**OCELOT: Overlapped Cell on Tissue Dataset for Histopathology**  
(Supplementary Material)

Jeongun Ryu∗ Aaron Valero Puche∗ JaeWoong Shin∗ Seonwook Park Biagio Brattoli Jinhee Lee Wonkyung Jung Soo Ick Cho Kyunghyun Paeng Chan-Young Ock Donggeun Yoo Sérgio Pereira  
Lunit Inc.  
{rjw0205, aaron.valero, jwoong.shin, spark, biagio, jinhee.lee, wkjung, sooickcho, khpaeng, ock.chanyoung, dgyoo, sergio}@lunit.io

Note: We use **blue** color to refer to section numbers in the main paper. All **red** and **green** characters refer to figures, tables, and citations in this supplementary material.

**Overview.** This supplementary material includes further information regarding the implementation details, results, and datasets discussed in the main paper, and summarized as,

- We detail how the cell detection task is posed as a segmentation task, and how cell detection and tissue segmentation tasks are handled simultaneously.

- We show tissue segmentation results to provide more insights on how large FoVs and the corresponding tissue label improve cell detection performance, as discussed in Subsection 6.4 and Tab. 7.

- We share qualitative results comparing the Cell-only and Pred-to-inter-2 models in OCELOT and CARP.

- Finally, regarding the datasets, we provide the amount of annotated cells and tissue pixels per data subset of OCELOT in Tab. A.1. For TIGER [1], we describe how the dataset is pre-processed in order to be used in our experiments.

**A. Annotation Protocol (Section 3.1)**

**Annotation rules.** For cell patches, annotators were asked to annotate the center point of each cell. For tissue patches, annotators drew contours as accurately as possible.

**Consensus strategy.** All data were annotated by board-certified pathologists. Each tissue patch is annotated by a single pathologist. Each cell patch is annotated by three pathologists with the following consensus strategy. First, two pathologists annotate the same cell patch independently.

Then, the third pathologist merges the two annotations taking the discrepancies into account. This strategy was specifically designed to reduce the naturally high inter-rater variability when annotating cells.

**B. Implementation Details (Section 5.1)**

**Cell detection as segmentation.** We define the cell detection task as a segmentation one, similarly to [7]. At training time, we provide the cell labels as a segmentation map by drawing a disk centered on each cell point annotation. We use a fixed radius of 1.4 μm, corresponding to 7 pixels at a resolution of 0.2 Microns-per-Pixel (MPP). Then, we assign the value of each pixel within each disk to the corresponding cell label, e.g., 1 for TC and 2 for BC in OCELOT; 0 for the remaining background pixels. We utilize the Dice loss [6] for both cell and tissue branches, which is a widely used loss function for semantic segmentation.

At inference time, we post-process the probabilistic cell segmentation map, i.e., the output of the cell branch, to obtain a set of points, corresponding to the detection of the cells. To that end, we apply

<table>
<thead>
<tr>
<th># Pixels</th>
<th># Cell</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Train</strong></td>
<td><strong>Val</strong></td>
</tr>
<tr>
<td>BG</td>
<td>237.4M</td>
</tr>
<tr>
<td>CA</td>
<td>171.0M</td>
</tr>
<tr>
<td>UNK</td>
<td>17.3M</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>425.7M</td>
</tr>
</tbody>
</table>

(a) Tissue Annotations  
(b) Cell Annotations  

Table A.1. **Annotation statistics of OCELOT.** In (a), BG, CA, and UNK denote Background, Cancer Area, and Unknown tissue classes, respectively. The pixel counts were computed from the down-sampled tissue patches (1024 × 1024). In (b), TC and BC denote Tumor cells and Background cells, respectively.
skimage.feature.peak_local_max\(^1\) on the cell segmentation map to get the set of predicted points (cells). Lastly, we retrieve the class probability values of each cell from the segmentation maps and determine their class through \(\text{argmax}\). The class probability is used as the confidence score.

Data augmentation. During training, five data augmentations are randomly applied, including three photometric (gaussian blur, gaussian noise, color jitter) and two geometric (horizontal flipping, rotation by a multiple of 90\(^\circ\)) transformations. In the case of geometric transformations, we apply the same transformation for cell and tissue patches within a pair to maintain the physical alignment between them (e.g. 90\(^\circ\) for both cell and tissue patches).

