Immune-guided AI for Reproducible Regions of Interest Selection in Multiplex Immunofluorescence Pathology Imaging

Tanishq Gautam[1], Karina P. Gonzalez[2], Maria E. Salvatierra[2], Alejandra Serrano[2], Pingjun Chen[1], Xiaoxi Pan[1,2], Yasin Shokrollahi[1], Sara Ranjbar[1], Leticia Rodriguez[1], Patient Mosaic Team[3], Luisa Solis-Soto[2], Yinyin Yuan[1,2], and Simon P. Castillo[1,2][0000−0002−0606−2160]

[1] The Institute for Data Science in Oncology, The University of Texas MD Anderson Cancer Center TX 77098, USA spcastillo@mdanderson.org
[2] Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center TX 77098, USA spcastillo@mdanderson.org
[3] The University of Texas MD Anderson Cancer Center, TX 77098, USA

Abstract. Selecting regions of interest (ROIs) in whole-slide histology images (WSIs) is a crucial step for spatial molecular profiling. As a general practice, pathologists manually select ROIs within each WSI based on morphological tumor markers to guide spatial profiling, which can be inconsistent and subjective. To enhance reproducibility and avoid inter-pathologist variability, we introduce a novel immune-guided end-to-end pipeline to automate the ROI selection in multiplex immunofluorescence (mIF) WSIs stained with three cell markers (Syto13, CD45, PanCK). First, we estimate immune infiltration ($CD45^+$ expression) scores at the grid level in each WSI. Then, we incorporate the Pathology Language and Image Pre-Training (PLIP) foundational model to extract features from each grid and further select a subset of grids representative of the whole slide that comparatively matches pathologists' assessment. Further, we implement state-of-the-art detection models for ROI detection in each grid, incorporating learning from pathologists' ROI selection. Our study shows a significant correlation between our automated method and pathologists' ROI selection across five different types of carcinomas, as evidenced by a significant Spearman's correlation coefficient (> 0.785, p < 0.001), substantial inter-rater agreement (Cohen's $\kappa > 0.671$), and the ability to replicate the ROI selection made by independent pathologists with excellent average performance (0.968 precision and 0.991 mean average precision at a 0.5 intersection-over-union). By minimizing manual intervention, our solution provides a flexible framework that potentially adapts to various markers, thus enhancing the efficiency and accuracy of digital pathology analyses.

Keywords: ROI selection · Automated detection · Immune infiltration · Multiplex immunofluorescence.

1   Introduction

Selecting representative ROIs within whole-slide images (WSIs) is a routine task in pathology, serving foundational roles in various analytical procedures, including molecular scoring, comprehensive *in situ* molecular characterization, diagnostic assessments, and biomarker identification [14]. Traditionally, the process of ROI selection has chiefly relied on manual methods, which can be time-consuming, subjective, and might fail to represent the whole tissue [12]. In addition, the advent of multiplex immunofluorescence (mIF) imaging demands higher precise quantification and localization of numerous biomarkers to elucidate cellular dynamics and represent the whole-slide context. Therefore, there is a clear need for objective, reproducible, and scalable techniques for ROI selection.

The field has witnessed significant strides in integrating computational models to automate the ROI selection process. Li *et al.* [11] pioneered the ROI selection automation and formulated it as a segmentation task. They proposed a superpixel-based ROI search algorithm combining boundary update and coarse-to-fine refinement. With the fast progress of deep learning, particularly convolutional neural networks(CNNs) [2, 5, 19], several studies have adopted either classification-based[1, 16] or detection-based [15] models to address this challenge. These automated frameworks demonstrated a reduction in pathologists' manual labor and increased precision [10]. Recently, Hossain *et al.* [6] employed the vision transformer [3] for ROI selection in WSIs and achieved superior performance compared to earlier CNN-based methodologies. Despite these advancements, the field faces specific challenges, such as the predominant application of these technologies to Hematoxylin and Eosin (H&E) stained images, limiting their use with other staining methods. Additionally, the effectiveness of these computational models often hinges on the structure of the ROI being well-defined, which may not always be the case in complex pathological conditions.

