Deep Multitask Learning for Automated Tissue Segmentation and Cell Detection

1st Mohamed Mounib Benimam
BIA Unit, Institut Pasteur
Paris, France
ORCID: 0009-0001-5185-495X

2nd Astri Frafjord
Department of Pathology
Oslo University Hospital
Oslo, Norway
ORCID: 0000-0001-7942-1411

3rd Alexandre Corthay
Department of Pathology
Oslo University Hospital
Oslo, Norway
ORCID: 0000-0001-9519-5725

4th Thibault Lagache
BIA Unit, Institut Pasteur
Paris, France
Email: thibault.lagache@pasteur.fr

5th Jean-Christophe Olivo-Marin
BIA Unit, Institut Pasteur
Paris, France
Email: jcolivo@pasteur.fr

6th Vannary Meas-Yedid
BIA Unit, Institut Pasteur
Paris, France
Email: vmeasyed@pasteur.fr

***Abstract*—We propose a unified multitasking deep learning method to identify and segment structural regions and detect immune cells of non-small cell lung cancer tissue. This approach combines multitask learning and deep supervision to significantly reduce training time compared with single-task models, while achieving competitive performance with 50% fewer parameters. We balance tasks using homoscedastic uncertainty and address class imbalance through a combination of focal loss and class weight adjustment.**

Tested on multiplexed WSI, our method overcomes challenges related to variability and limited annotations, while extracting robust spatial information from the tumor microenvironment (TME). This data can be leveraged to quantify tumor–immune dynamics and intercellular communication, thereby providing objective insights into the TME's structural and functional complexity.

***Index Terms*—Multitask learning, Multi-label learning, deep supervision, segmentation, cell detection, heatmap regression, imbalanced data, computational pathology, cancer**

I. Introduction

The tumor microenvironment (TME) plays a crucial role in influencing cancer cell responses to treatment. A thorough understanding of the TME and its modifications benefits from a quantitative, reproducible, and comprehensive analysis of its constituent components. Multiplex images provide rich, high-content information and can be effectively analyzed using efficient, automated processing tools. Analyzing these multiplexed whole slide images (WSI) presents the opportunity to address the significant intra-class variance in the size, color, and texture of objects of interest (see Fig.1 and Fig.2) through deep learning (DL) approaches. In the framework of machine learning, additional challenges such as dataset imbalance (see Tables I and II) and limited annotations (Fig. 3) can be met with robust solutions.

This work was supported by the program PIA INCEPTION (ANR-16-CONV-0005). This project has received funding from by grants from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 945358. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA. www.imi.europe.eu

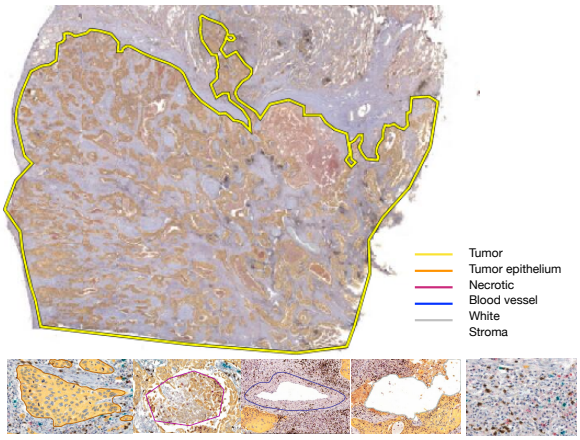

Fig. 1. Image of a whole slide image of NSCLC (size: 95310 x 205272 at the highest resolution of 0.243 $\mu m/\text{px}$) and examples of 5 types of regions to segment: tumor epithelium, necrotic, blood vessel, empty, and stroma.

Traditional image analysis techniques, while essential, often require considerable, image-dependent parameter adjustments, which can limit scalability and introduce subjectivity. They also tend to employ different algorithms for each task—using one algorithm for region segmentation and another for immune cell identification. This separation can result in fragmented workflows that hinder the coordination between multi-scale features such as regional morphology and cellular spatial distributions. Additionally, the need for expert knowledge in each task can make it challenging to translate these methodologies to other parts of the TME. These challenges become even more pronounced in large-scale studies, particularly in the presence of staining variability.

