegGini	$\mathbf{90.0_{\pm 0.1}}$	$\mathbf{39.4_{\pm 3.3}}$	$\mathbf{40.2_{\pm 2.7}}$	$7.4_{\pm 2.4}$	$\mathbf{44.3_{\pm 2.0}}$	$\mathbf{62.0_{\pm 3.6}}$
$\mathcal{IE}$	CLAM[18]	-	-	-	-	-	$47.5_{\pm 4.3}$
	NIC[30]	-	-	-	-	-	$32.4_{\pm 10.0}$
	CACNN[25]	-	-	-	-	-	$21.8_{\pm 4.7}$
	HistoSegNet[9]	$\mathbf{78.1_{\pm 1.4}}$	$1.5_{\pm 0.7}$	$8.4_{\pm 0.9}$	$1.6_{\pm 0.3}$	$22.4_{\pm 0.3}$	$16.7_{\pm 4.3}$
	SegGini	$55.9_{\pm 12.0}$	$\mathbf{19.5_{\pm 6.7}}$	$\mathbf{20.7_{\pm 2.9}}$	$\mathbf{8.0_{\pm 4.2}}$	$\mathbf{26.0_{\pm 5.0}}$	$\mathbf{48.7_{\pm 6.3}}$
	Pathologist	-	-	-	-	-	63.00

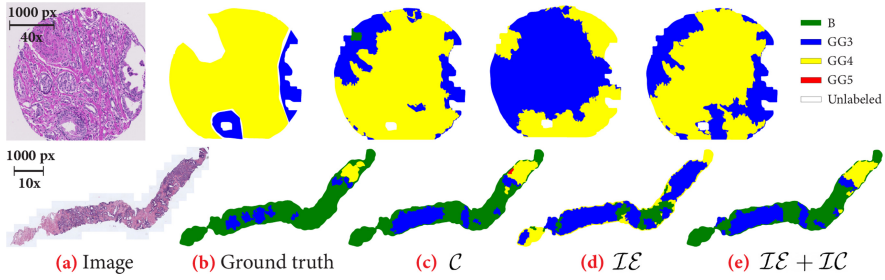| Annot.	Method	avg. Dice			
		10% pixel	25% pixel	50% pixel	100% pixel
$\mathcal{IC}$	SegGini	$37.8_{\pm 1.1}$	$\mathbf{41.9_{\pm 1.0}}$	$42.4_{\pm 0.8}$	$44.3_{\pm 2.0}$
$\mathcal{IC} + \mathcal{IE}$	SegGini	$\mathbf{39.6_{\pm 1.2}}$	$41.8_{\pm 0.6}$	$\mathbf{46.0_{\pm 0.6}}$	$\mathbf{47.0_{\pm 1.8}}$

Fig. 3. Example of predicted segmentation maps on UZH and SICAPv2 datasets for various annotation settings. $\mathcal{I}_{\mathcal{C}}$ is 10% and 25% for the datasets, respectively.

4   Conclusion

We proposed a novel WSS method, SEGGINI, to perform semantic segmentation of histology images by leveraging complementary information from weak multiplex supervision, *i.e.*, *inexact* image labels and *incomplete* scribbles. SEGGINI employs a graph-based classification that can directly operate on large histology images, thus utilizing local and global context for improved segmentation. SEGGINI is a generic method that can be applied to different tissues, organs, and histology tasks. We demonstrated state-of-the-art segmentation performance on two prostate cancer datasets for various annotation settings, while not compromising on classification results. Future research will focus on studying the generalizability of our method to previously unseen datasets.

References

1. Achanta, R., et al.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 2274–2282 (2012)
2. Ahn, J., et al.: Weakly supervised learning of instance segmentation with interpixel relations. In: IEEE CVPR, pp. 2204–2213 (2019)
3. Aresta, G., et al.: Bach: grand challenge on breast cancer histology images. Med. Image Anal. **56**, 122–139 (2019)
4. Arvaniti, E., et al.: Automated gleason grading of prostate cancer tissue microarrays via deep learning. In: Scientific Reports, vol. 8, p. 12054 (2018)
5. Bandi, P., et al.: Comparison of different methods for tissue segmentation in histopathological whole-slide images. In: IEEE ISBI, pp. 591–595 (2017)
6. Bejnordi, B., et al.: A multi-scale superpixel classification approach to the detection of regions of interest in whole slide histopathology images. In: SPIE 9420, Medical Imaging 2015: Digital Pathology, vol. 94200H (2015)
7. Bejnordi, B., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA **318**, 2199–2210 (2017)
8. Binder, T., et al.: Multi-organ gland segmentation using deep learning. In: Frontiers in Medicine (2019)

9. Chan, L., et al.: Histosegnet: semantic segmentation of histological tissue type in whole slide images. In: IEEE ICCV, pp. 10661–10670 (2019)