In this research, we present an immune-guided automated system for selecting ROIs in mIF images to identify ROIs in the epithelial/neoplastic immune interface in carcinomas. From a pathological standpoint, quantifying immune infiltration is relevant for clinical prognosis [4, 18]; hence, developing an immune-guided approach contributes to the explainability and impact of an automated approach. Unlike previous models designed for applications in regions with clear structural definitions, our study embraces the inherently subjective nature of identifying ROIs in mIF images where no clear ground truth exists closely matching pathologist's selections in terms of consistency and coverage; this being a crucial step in molecular profiling, impacting the detection and quantification of biomarkers. While our current pipeline is optimized for mIF, it contains flexible components in pre-processing stages that can be adapted to different staining techniques. Furthermore, our approach offers a level of selection flexibility previously unattainable in manual processes. This innovative pipeline significantly enhances the explainability, reproducibility, scalability, and efficiency of digital pathology analyses by optimizing multiplex imaging through streamlined ROI selection by applying SOTA models within a rich dataset, and demonstrating the transformative potential of an automated approach for mIF imaging.

2   Methods

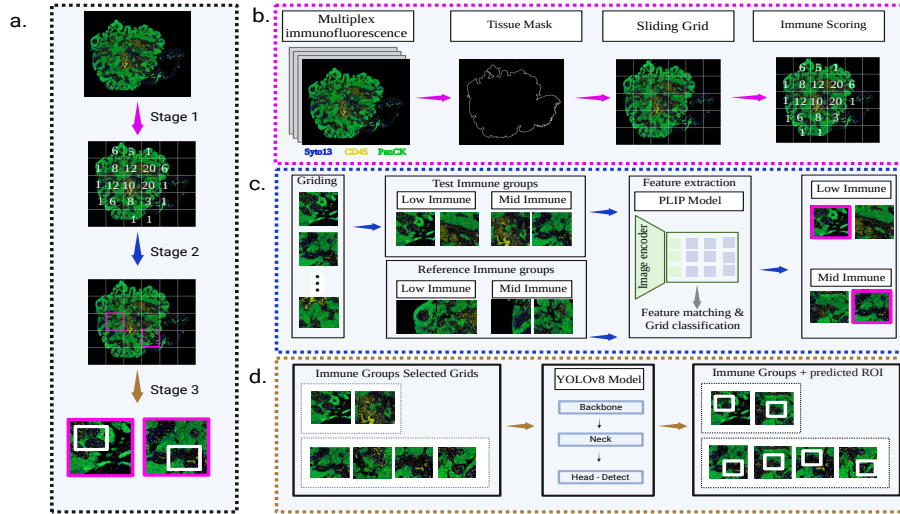## 2.1   Dataset Preparation and Pathological Annotation

Fig. 1. (**a**) Immune-guided AI ROI selection workflow. In stage 1 (**b.**), we generate non-overlapping grids from tissue masks and calculate the automated immune score. In stage 2 (**c.**), we use the PLIP model to extract features and select grids from low-immune (<10%) and mid-immune (10%-50%) groups based on their similarity with reference pathologist-scored grids. Finally, in stage 3 (**d.**), ROIs are generated on selected grids by applying a fine-tuned object detection YOLOv8 model.

We analyzed high-resolution (0.4 μm/pixel) mIF WSIs obtained from surgical resection samples from patients diagnosed with papillary urothelial carcinoma (PUC), penile squamous cell carcinoma (PSCC), urothelial carcinoma (UC), cholangiocarcinoma (CC), and rectal squamous cell carcinoma (RSCC). We performed a 5-fold cross-validation (80/20) for model training with 25 PUC images containing 300 ROIs selected by two pathologists. The remaining four tumor types, totaling 12 images and 144 ROIs, provide a diverse test set to evaluate the generalizability and robustness of our approach.

Our ROI-training dataset was curated using pathologists' annotations on mIF slides, guided by tumor-infiltrating lymphocytes (TILs) scoring on H&E [18] and immune-scored ROIs on mIF. Pathologists identified ROIs and assessed immune scores by quantifying the $CD45^+$ immune cells in the stromal area. This evaluation excluded non-cellular regions, tumor nests, and artifacts, with the immune score calculated as the area of $CD45^+$ cells relative to the total stromal area.

The evaluation dataset was compiled to focus on automated mIF immune scoring, an alternative to the H&E TILs scoring. This transfer learning approach underscores the depth of our analysis, ensuring that our evaluation is flexible and grounded in distinct yet complementary immunopathological assessments. Data use follows MD Anderson's Institutional Review Board (IRB) protocol (#2022-0404).