Alternatively, unifying multiple tasks under a single model offers clear benefits. By aggregating tasks such as segmenta-

Fig. 2. Examples of 3 immune cell types : T cell (brown), mast cell (cyan), Nk cell (red) and their respective false detection

tion and immune cell detection, the model can share valuable information between them—where segmentation of the tumor area can aid in detecting immune cells, and vice versa. This combined approach not only streamlines the analysis process but also enhances the model's flexibility in adapting to the variability and complexity of the TME, ultimately yielding a more robust and generalizable tool.

In this work, we propose a unified method for extracting structures of interest in the TME of non-small cell lung cancer (NSCLC)—including TME regions and immune cell locations—through a multitask approach that simultaneously performs region segmentation and immune cell detection.

II. Dataset

Our dataset comprises 10 multiplex WSI that include 5 distinct regions (tumor epithelium, empty, necrotic, stroma regions, and blood vessels; see Fig. 1) and 3 immune cell types (T cells, mast cells, and NK cells; see Fig. 2).

Generation of annotations

Given the novel staining protocol [1], we have developed our own training dataset. To streamline the training data generation process, we use Icy, an user-friendly platform [2] that integrates computer assisted tools. Manual annotation for supervised deep learning methods can be labor-intensive, so we employ an semi automatic approach under human expert oversight to segment regions and detect immune cells. This segmentation method relies on color quantization [3] using a split and merge color algorithm, with the generated masks polygonalized. For cell detection, the staining image is decomposed according to the Vahadane method [4], and nuclei detection is performed via wavelet transform [5] on the nucleus staining image (blue staining). These approaches, while not requiring a training step, require human expert supervision to effectively manage the large intra-class variation present in the dataset. To promote reproducibility, the dataset has been made publicly available in zenodo [6]

TABLE I
Class distribution of annotated regions in training data

Region	TER	Necrotic	Blood vessels	White	Stroma
Percentage	78.8%	6.9%	2.53%	7.9%	3.3%

TABLE II
Class distribution of immune cell annotations

Cell Type	T cells	Mast cells	Nk cells
Percentage	96.78%	2.02%	1.1%

III. Related Works

Deep learning has revolutionized biomedical image analysis, particularly for detection and segmentation tasks. However, opportunities remain to advance multi-objective architectures that handle heterogeneous annotations. U-Net [7] and Mask R-CNN [8] have proven highly effective in semantic and instance segmentation, yet they are primarily designed for single-task objectives under fully annotated data. SOLO [9] advances instance segmentation with location-aware embeddings, though it relies on accurate boundary annotations, which can restrict its application in weakly labeled datasets. These models face challenges in scenarios where tasks differ in annotation granularity levels—such as combining pixel-level region segmentation with weakly supervised cell detection using sparse point annotations.

Deep supervision methods, which enhance intermediate layer training with auxiliary losses, have accelerated convergence in applications like cell density estimation [10]. While these methods effectively address single-task optimization, they do not provide mechanisms to balance competing objectives in multi-task settings. Multi-task learning (MTL) addresses this by optimizing shared representations jointly. Kendall et al.'s [11] uncertainty-aware architecture scales losses homoscedastically, offering a principled mechanism to balance heterogeneous tasks with varying complexities. This approach remains underexplored in histopathology.

Recent MTL models, such as Cerberus [12], demonstrate performance gains in both segmentation and regression when combining multiple objectives compared to single-objective approaches. However, the Cerberus model is stain-specific, which can limit its flexibility for applications involving IHC stains.

IV. Methodology

Our framework successfully addresses both cell detection and region segmentation through a unified approach.

A. Pre-processing

To enhance model robustness against staining variations and rotations, we apply the following data augmentations during training:

- Geometric transformations: random crops (1024px), mirroring (50% chance), 90° rotations (50% chance)
- Appearance variations: color jitter, elastic deformations
- Normalization: zero-mean, unit-variance scaling (applied consistently during inference)

B. Cell Heatmap Generation

We convert sparse cell coordinates into a probability heatmap using the following three steps:

1) Place a Gaussian distribution with variance σ on each dot (x, y coordinates).
2) Apply an element-wise maximum operation to preserve distinct cell peaks.
3) Normalize the values to the range [0, 1].

