

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.

*: Equal contribution

	# Pixels			# Cell			
	Train	Val	Test	Train	Val	Test	
BG	237.4M	79.1M	71.6M	TC	43.8K	16.3K	12.9K
CA	171.0M	57.8M	58.8M	BC	23.6K	8.4K	9.7K
UNK	17.3M	6.7M	6.1M				
Total	425.7M	143.6M	136.3M	Total	67.4K	24.7K	22.6K

(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.

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 $1.4 \mu\text{m}$, 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

`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 `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°) 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° 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 added at the end of 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 MFoVCENet [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

¹https://scikit-image.org/docs/stable/api/skimage.feature.html#skimage.feature.peak_local_max

Method	Architecture	Val	Test
Cell-only	DeepLabV3+ [3]	68.87±1.76	64.44±1.82
Cell-only	U-Net	67.75±1.42	63.46±4.59
Cell-only†	MFoVCE-Net [2]	69.14±0.52	67.12±1.96
<i>Pred-to-inter-2</i>	DeepLabV3+ [3]	72.68±1.58	71.23±0.96

Table C.1. **More cell detection baselines.** Comparison with various cell detection methods. † denotes that a large FoV patch is also utilized as an input.

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.

Further utilize		<i>Cell (mFI)</i>		<i>Tissue (mIoU)</i>	
Tissue label	Large FoV	Val	Test	Val	Test
-	-	68.87 \pm 1.76	64.45 \pm 1.82	N/A	N/A
✓	-	70.36 \pm 1.34	66.58 \pm 2.02	75.27 \pm 3.10	73.75 \pm 3.74
-	✓	69.61 \pm 1.59	66.22 \pm 1.53	N/A	N/A
✓	✓	72.30\pm0.73	68.91\pm2.52	77.48\pm1.96	81.97\pm1.75

Table D.1. **Ablation study on the tissue segmentation performance.** Tissue segmentation performance is further reported beyond the ablation study in Tab. 7. The first row corresponds to the *Cell-only* model, and the third row includes the tissue branch with large FoV, but without tissue supervision. Since the models in two of the rows do not consider the tissue labels, we denote their performance as *N/A*. Note that the *Feature-sharing* model and the OCELOT dataset are used.

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

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

Dataset	Cell FoV	Tissue FoV	MPP	# of patch pairs
OCELOT	1024	4096	\sim 0.2	673
CARP	1024	4096	\sim 0.2	6,480
TIGER*	128	512	\sim 0.5	9,888

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 [4] 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 256×256 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 512×512 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 128×128 and 512×512 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. Please, refer to Tab. F.2 for a comparison of the statistics across OCELOT, CARP, and the pre-processed TIGER datasets.

References

- [1] TIGER: Grand Challenge. <https://tiger.grand-challenge.org>, 2022. [Online; accessed Nov-2022]. 1
- [2] Tian Bai, Jiayu Xu, and Fuyong Xing. Multi-field of view aggregation and context encoding for single-stage nucleus recognition. In *MICCAI*, pages 382–392. Springer, 2020. 2
- [3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, pages 801–818, 2018. 2
- [4] Carolyn Hutter and Jean Claude Zenklusen. The cancer genome atlas: creating lasting value beyond its data. *Cell*, 173(2):283–285, 2018. 3
- [5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. 2
- [6] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sébastien Ourselin, and M. Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *CoRR*, abs/1707.03237, 2017. 1
- [7] Zaneta Swiderska-Chadaj, Hans Pinckaers, Mart van Rijthoven, Maschenka Balkenhol, Margarita Melnikova, Oscar Geessink, Quirine Manson, Mark Sherman, Antonio Polonia, Jeremy Parry, et al. Learning to detect lymphocytes in immunohistochemistry with deep learning. *Medical Image Analysis*, 58:101547, 2019. 1
- [8] Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 648–656, 2015. 2