2.2   Pipeline Overview

Our methodology has three key stages (Fig. 1). First, we preprocessed WSI to quantify immune infiltration using CD45 expression, setting a foundation for ROI selection. Next, we refined this selection by aligning with pathologist-scored grids via feature similarity optimization, utilizing the PLIP model [7] for feature extraction. Finally, we employed the YOLOv8 object detection model, fine-tuned with pathologist annotations, for precise ROI detection in immune-selected grids.

2.3   Stage 1 - Automated Immune Scoring

For computing an accurate automated immune scoring (aIS), based on their distinct color ranges, we identified and excluded auto-fluorescent regions, such as those produced by hemorrhages and blood vessels, which could confound the analysis. Then, for clear tissue segmentation, we utilized hole-filling techniques to construct a comprehensive contour mask, allowing us to distinguish between glass and tissue regions within the tissue sample since the latter may not express any of the markers considered here (Syto13, CD45, or panCK). In our case, the stroma was defined as the region within the tumor devoid of epithelial cells, artifacts, or necrosis from where we calculate aIS.

The subsequent phase involved generating 3000×3000-pixel non-overlapping grids throughout the WSI. Within each selected grid (grid with $< 50\%$ black pixel), aIS is determined by classifying pixels within the stroma as immune (yellow, $CD45^+$), excluding those classified as epithelial (green, $PanCK^+$). We verified that the association of color and marker expression was consistent across our dataset.

$$aIS = \frac{CD45^+ Area}{GridArea - BackgroundArea - panCK^+ Area} \qquad (1)$$

In each case, the area was calculated by selecting specific color ranges to accurately capture various intensities of the indicated colors, considering that both strong and weak expressions of yellow represent $CD45^+$ immune infiltration. To evaluate the suitability of this approach for quantifying immune infiltration, we designed an evaluation step that included manual immune scoring (mIS) by two independent pathologists blinded to the aIS. We evaluate the association between aIS and mIS with Spearman's correlation for continuous values and Cohen's κ score for inter-rater agreement on categorized scores orthogonally comparing aIS, mIS-1, and mIS-2.

2.4   Stage 2 - Group Classification and Similarity Matching

Upon calculating aIS for each grid, we classified them into two immune classes where aIS $< 10\%$ (low immune) or aIS between 10% and 50% (mid immune). Our tissue samples did not have grids with aIS $> 50\%$, and since the grids correspond to tumor regions, we do not expect grids with a predominance of immune cells. These criteria mirror a repository of pathologist-annotated grids consisting of grids (3000 × 3000 pixels) previously categorized into the same groups where no grids were scored mIS $> 50\%$. To define the number of grids per immune class per slide, we compute the global ratio for each immune class over the total number of grids, keeping the same ratio for the selected grids.

After defining the number of grids containing the ROIs, we employ the PLIP model to perform feature extraction and select the *best* grids. Selecting similar regions based on feature similarity from a pre-trained model is standard practice for image retrieval [20]. We applied feature extraction on each grid with aIS and on the reference grids with mIS in each immune class. For each grid in a sample, we calculated the feature similarity (cosine similarity) with the reference grids. The selected grids are the ones that maximize the feature similarity with the reference set. We confirmed PLIP's effectiveness using dispersion along the principal components in PCA, showing it captures key variances in H&E images that apply to mIF images.

2.5   Stage 3 - Object Detection Training and ROI Selection

The grids selected from the previous stage that matched features with the reference set were subjected to analysis by a finely tuned object detector, primarily to predict potential new ROIs of size 900×900 pixels (corresponding to the average size of pathologist-annotated ROIs).

To provide a robust evaluation of our framework, we evaluated the performance of 4 detector models for the task of ROI selection. We assessed the performance of RetinaNet [13], Faster R-CNN [17], YOLOv5 [8], and YOLOv8 [9].

For ROI prediction, the selected low and mid-immune grids from stage 2 were fed as input into the detector models, generating ROIs with the highest confidence scores. To evaluate each detector, we calculated the overlap of the prediction with manually selected ROIs by two independent pathologists who were blinded to the model's predictions. We quantified the mean average precision (mAP) and the intersection over union (IOU) scores between predictions and pathologists' ROI selection. This approach ensures that our study identifies ROIs with a high degree of accuracy while quantifying the confidence of these predictions.