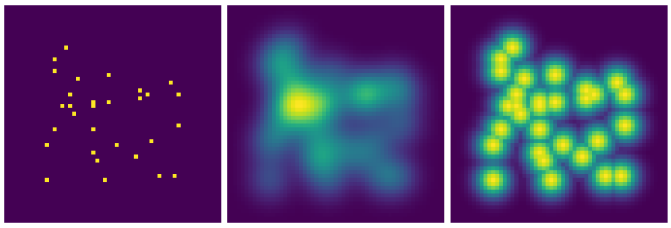

Fig. 3. Cell localization visualization: (A) Original keypoints, (B) Density map obtained using Gaussian kernel convolution, (C) Heatmap of maximum Gaussian responses.

The Gaussian spread is controlled by:

$$\sigma = \frac{\text{desired\_diameter}}{2\sqrt{2\ln 2}} \tag{1}$$

where σ controls the extent of cell influence in the density map. Each location (x, y) in the final map represents the maximum Gaussian value from nearby cells:

$$I(x,y) = \max_i \exp\left(-\frac{(x-x_i)^2 + (y-y_i)^2}{2\sigma^2}\right) \tag{2}$$

where $I(x, y)$ is the resulting heatmap at position (x, y), representing the maximum influence of all Gaussian centers at that location; x_i, y_i are the coordinates of the i-th Gaussian cell; σ is the standard deviation of the Gaussian, controlling its spread; $e^{-\frac{(x-x_i)^2+(y-y_i)^2}{2\sigma^2}}$ defines the influence of the i-th Gaussian cell at (x, y); and $\max_i$ selects the maximum contribution among all Gaussian cells for each point.

C. Multi-scale Supervision

Following [10], we generate auxiliary supervision signals by downsampling ground truth maps with scaling factors of 2, 4, and 8. We do so by applying max pooling instead of sum pooling in order to preserve the peaks for the downsampled maps as well. This multi-scale approach enhances detection accuracy across different resolutions.

D. Cell Detection via Density Regression

Building on density estimation methods [13], [14], we formulate cell detection as a regression task that predicts cell heatmaps $Y \in \mathbb{R}^{M\times N}$ from input images X. The regression loss integrates mean squared error with class-specific weights w_c derived from cell frequency:

$$\mathcal{L}_{\text{reg}} = \sum_c w_c \cdot \|y_c - \hat{y}_c\|^2 \tag{3}$$

where y_c and $\hat{y}_c$ denote the ground truth and predicted cell heatmaps for class c.

E. Region Segmentation loss

Weighted Focal Loss: Emphasizes learning on challenging examples while balancing classes:

$$\mathcal{L}_{\text{Focal}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} w_j y_{ij}(1-p_{ij})^{\gamma}\log(p_{ij}) \tag{4}$$

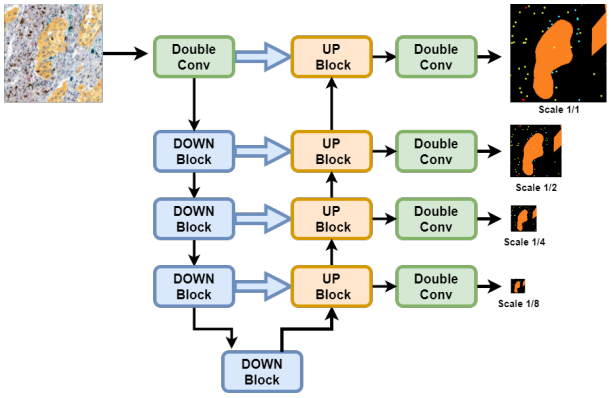

Fig. 4. Multi-task architecture with shared encoder and task-specific decoders.

where $\mathcal{L}_{\text{Focal}}$ is the focal loss value; N is the total number of samples in the dataset; C is the number of classes; $y_{ij} \in \{0, 1\}$ indicates whether class j is the true classification for sample i; p_{ij} represents the predicted probability that sample i belongs to class j; w_j is the class weight assigned to address data imbalance; and γ is the focusing parameter that modulates the loss contribution of well-classified examples.

F. Unified Model

The proposed unified model jointly performs cell detection and region segmentation through a shared architecture (Fig. 4). Building on the U-Net framework [7], we integrate auxiliary supervision branches inspired by deep supervision principles [10]. This design enables weight sharing between tasks while addressing gradient vanishing through multi-scale feature learning. The architecture combines a primary segmentation branch with cell heatmaps regression outputs at multiple resolutions.