Algorithm 1 TIGER pre-processing step for TCGA samples

```
1: Input  $D_{TCGA}$  ▷ TCGA dataset
2: Output  $D_{proc}$  ▷ Pre-processed dataset
3:  $Sz_c, Sz_t \leftarrow 128, 512$  ▷ Cell and tissue patch sizes, respectively
4:  $D_{proc} \leftarrow []$  ▷ Initialize pre-processed dataset to empty list
5: for all  $(img_t, imgs_c^{ROIs}, masks_t, bboxes_c^{ROIs}, info^{ROIs})$  in  $D_{TCGA}$  do ▷ Loop over the dataset
6:    $H_t, W_t \leftarrow \text{Size}(img_t)$  ▷ Tissue image dimension
7:   for all  $(img_c, bboxes_c, info_c)$  in  $(imgs_c^{ROIs}, bboxes_c^{ROIs}, info^{ROIs})$  do ▷ Loop over cell ROIs in a sample
8:      $H_c, W_c \leftarrow \text{Size}(img_c)$  ▷ Cell image dimension
9:      $y_c, x_c \leftarrow \text{GetROILocation}(info_c)$  ▷ 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:     $img_c \leftarrow \text{Crop}(img_c, (0, 0), Sz_c)$  ▷ Cropping cell ROI from the top-left corner (0,0) and size  $Sz_c$ 
13:     $bboxes_c \leftarrow \text{FilterBboxes}(bboxes_c, y_c, x_c, Sz_c)$  ▷ Removing cell bounding boxes due to previous cropping
14:    for all  $(i_t, j_t) \in [0..Sz_t/Sz_c] \times [0..Sz_t/Sz_c]$  do ▷ Loop over 16 surrounding tissue patches per cell ROI
15:       $y_t, x_t \leftarrow y_c - i_t \cdot Sz_c, x_c - j_t \cdot Sz_c$  ▷ Defining surrounding tissue coordinates
16:      if  $\text{CheckTissueExcelsImg}(y_t, x_t, Sz_t, H_t, W_t)$  then ▷ Ignore tissue patches exceeding the image
17:        continue
18:       $curImg_t \leftarrow \text{Crop}(img_t, (y_t, x_t), Sz_t)$  ▷ Cropping tissue surrounding patch
19:       $curMask_t \leftarrow \text{Crop}(mask_t, (y_t, x_t), Sz_t)$ 
20:       $resImg_t \leftarrow \text{Resize}(curImg_t, (Sz_c, Sz_c))$  ▷ Matching tissue to cell size
21:       $resMask_t \leftarrow \text{Resize}(curMask_t, (Sz_c, Sz_c))$ 
22:       $\text{AppendTo}(D_{proc}, (img_c, bboxes_c, resImg_t, resMask_t))$  ▷ 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:  $D_{proc} \leftarrow []$ 
4:  $Sz_c, Sz_t \leftarrow 128, 512$  ▷ Cell and tissue patch sizes, respectively
5: for all  $(img, mask, bboxes)$  in  $D_{nonTCGA}$  do ▷ Looping over perfectly overlapping cell-tissue images
6:    $H, W \leftarrow \text{Size}(img)$  ▷ Image dimension
7:   if  $H > 1024$  and  $W > 1024$  then ▷ Consider only large samples
8:     for all  $(i_t, j_t) \in [0..H_t/Sz_t] \times [0..W_t/Sz_t]$  do ▷ Tissue 2D patch loop
9:        $y_t, x_t \leftarrow i_t \cdot Sz_t, j_t \cdot Sz_t$  ▷ Define top-left coordinates of the tissue patch
10:       $img_t \leftarrow \text{Crop}(img, (y_t, x_t), Sz_t)$  ▷ Cropping tissue patch
11:       $mask_t \leftarrow \text{Crop}(mask, (y_t, x_t), Sz_t)$ 
12:       $resImg_t \leftarrow \text{Resize}(img_t, (Sz_c, Sz_c))$  ▷ Matching tissue to cell size
13:       $resMask_t \leftarrow \text{Resize}(mask_t, (Sz_c, Sz_c))$ 
14:      for all  $(i_c, j_c) \in [0..Sz_t/Sz_c] \times [0..Sz_t/Sz_c]$  do ▷ Cell 2D sub-patch loop
15:         $y_c, x_c \leftarrow i_c \cdot Sz_c, j_c \cdot Sz_c$  ▷ Define cell patch coordinates
16:         $img_c \leftarrow \text{Crop}(img_t, (y_c, x_c), Sz_t)$  ▷ Cropping cell sub-patch in tissue patch
17:         $bboxInPatch \leftarrow \text{FilterBboxes}(bboxes, y_c + y_t, x_c + y_t, Sz_c)$  ▷ Removing bboxes off the cell sub-patch
18:         $\text{AppendTo}(D_{proc}, (img_c, bboxInPatch, resImg_t, resMask_t))$  ▷ Save sample in pre-processed dataset
```