3   Results

3.1   Validation of Automated Immune Scoring

The automated immune score (aIS) is shown to be apt for inferring immune infiltration, as shown by having a significant association with the manual scores
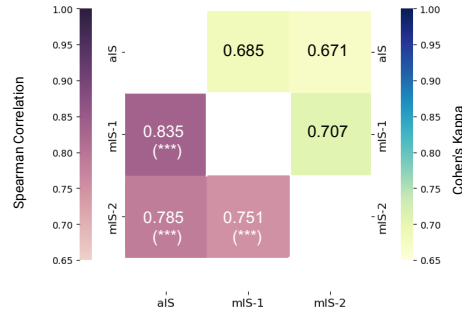
Fig. 2. Correlation between automated (aIS) and manual (mIS) immune scores. Two independent pathologists scored N=125 grids. The lower triangular shows Spearman's Correlation and P-values. The upper triangular shows the inter-rater agreement, which is measured with Cohen's κ score on immune classes. P-values for Spearman's correlation *p<0.05, **p<0.01, ***p<0.001.

(mIS) from two independent pathologists (Fig 2). There is a strong and significant correlation between AI and the pathologists' scores (Fig 2a, Spearman's correlation aIS vs mIS-1 = 0.835, p-value=9.310×10^{-34}; AI vs mIS-2 = 0.785, p-value=2.116×10^{-27}), suggesting the suitability of the aIS to estimate mIF immune infiltration.

The relevance of the automated score is further supported by the high inter-rater agreement analysis between aIS and mIS (Fig 2b). Cohen's κ provides insights into the agreement between raters focusing on categorized scores ($< 10\%$ or 10%-50%,). The observed values of 0.685 (aIS *vs* mIS-1) and 0.671 (aIS *vs* mIS-1) indicate a substantial agreement having as a reference that the concordance between pathologists (mIS-1 *vs* mIS-2) is $\kappa = 0.707$ and a theoretical value of $\kappa = 1$ reflects a complete agreement between the raters.

3.2   Assessment of Feature-based Grid Similarity

To select the grids from a WSI candidate for containing ROIs, we applied feature extraction of grids and, stratified by immune class, selected those with the highest similarity to the pathologist's reference set. Implementing a foundational PLIP model pre-trained on H&E images facilitates the extraction of high-dimensional feature embeddings, providing a granular analysis of the cellular and morphological attributes discernible within the tissue specimens.

Grid selection prioritized those with higher cosine similarity to pathologist-annotated reference grids (Fig. 3). Both Low-Immune and Mid-Immune selected grids (*best* grids) demonstrated higher median similarity scores compared to Not-Selected grids. This trend highlights the effectiveness of our selection criteria in capturing expert evaluations, particularly in selecting grids that pathologists consider more indicative of the immune context within the tissue.
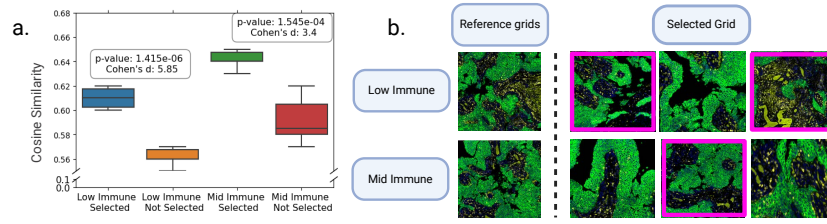
Fig. 3. (**a**) Quantitative comparison of the selected and not-selected grids cosine similarity with the reference set for low immune (<10%) and mid immune (10%-50%) scores. Boxplots show median (thick line), interquartile range (box limits), and minimum-maximum values (whiskers). (**b**) Examples of selected grids and the reference for both immune classes. The purple boxes indicate the grids that are most similar to the reference grids based on feature embeddings to be selected based on a 2:1, low:medium immune ratio.

3.3   Evaluation of Predicted ROI Selection

In our comparative evaluation for object-detection models trained and validated on tumor-infiltrating lymphocytes (TIL)-based ROI (Table 1), YOLOv8 outperformed competitors models having the highest mAP scores (mAP@50=0.903 $\pm$ 0.005, mAP@50-95= 0.713 $\pm$ 0.007). Based on the superiority in accuracy, particularly over varied IOU thresholds, we select YOLOv8 as the last layer of our framework and as the baseline model for further fine-tuning. A YOLOv8-S model with batch size 64, image size 1280 pixels, learning rate 0.01, momentum 0.937, and weight decay 0.0005 was trained for 1000 epochs with a patience of 200. At 20x magnification, the final scores were: box loss 0.523, class loss 0.399, and distribution focus loss 1.001. Validation scores were: box loss 1.686, class loss 0.985, and distribution focus loss 1.51.
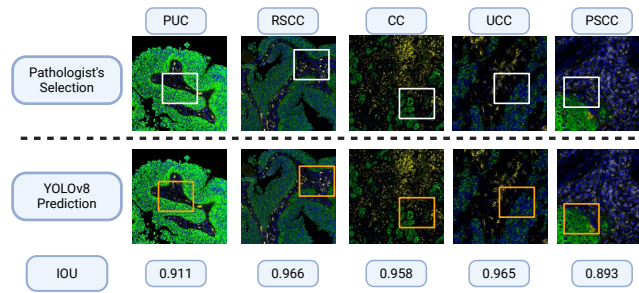