1) Deep Supervision: Following [10], we generate downsampled ground truth maps using sum pooling (factors 2, 4, 8) for auxiliary supervision. The total loss combines multi-level predictions through uncertainty-weighted task balancing:

$$\mathcal{L}_{HL} = \frac{1}{2\sigma_1^2}\mathcal{L}_{reg} + \frac{1}{\sigma_2^2}\mathcal{L}_{seg} + \log\sigma_1\sigma_2, \tag{5}$$

where $\mathcal{L}_{reg}$ and $\mathcal{L}_{seg}$ denote regression and segmentation losses, with σ_1, σ_2 as learnable uncertainty parameters. Deep supervision aggregates scaled losses across resolution levels:

$$\mathcal{L}_{DS} = \sum_{i=0}^{3}\frac{1}{2^i}\mathcal{L}_{HL}^i. \tag{6}$$

G. Implementation Details

The model is implemented in PyTorch Lightning. Training employs the Adam optimizer ($\beta = 0.9$, initial learning rate 10^{-4}) for 400 epochs with cosine annealing, following parameters from [15]. Batches of four patches (1024x1024px) are processed through 3x3 convolutions.

H. Post Processing

To enhance multi-scale consistency, we multiply probability maps element-wise from deep supervision auxiliary branches at various resolution levels. This **MF** (multi-scale fusion) leverages complementary contextual information by reinforcing features consistent across scales.

The combined probability map is passed through softmax normalization to ensure valid probabilities and argmax classification to generate the final segmentation. A 3×3 mode filter is subsequently applied to smooth region boundaries while preserving anatomical edges through majority voting of neighboring labels. In cell detection tasks, non-maximum suppression is applied to identify local maxima within predicted cell heatmaps for accurate cellular centroid localization while eliminating duplicate detections.

I. Evaluation Metrics

Region segmentation performance is measured using class-wise F1-score. Cell detection accuracy is assessed by matching predictions to ground truth centroids within an 8px tolerance: true positives (TP) represent matched pairs, false positives (FP) represent unmatched predictions, and false negatives (FN) represent unmatched ground truths. Precision ($\frac{TP}{TP+FP}$), recall ($\frac{TP}{TP+FN}$), and average precision are calculated from these counts.

V. Discussion

The results validate the efficacy of the proposed method in improving precision-recall performance and F1 scores through multiscale fusion.

a) Precision-Recall Analysis: Average precision (AP) values indicate strong performance for all three biomarkers: Nkp46, Tryptase, and CD3. The AP for CD3 was highest at 0.88, followed by Tryptase at 0.86 and Nkp46 at 0.76. These results demonstrate that the model effectively balances precision and recall, particularly for CD3, which exhibits the highest discriminative power among the tested biomarkers. This finding suggests that the model handles class imbalances effectively and reduces false positives for critical biomarkers.

b) Impact of Multiscale Fusion: The F1 scores reveal that multiscale fusion consistently enhances model performance across most region categories. In the tumor epithelial region (TER), the F1 score improved from 0.907 to 0.908, underscoring the robustness of the method in maintaining high performance across categories. A significant improvement was observed in the blood vessel category, where the F1 score increased from no result to 0.656, highlighting the ability of multiscale fusion to handle challenging regions that were previously undetectable.

Moreover, stroma and white regions also exhibited slight performance gains, with F1 scores increasing from 0.791 to 0.793 and 0.969 to 0.971, respectively. These results demonstrate the success of the multiscale approach in refining predictions for diverse region types, even for those with already strong baseline performance.