Learning rate and dropout for cell and tissue branches. During experiments, we find that the convergence speeds of the cell detection and tissue segmentation tasks are different. The cell branch starts overfitting while the tissue branch is still learning. To address this behavior, we use different dropout probabilities and learning rates (LRs) for each branch. In the case of dropout, a fixed probability value of 0.1 is used for the tissue branch. Conversely, we tune the cell branch by performing a grid search with 3 dropout probability values: 0.1, 0.3, and 0.5. Note that the dropout layer is still learning. To address this behavior, we use different dropout layer values for each ResNet block. We use spatial dropout [8]. In the case of the LR, while searching for the best hyper-parameter values, we constrain the LR of the cell branch to be the same or half of the LR of the tissue branch. This constraint is applied to reduce the search space.

C. More Cell Detection Baselines (Section 6.1)

We provide more cell detection baselines (U-Net [5] and MFoVCE-Net [2]) on the OCELOT dataset. MFoVCE-Net is a strong baseline that further utilizes a large FoV patch as an input, but not a corresponding tissue annotation. Tab. C.1 shows that the proposed Pred-to-inter-2 model still outperforms all the baselines by a large margin. This emphasizes the importance of the additional larger FoV input and associated tissue label. In addition, the U-Net architecture shows lower performance than DeepLabV3+ [3].

D. Ablation Study: Tissue Performance (Section 6.4)

Through the ablation study in the Tab. 7 of the main document, we observe improvements in cell detection performance by utilizing a large FoV or tissue segmentation label. Moreover, utilizing both components simultaneously shows synergy, leading to an even better performance improvement. In Tab. D.1, we investigate the tissue segmentation performance to better understand the reason for such synergy. By comparing the second and last rows in Tab. D.1, we observe that training with large input/label FoV tissue results in a better tissue model, which achieves higher mIoU in both validation and test sets. Therefore, the cell detection performance boost can be justified by the fact that the tissue model shares more accurate tissue information to the cell branch.

E. Qualitative Results (Section 6.5)

We provide more examples for qualitative comparison between Cell-only and Pred-to-inter-2 models. Visualizations of OCELOT can be found in Fig. E.1 and CARP in Fig. E.2. We use a different color scheme for each figure since each dataset is based on different staining methods. The color scheme can be found in the captions.

F. Details about TIGER (Section 6.1)

Annotations. There is a single class annotation for the cell task, namely, lymphocyte cells. In contrast, 7 classes are considered for the tissue task: Invasive Tumor, Tumor-associated Stroma, In-situ Tumor, Healthy Glands, Necrosis not in-situ, Inflamed Stroma, and Rest. In addition, TIGER considers the tissue class Excluded, which has the same role as Unknown in OCELOT.

Based on the statistics in Tab. F.1, we observe that most of the lymphocyte cells are located within stroma tissue areas, i.e., Tumor-associated Stroma and Inflamed Stroma. Also, the tissue annotations suffer from severe class imbalance. In fact, the frequencies of 4 out of 7 classes are lower than 5%. To make the tissue task more straightforward, while maintaining the interrelation between lymphocyte cells and stroma tissue, we remap the tissue classes as follows: Tumor-associated Stroma and Inflamed Stroma are grouped into the Stroma (ST) class, and the remaining labels are remapped to BG class. Note that the main goal of this work is to explore cell-tissue relationships for improving the cell detection task, not to tackle the tissue segmentation task explicitly.

<table>
<thead>
<tr>
<th>Method</th>
<th>Architecture</th>
<th>Val</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell-only</td>
<td>DeepLabV3+ [3]</td>
<td>68.87±1.76</td>
<td>64.44±1.82</td>
</tr>
<tr>
<td>Cell-only</td>
<td>U-Net</td>
<td>67.75±1.42</td>
<td>63.46±1.59</td>
</tr>
<tr>
<td>Cell-only†</td>
<td>MFoVCE-Net [2]</td>
<td>69.14±0.52</td>
<td>67.12±1.96</td>
</tr>
<tr>
<td>Pred-to-inter-2</td>
<td>DeepLabV3+ [3]</td>
<td>72.68±1.58</td>
<td>71.23±0.96</td>
</tr>
</tbody>
</table>

\(^1\)https://scikit-image.org/docs/stable/api/skimage.feature.html#skimage.feature.peak_local_max
A pre-processing step is necessary for the TIGER dataset due to the inconsistent annotated ROI sizes for both cell and tissue samples. We can identify two different subsets in TIGER: 1) the sample pairs from the TIGER dataset after pre-processing, and 2) the pairs from other sources, which we denote as non-TIGA pairs. On one hand, TCGA samples (see Algo. 1) and non-TCGA samples (see Algo. 2). As a result of this pre-processing step, each non-TCGA tissue patch is paired to \(4^2\) different cell sub-patches. In contrast, for each cell ROI in TCGA, there can be up to \(4^2\) surrounding tissue patches. Please, refer to Tab. F.2 for a comparison of the statistics across OCELOT, CARP, and the pre-processed TIGER datasets.