Fig. 4. Examples of immune-guided ROI detection and comparison with manual pathologists selection. The intersection-over-union (IOU) score between pathologists and automated selected ROIs is shown for each example. PUC: Papillary Urothelial Carcinoma, PSCC: Penile Squamous Cell Carcinoma, UC: Urothelial Carcinoma, CC: Cholangiocarcinoma, and RSCC: Rectal Squamous Cell Carcinoma.

Table 1. Performance comparison of state-of-the-art detection models on ROI selection trained with TIL-based annotations. The mean average precision (mAP) is calculated at 50 and 50-95 IOU. Mean mAP ± SD was calculated across five distinct cancer types.

Metric	RetinaNet	Faster R-CNN	YOLOv5	YOLOv8
mAP@50	0.745 ± 0.101	0.801 ± 0.009	0.848 ± 0.007	**0.903 ± 0.005**
mAP@50-95	0.504 ± 0.111	0.601 ± 0.100	0.655 ± 0.008	**0.713 ± 0.007**

To evaluate the performance of immune-based ROI detection, we performed 5-fold cross-validation with pathologists-selected ROIs. The analysis showcases the efficacy of transfer learning methodologies, where models pre-trained on TIL-scored ROIs are fine-tuned to evaluate immune-scored ROIs (Table 2). These results also demonstrate the effectiveness of fine-tuning to enhance the model's sensitivity and specificity across a diverse set of tumor types.

Table 2. Performance comparison of baseline and fine-tuned YOLOv8 models evaluated across different tumor types. PUC: papillary urothelial carcinoma, PSCC: penile squamous cell carcinoma, UC: urothelial carcinoma, CC: cholangiocarcinoma, and RSCC: rectal squamous Cell Carcinoma. The mean average precision (mAP) is calculated at 50 and 50-95 IOU. Mean mAP ± SD calculated across folds (PUC) and across tumor types for PSCC/RSCC/UCC/CC.

	PUC		PSCC/RSCC/UCC/CC	
Metric	Baseline	Fine-tuned	Baseline	Fine-tuned
Precision	0.833 ± 0.035	0.971 ± 0.043	0.825 ± 0.030	0.965 ± 0.041
Recall	0.885 ± 0.014	0.993 ± 0.006	0.891 ± 0.021	0.980 ± 0.003
mAP@50	0.901 ± 0.110	0.991 ± 0.005	0.888 ± 0.104	0.992 ± 0.004
mAP@50-95	0.888 ± 0.107	0.897 ± 0.009	0.901 ± 0.005	0.911 ± 0.002

The resulting ROI selection with a fine-tuned YOLOv8s model as the last layer of our approach shows considerable pathologist-validated generalizability across different tumor types (Table 2, Fig 4). These results reflect a high degree of congruity between the YOLOv8 predictions and the pathologist's annotations, implying a robust model performance across different tumor morphologies and reliability in complex and varied pathological scenarios.

4   Conclusion

We propose a comprehensive framework for immune-guided ROI selection on whole-slide mIF images, which would impact research and overcome manual processing due to the different challenges such an approach may entail. Our approach showcases the reproducibility and efficiency of automated ROI selection without sacrificing explainability, preserving biologically relevant features, such as immune infiltration, and with a good overlap with independent pathologists'

evaluation. Beyond automation, this approach can improve the reliability of downstream molecular analyses.