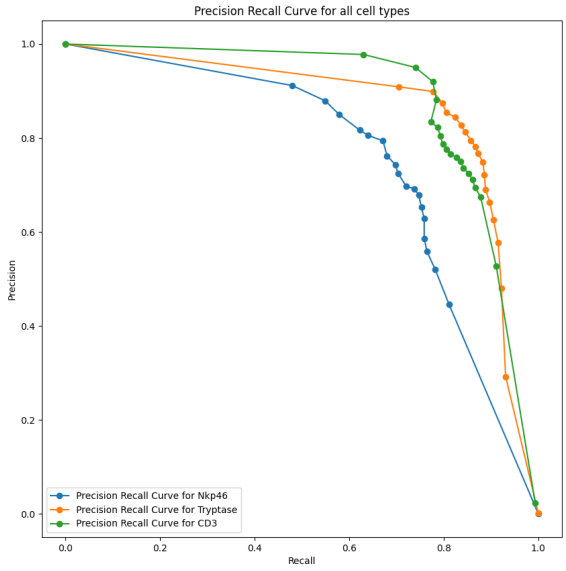

Fig. 5. Cell detection: precision recall curves for CD3, Tryptase, Nkp46

c) Comparison to Single Task Learning: It is important to note that the Single Task Learning (STL) model sometimes outperforms the proposed multi-task models, particularly in categories like **TER** and **Necrotic**. However, the proposed multi-task method (MTL + MF) delivers competitive performance with **50% fewer parameters** than training two separate models (7 million parameters comparred to 14 millions for the case of 2 separate models). This highlights the efficiency of the multi-task approach in terms of computational resources, while maintaining comparable or even improved performance in certain regions.

d) Overall Significance: The incorporation of multiscale fusion proves particularly advantageous for addressing segmentation challenges in regions with lower baseline performance. The results indicate the versatility of the technique in improving F1 score. These findings suggest that multiscale fusion is a valuable enhancement to segmentation pipelines, offering complementary gains that enhance the model's applicability in various contexts. Furthermore, the reduced parameter requirement in the multi-task approach provides an added benefit, making it a more resource-efficient solution for large-scale deployment.

e) Benchmarking: The proposed multitask framework demonstrates competitive performance for tissue segmentation and cell detection in histopathology images. To strengthen the validity and relevence of our model, we intend to extend this work by conducting a more comprehensive benchmarking against recent state of the art multitask learning approaches.

VI. Conclusion

We introduced a unified model that leverages a multi-task framework for segmenting regions of interest and detecting

TABLE III
F1 SCORES FOR SINGLE TASK, MULTI-TASK WITHOUT MULTISCALE FUSION, AND MULTI-TASK WITH MULTISCALE FUSION

Model Variant	TER	Necrotic	Blood Vessels	White	Stroma
STL (U-net)	**0.931**	**0.927**	/	0.782	0.960
MTL	0.907	**0.9142**	/	0.791	0.969
MTL + MF	**0.908**	0.894	**0.656**	**0.793**	**0.971**

TABLE IV
AREA UNDER THE CURVE (AUC) FOR PRECISION-RECALL CURVES

AUC (Average Precision)	Nkp46	Tryptase	CD3
Single Task	**0.78**	**0.90**	**0.93**
Multi Task	0.76	0.86	0.88

immune cells in WSI. By incorporating deep supervision, our approach accelerates training compared to training separate models individually. To maintain a balance across the multiple tasks, we employ homoscedastic uncertainty, ensuring task equilibrium. To address class imbalance, we apply focal loss for segmentation and weighted mean squared error loss for cell detection. The results demonstrate competitive performance in segmentation tasks, with notable improvement in the immune cell detection task. Although Single Task Learning (STL) outperforms in some segmentation categories, the multi-task model achieves similar performance with **50% fewer parameters**, offering a more efficient solution without compromising quality. The extraction of spatial structures facilitates spatial statistical analysis [16].

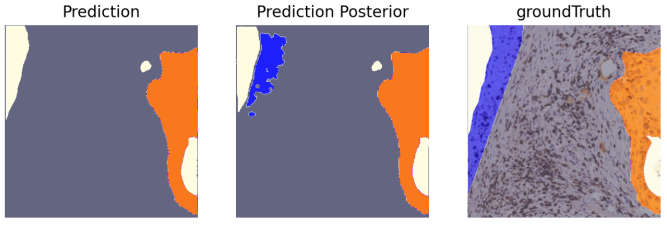

Fig. 6. Qualitative results comparison between with and without Multiscale fusion