### References


<table>
<thead>
<tr>
<th>Tissue Class Name</th>
<th>Class Ratio</th>
<th>LC in Tissue Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasive Tumor</td>
<td>27.11 %</td>
<td>4.14 %</td>
</tr>
<tr>
<td>Tumor-associated Stroma</td>
<td>27.48 %</td>
<td>30.36 %</td>
</tr>
<tr>
<td>In-situ Tumor</td>
<td>4.86 %</td>
<td>0.14 %</td>
</tr>
<tr>
<td>Healthy Glands</td>
<td>3.05 %</td>
<td>0.66 %</td>
</tr>
<tr>
<td>Necrosis not in-situ</td>
<td>1.48 %</td>
<td>0.09 %</td>
</tr>
<tr>
<td>Inflamed Stroma</td>
<td>3.31 %</td>
<td>58.98 %</td>
</tr>
<tr>
<td>Rest</td>
<td>31.67 %</td>
<td>4.65 %</td>
</tr>
<tr>
<td>Excluded</td>
<td>1.04 %</td>
<td>0.98 %</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Cell FoV</th>
<th>Tissue FoV</th>
<th>MPP</th>
<th># of patch pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCELOT</td>
<td>1024</td>
<td>4096</td>
<td>~0.2</td>
<td>673</td>
</tr>
<tr>
<td>CARP</td>
<td>1024</td>
<td>4096</td>
<td>~0.2</td>
<td>6,480</td>
</tr>
<tr>
<td>TIGER*</td>
<td>128</td>
<td>512</td>
<td>~0.5</td>
<td>9,888</td>
</tr>
</tbody>
</table>

Table F.1. **TIGER class ratio.** LC denotes lymphocyte cell. We observe that most LCs are located within stroma regions.

Table F.2. **Dataset comparison after TIGER pre-processing.** TIGER* denotes the TIGER dataset after pre-processing. # of patch pairs includes all the samples in training, validation, and test.

### Data pre-processing

A pre-processing step is necessary for the TIGER dataset due to the inconsistent annotated ROI sizes for both cell and tissue samples. We can identify two different subsets in TIGER: 1) the sample pairs from the TCGA database, and 2) the pairs from other sources, which we denote as non-TCGA pairs. On one hand, TCGA samples are composed of large annotated tissue patches that contain several smaller cell annotated ROIs within their region. The number of cell ROIs per sample highly varies, reaching up to 58. These cell ROIs are variable in size and most of them are smaller than \(256 \times 256\) pixels. On the other hand, non-TCGA samples have a complete overlap between the cell and tissue patches, and the size of these patches is larger than \(512 \times 512\) pixels.

To maximize the amount of usable cell-tissue sample pairs, while maintaining the 4 times FoV difference across the cell and tissue tasks (as done in OCELOT and CARP), we define the cell and tissue FoVs to be \(128 \times 128\) and \(512 \times 512\) pixels, respectively. Note that the image patch size is considerably smaller than in OCELOT and CARP mainly because of the limited size of cell ROIs in TCGA samples. In addition, the pre-processing step is implemented differently according to the data source; TCGA samples (see Algo. 1) and non-TCGA samples (see Algo. 2). As a result of this pre-processing step, each non-TCGA tissue patch is paired to \(4^2\) different cell sub-patches. In contrast, for each cell ROI in TCGA, there can be up to \(4^2\) surrounding tissue patches.
Algorithm 1 TIGER pre-processing step for TCGA samples