Our study advances automated ROI selection but faces limitations that guide our future directions. The primary constraint is the dataset size, which restricts the current approach to a limited number of markers. Additionally, the foundational model, trained exclusively on H&E images, may not seamlessly apply to broader pathology practices. Future efforts will aim to enlarge the dataset, incorporate artifact correction techniques, broaden model training across imaging methods, and integrate spatial analysis into ROI selection. Although limited by the minimal set of markers and tumor types availability, this research underscores the significant impact such automated systems could have on digital pathology, ultimately showcasing the co-evolution between pathologists' needs and AI developments.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, M., Zhang, B., Topatana, W., Cao, J., Zhu, H., Juengpanich, S., Mao, Q., Yu, H., Cai, X.: Classification and mutation prediction based on histopathology HE images in liver cancer using deep learning. npj Precision Oncology **4**(1), 14 (6 2020)
2. Chollet, F.: Xception: Deep Learning with Depthwise Separable Convolutions (10 2016)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (10 2020)
4. Gooden, M.J., de Bock, G.H., Leffers, N., Daemen, T., Nijman, H.W.: The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. British journal of cancer **105**(1), 93–103 (2011)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE (6 2016). https://doi.org/10.1109/CVPR.2016.90
6. Hossain, M.S., Shahriar, G.M., Syeed, M.M., Uddin, M.F., Hasan, M., Shivam, S., Advani, S.: Region of interest (ROI) selection using vision transformer for automatic analysis using whole slide images. Scientific Reports **13**(1) (12 2023). https://doi.org/10.1038/s41598-023-38109-6

7. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., Zou, J.: A visual-language foundation model for pathology image analysis using medical Twitter. Nature medicine **29**(9), 2307–2316 (9 2023). https://doi.org/10.1038/s41591-023-02504-3
8. Jocher, G.: Ultralytics yolov5 (2020). https://doi.org/10.5281/zenodo.3908559, https://github.com/ultralytics/yolov5
9. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics yolov8 (2023), https://github.com/ultralytics/ultralytics
10. Korkut, S., Erkan, C., Aksoy, S.: On the benefits of region of interest detection for whole slide image classification. In: Tomaszewski, J.E., Ward, A.D. (eds.) Medical Imaging 2023: Digital and Computational Pathology. p. 23. SPIE (4 2023). https://doi.org/10.1117/12.2654193
11. Li, R., Huang, J.: Fast Regions-of-Interest Detection in Whole Slide Histopathology Images. pp. 120–127 (2015). https://doi.org/10.1007/978-3-319-28194-0{_}15
12. Lin, J.R., Wang, S., Coy, S., Chen, Y.A., Yapp, C., Tyler, M., Nariya, M.K., Heiser, C.N., Lau, K.S., Santagata, S., et al.: Multiplexed 3d atlas of state transitions and immune interaction in colorectal cancer. Cell **186**(2), 363–381 (2023)
13. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal Loss for Dense Object Detection (8 2017)
14. Nagarkar, D.B., Mercan, E., Weaver, D.L., Brunyé, T.T., Carney, P.A., Rendi, M.H., Beck, A.H., Frederick, P.D., Shapiro, L.G., Elmore, J.G.: Region of interest identification and diagnostic agreement in breast pathology (9 2016). https://doi.org/10.1038/modpathol.2016.85
15. Nugaliyadde, A., Wong, K.W., Parry, J., Sohel, F., Laga, H., Somaratne, U.V., Yeomans, C., Foster, O.: RCNN for Region of Interest Detection in Whole Slide Images. pp. 625–632 (2020). https://doi.org/10.1007/978-3-030-63823-8{_}71
16. Qaiser, T., Lee, C.Y., Vandenberghe, M., Yeh, J., Gavrielides, M.A., Hipp, J., Scott, M., Reischl, J.: Usability of deep learning and H&E images predict disease outcome-emerging tool to optimize clinical trials. npj Precision Oncology **6**(1), 37 (6 2022). https://doi.org/10.1038/s41698-022-00275-7
17. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(6), 1137–1149 (6 2017). https://doi.org/10.1109/TPAMI.2016.2577031
18. Salgado, R., Denkert, C., Demaria, S., Sirtaine, N., Klauschen, F., Pruneri, G., Wienert, S., Van den Eynden, G., Baehner, F.L., Pénault-Llorca, F., et al.: The evaluation of tumor-infiltrating lymphocytes (tils) in breast cancer: recommendations by an international tils working group 2014. Annals of oncology **26**(2), 259–271 (2015)
19. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition (9 2014)
20. Yu, J., Wu, Z., Mayer, A., Trevino, A., Zou, J.: A multi-granularity approach to similarity search in multiplexed immunofluorescence images (05 2024)