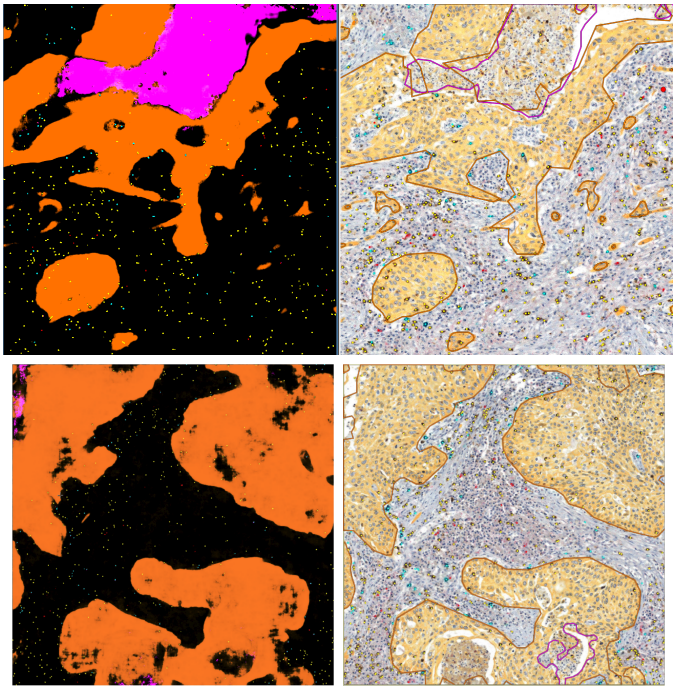

Fig. 7. Qualitative results raw output vs overlayed post processed output

REFERENCES

[1] A. Frafjord, L. Buer, C. Hammarström, H. Aamodt, P. R. Woldbæk, O. T. Brustugun, Åslaug Helland, I. Øynebråten, and A. Corthay, "The immune landscape of human primary lung tumors is th2 skewed," *Frontiers in immunology*, vol. 12, 11.

[2] F. de Chaumont, S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. L. Montagner, T. Lagache, A. Dufour, and J. C. Olivo-Marin, "Icy: an open bioimage informatics platform for extended reproducible research," *Nature Methods 2012 9:7*, vol. 9, pp. 690–696, 6.

[3] S. Mukherjee, V. Meas-Yedid, M. Bokobza, T. Lagache, A. Corthay, and J.-C. Olivo-Marin, "Spatial analysis for histopathology: A statistical approach," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2022.

[4] A. Vahadane, T. Peng, S. Albarqouni, M. Baust, K. Steiger, A. Schlitter, A. Sethi, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Transactions on Medical Imaging*, pp. 1962–1971, 2016.

[5] J.-C. Olivo-Marin, "Extraction of spots in biological images using multiscale products," *Pattern Recognition*, vol. 35, no. 9, pp. 1989–1996, 2002.

[6] M. M. Benimam, A. Frafjord, A. Corthay, T. Lagache, J.-C. Olivo-Marin, and V. Meas-Yedid, "Nsclc tumor microenvironment: Annotated histopathology image dataset," June 2025.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *MICCAI2015*, 2015.

[8] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, pp. 386–397, 3.

[9] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, "Solov2: Dynamic and fast instance segmentation," *Advances in Neural Information Processing Systems*, vol. 2020-Decem, 3.

[10] S. He, K. T. Minn, L. Solnica-Krezel, M. A. Anastasio, and H. Li, "Deeply-supervised density regression for automatic cell counting in microscopy images," *Medical Image Analysis*, vol. 68, p. 101892, 2021.

[11] R. Cipolla, Y. Gal, and A. Kendall, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7482–7491, 2018.

[12] S. Graham, Q. D. Vu, M. Jahanifar, S. E. A. Raza, F. Minhas, D. Snead, and N. Rajpoot, "One model is all you need: Multi-task learning enables simultaneous histology image segmentation and classification," 2022.

[13] W. Xie, J. A. Noble, and A. Zisserman, "Microscopy cell counting and detection with fully convolutional regression networks," in *MICCAI*, (Cham), pp. 283–292, Springer International Publishing, 2015.

[14] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, "Beyond classification: Structured regression for robust cell detection using convolutional neural network," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), (Cham), pp. 358–365, Springer International Publishing, 2015.

[15] J. M. J. Valanarasu and V. M. Patel, "Unext: Mlp-based rapid medical image segmentation network," 2022.

[16] M. M. Benimam, V. Meas-Yedid, S. Mukherjee, A. Frafjord, A. Corthay, T. Lagache, and J. C. Olivo-Marin, "Statistical analysis of spatial patterns in tumor microenvironment images:," *Nature Commun*, vol. 16, p. 3090, 3.