10. Chan, L., et al.: A comprehensive analysis of weakly-supervised semantic segmentation in different image domains. IJCV **129**, 361–384 (2021)

11. Deng, J., et al.: Imagenet: a large-scale hierarchical image database. In: IEEE CVPR, pp. 248–255 (2009)

12. Dwivedi, V., et al.: Benchmarking graph neural networks. In: arXiv (2020)

13. Ho, D., et al.: Deep multi-magnification networks for multi-class breast cancer image segmentation. In: Computerized Medical Imaging and Graphics, vol. 88, p. 101866 (2021)

14. Hou, L., et al.: Patch-based convolutional neural network for whole slide tissue image classification. In: IEEE CVPR, pp. 2424–2433 (2016)

15. Jaume, G., et al.: Quantifying explainers of graph neural networks in computational pathology. In: IEEE CVPR (2021)

16. Jia, Z., et al.: Constrained deep weak supervision for histopathology image segmentation. IEEE Trans. Med. Imaging **36**, 2376–2388 (2017)

17. Kipf, T., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR (2017)

18. Ming, Y., et al.: Data efficient and weakly supervised computational pathology on whole slide images. In: Nature Biomedical Engineering (2020)

19. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: NeurIPS, pp. 8024–8035 (2019)

20. Pati, P., et al.: Hact-net: A hierarchical cell-to-tissue graph neural network for histopathological image classification. In: MICCAI, Workshop on GRaphs in biomedicAl Image anaLysis (2020)

21. Pope, P., et al.: Explainability methods for graph convolutional neural networks. In: IEEE CVPR, pp. 10764–10773 (2019)

22. Potjer, F.: Region adjacency graphs and connected morphological operators. In: Mathematical Morphology and its Applications to Image and Signal Processing. Computational Imaging and Vision, vol. 5, pp. 111–118 (1996)

23. Sandler, M., et al.: Mobilenetv2: inverted residuals and linear bottlenecks. In: IEEE CVPR, pp. 4510–4520 (2018)

24. Selvaraju, R., et al.: Grad-cam : visual explanations from deep networks. In: IEEE ICCV, pp. 618–626 (2017)

25. Shaban, M., et al.: Context-aware convolutional neural network for grading of colorectal cancer histology images. IEEE Trans. Med. Imaging **39**, 2395–2405 (2020)

26. Shi, Y., et al.: Building segmentation through a gated graph convolutional neural network with deep structured feature embedding. ISPRS J. Photogramm. Remote. Sens. **159**, 184–197 (2020)

27. Silva-Rodríguez, J., et al.: Going deeper through the Gleason scoring scale: an automatic end-to-end system for histology prostate grading and cribriform pattern detection. In: Computer Methods and Programs in Biomedicine, vol. 195 (2020)

28. Silva-Rodríguez, J., et al.: Weglenet: a weakly-supervised convolutional neural network for the semantic segmentation of Gleason grades in prostate histology images. In: Computerized Medical Imaging and Graphics, vol. 88, p. 101846 (2021)

29. Sirinukunwattana, K., et al.: Gland segmentation in colon histology images: the Glas challenge contest. Med. Image Anal. **35**, 489–502 (2017)

30. Tellez, D., et al.: Neural image compression for gigapixel histopathology image analysis. IEEE Trans. Pattern Anal. Mach. Intell. **43**, 567–578 (2021)

31. Vahadane, A., et al.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. Med. Imaging **35**, 1962–1971 (2016)

32. Wang, M., et al.: Deep graph library: towards efficient and scalable deep learning on graphs. In: CoRR, vol. abs/1909.01315 (2019)
33. Wang, S., et al.: Pathology image analysis using segmentation deep learning algorithms. Am. J. Pathol. **189**, 1686–1698 (2019)
34. Xie, J., et al.: Deep learning based analysis of histopathological images of breast cancer. In: Frontiers in Genetics (2019)
35. Xu, G., et al.: Camel: a weakly supervised learning framework for histopathology image segmentation. In: IEEE ICCV, pp. 10681–10690 (2019)
36. Xu, K., et al.: Representation learning on graphs with jumping knowledge networks. ICML **80**, 5453–5462 (2018)
37. Xu, K., et al.: How powerful are graph neural networks? In: ICLR (2019)
38. Xu, Y., et al.: Weakly supervised histopathology cancer image segmentation and classification. Med. Image Anal. **18**, 591–604 (2014)
39. Xu, Y., et al.: Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. In: BMC bioinformatics, vol. 18 (2017)
40. You, J., et al.: Design space for graph neural networks. In: NeurIPS (2020)
41. Zhang, L., et al.: Dual graph convolutional network for semantic segmentation. In: BMVC (2019)
42. Zhong, Q., et al.: A curated collection of tissue microarray images and clinical outcome data of prostate cancer patients. In: Scientific Data, vol. 4 (2017)
43. Zhou, Z.: A brief introduction to weakly supervised learning. Natl. Sci. Rev. **5**, 44–53 (2017)