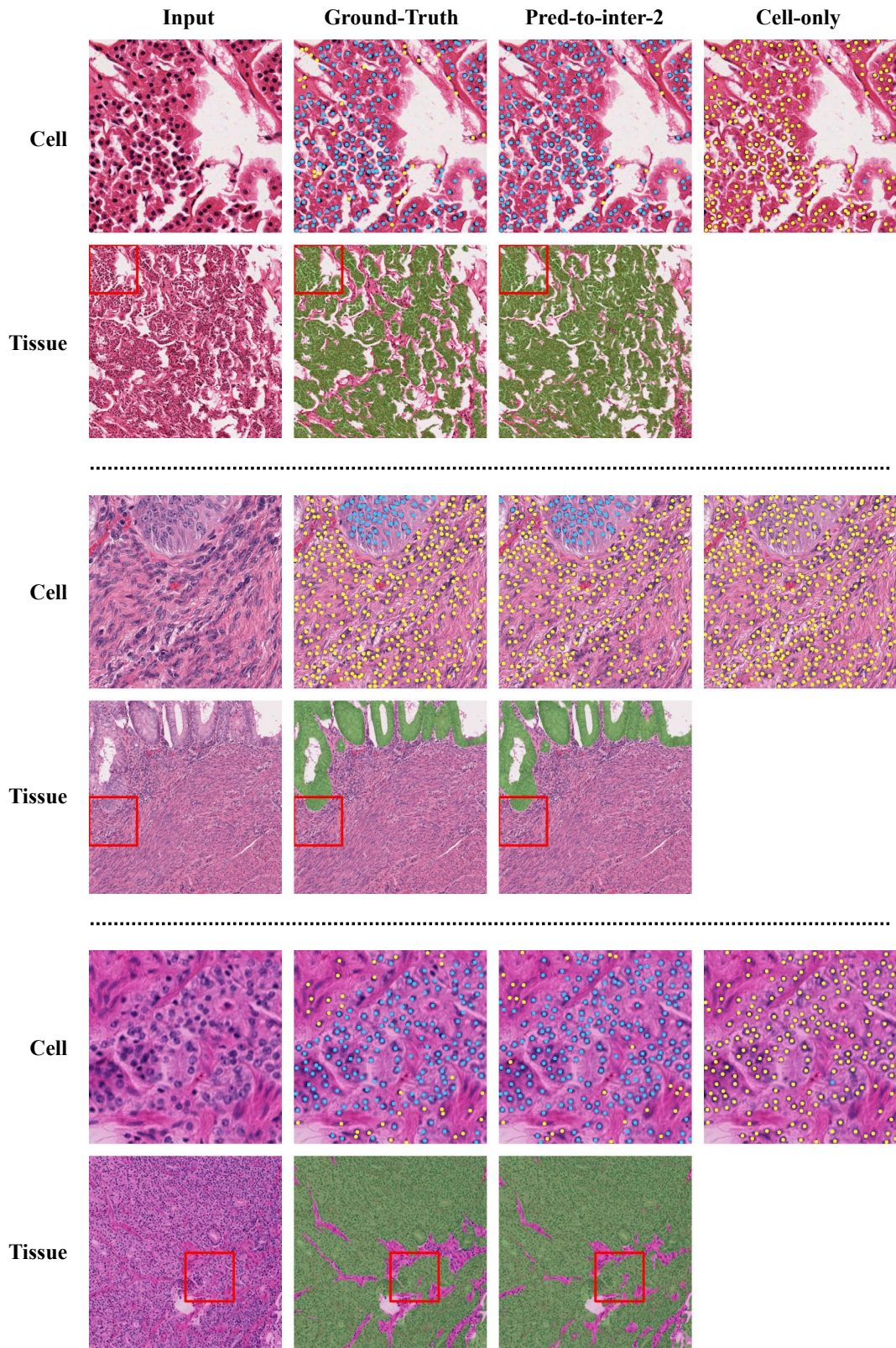


Figure E.1. **Qualitative results - OCELOT.** The *Pred-to-inter-2* model can correct the mistakes of the *Cell-only* model by incorporating tissue prediction information during cell prediction. The colors represent the following classes: ● Tumor Cells (TC), ● Background Cells (BC), and ■ Cancer Area (CA).