1: **Input** $D_{TCGA}$  \> TCGA dataset
2: **Output** $D_{proc}$  \> Pre-processed dataset
3: $Sz_c$, $Sz_t$ ← 128, 512  \> Cell and tissue patch sizes, respectively
4: $D_{proc}$ ← [ ]  \> Initialize pre-processed dataset to empty list
5: **for all** $(img_t, img_cROIs, masks_c, boxes_cROIs, infoROIs)$ in $D_{TCGA}$ do  \> Loop over the dataset
6: $H_t, W_t$ ← Size($img_t$)  \> Tissue image dimension
7: **for all** $(img_c, boxes_c, info_c)$ in $(img_cROIs, boxes_cROIs, infoROIs)$ do  \> Loop over cell ROIs in a sample
8: $H_c, W_c$ ← Size($img_c$)  \> Cell image dimension
9: $y_c, x_c$ ← GetROILocation($info_c$)  \> Getting the top-left coordinates of the cell ROI
10: if $W_c < Sz_c$ or $H_c < Sz_c$ then  \> Ignore small cell ROIs
11:     continue
12: $img_c$ ← Crop($img_c$, (0, 0), $Sz_c$)  \> Cropping cell ROI from the top-left corner (0,0) and size $Sz_c$
13: $boxes_c$ ← FilterBoxes($boxes_c$, $y_c, x_c, Sz_c$)  \> Removing cell bounding boxes due to previous cropping
14: **for all** $(i_c, j_c) ∈ [0..Sz_c/Sz_t] × [0..Sz_t/Sz_c]$ do  \> Loop over 16 surrounding tissue patches per cell ROI
15:     $y_c, x_c$ ← $y_c - i_c ∙ Sz_c, x_c - j_c ∙ Sz_c$  \> Defining surrounding tissue coordinates
16:     if CheckTissueExceedsImg($y_c, x_c, Sz_t, H_t, W_t$) then  \> Ignore tissue patches exceeding the image
17:         continue
18: $curImg_t$ ← Crop($img_t$, ($y_c, x_c$), $Sz_c$)  \> Cropping tissue surrounding patch
19: $curMask_t$ ← Crop($mask_c$, ($y_c, x_c$), $Sz_c$)  \> Matching tissue to cell size
20: $resImg_t$ ← Resize($curImg_t$, ($Sz_c, Sz_c$))  \> Tissue image dimension
21: $resMask_t$ ← Resize($curMask_t$, ($Sz_c, Sz_c$))  \> Consider only large samples
22: AppendTo($D_{proc}$, ($img_c$, $boxes_c$, $resImg_t$, $resMask_t$))  \> Save sample in pre-processed dataset

Algorithm 2 TIGER pre-processing step for non-TCGA samples

1: **Input** $D_{nonTCGA}$  \> Non-TCGA dataset
2: **Output** $D_{proc}$  \> Pre-processed dataset
3: $D_{proc}$ ← [ ]  \> Cell and tissue patch sizes, respectively
4: $Sz_c$, $Sz_t$ ← 128, 512  \> Looping over perfectly overlapping cell-tissue images
5: **for all** $(img, mask, boxes)$ in $D_{nonTCGA}$ do  \> Tissue 2D patch loop
6: $H, W$ ← Size($img$)  \> Define top-left coordinates of the tissue patch
7: if $H > 1024$ and $W > 1024$ then  \> Cropping tissue patch
8:     for all $(i_c, j_c) ∈ [0..H_t/Sz_t] × [0..W_t/Sz_c]$ do  \> Matching tissue to cell size
9:         $y_c, x_c$ ← $i_c ∙ Sz_c, j_c ∙ Sz_t$  \> Cell 2D sub-patch loop
10: $img_t$ ← Crop($img$, ($y_c, x_c$), $Sz_c$)  \> Define cell patch coordinates
11: $mask_t$ ← Crop($mask$, ($y_c, x_c$), $Sz_c$)  \> Cropping cell sub-patch in tissue patch
12: $resImg_t$ ← Resize($img_t$, ($Sz_c, Sz_c$))  \> Removing bboxes off the cell sub-patch
13: $resMask_t$ ← Resize($mask_t$, ($Sz_c, Sz_c$))  \> Save sample in pre-processed dataset
14: $boxInPatch$ ← FilterBoxes($boxes$, $y_c + y_t, x_c + x_t, Sz_c$)  \> Removing cell bounding boxes
15: AppendTo($D_{proc}$, ($img_c$, $boxes_c$, $resImg_t$, $resMask_t$))  \> Save sample in pre-processed dataset
Figure E.1. **Qualitative results - OCELOT.** The *Pred-to-inter-2* model can correct the mistakes of the *Cell-only* model by incorporating tissue prediction information during cell prediction. The colors represent the following classes: ● Tumor Cells (TC), ○ Background Cells (BC), and ■ Cancer Area (CA).
Figure E.2. **Qualitative results - CARP.** The *Pred-to-inter-2* model can correct the mistakes of the *Cell-only* model by incorporating tissue prediction information during cell prediction. The colors represent the following classes: • PD-L1 positive tumor cells (TC+), ● PD-L1 negative tumor cells (TC-), and ▲ Cancer Area (CA).