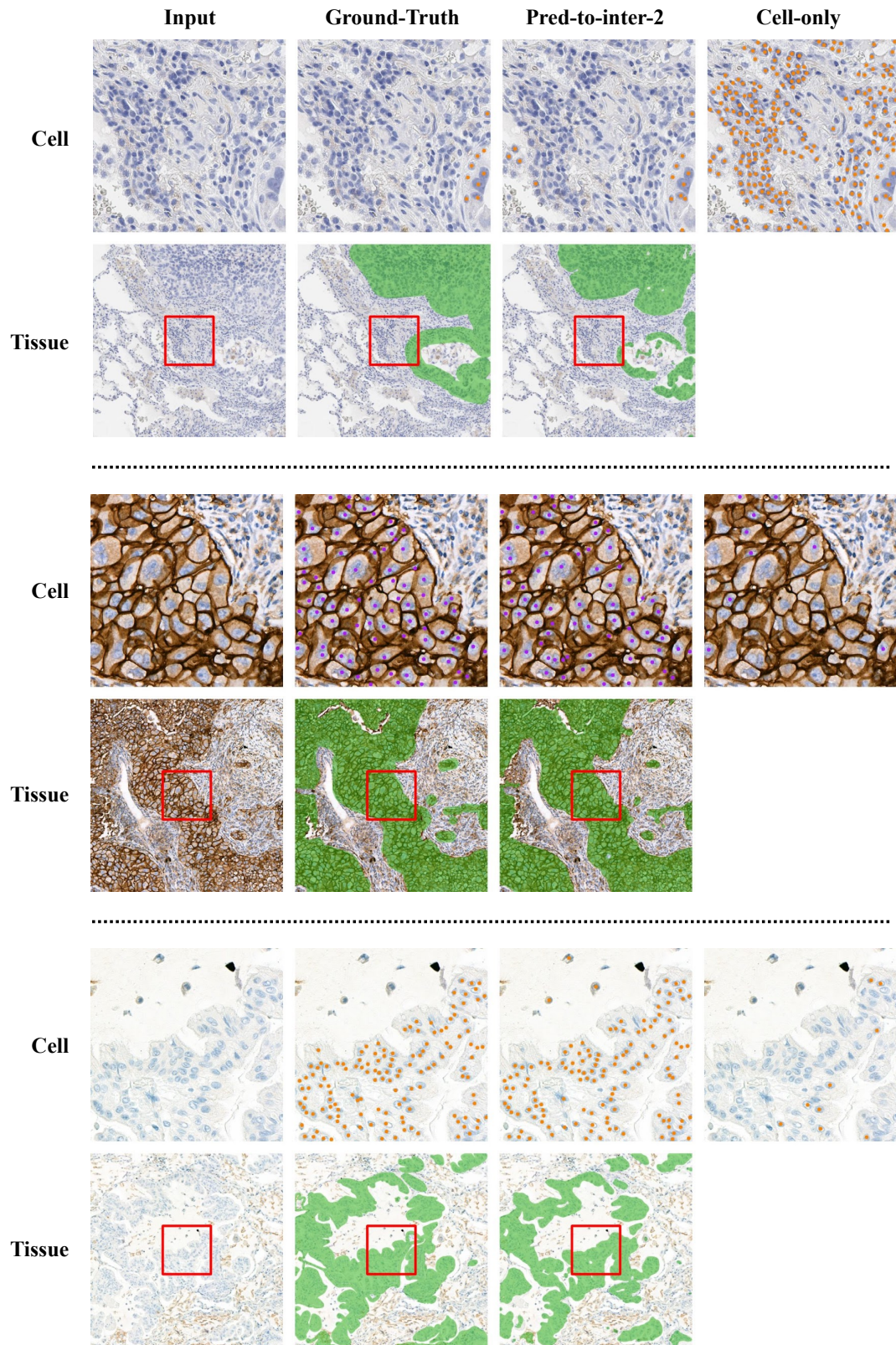


Figure E.2. **Qualitative results - CARP.** The *Pred-to-inter-2* model can correct the mistakes of the *Cell-only* model by incorporating tissue prediction information during cell prediction. The colors represent the following classes: ● PD-L1 positive tumor cells (TC+), ● PD-L1 negative tumor cells (TC-), and ■ Cancer Area (CA).