nature medicine

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41591-020-0900-x

In the format provided by the authors and unedited.

Geospatial immune variability illuminates differential evolution of lung adenocarcinoma

Khalid AbdulJabbar[1,2,53], Shan E. Ahmed Raza[1,2,53], Rachel Rosenthal[3,4], Mariam Jamal-Hanjani[3,5], Selvaraju Veeriah[3,4], Ayse Akarca[6], Tom Lund[7], David A. Moore[3,6], Roberto Salgado[8,9], Maise Al Bakir[4], Luis Zapata[1,2], Crispin T. Hiley[3,4], Leah Officer[10], Marco Sereno[11], Claire Rachel Smith[11], Sherene Loi[9], Allan Hackshaw[12], Teresa Marafioti[6], Sergio A. Quezada[13], Nicholas McGranahan[3,14], John Le Quesne[10,11,15] ✉, TRACERx Consortium*, Charles Swanton[3,4,5,54] ✉ and Yinyin Yuan[1,2,54] ✉

[1]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK. [2]Division of Molecular Pathology, The Institute of Cancer Research, London, UK. [3]Cancer Research UK Lung Cancer Centre of Excellence, University College London Cancer Institute, London, UK. [4]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK. [5]Department of Medical Oncology, University College London Hospitals NHS Foundation Trust, London, UK. [6]Department of Cellular Pathology, University College London, University College Hospital, London, UK. [7]Translational Immune Oncology Group, Centre for Molecular Medicine, Royal Marsden Hospital NHS Trust, London, UK. [8]Department of Pathology, GZA-ZNA-Ziekenhuizen, Antwerp, Belgium. [9]Division of Research, Peter MacCallum Cancer Centre, University of Melbourne, Melbourne, Victoria, Australia. [10]MRC Toxicology Unit, University of Cambridge, Leicester, UK. [11]Leicester Cancer Research Centre, University of Leicester, Leicester, UK. [12]Cancer Research UK & University College London Cancer Trials Centre, University College London, London, UK. [13]Cancer Immunology Unit, University College London Cancer Institute, London, UK. [14]Cancer Genome Evolution Research Group, University College London Cancer Institute, University College London, London, UK. [15]Glenfield Hospital, University Hospitals Leicester NHS Trust, Leicester, UK. [53]These authors contributed equally: Khalid Abdul Jabbar, Shan E. Ahmed Raza. [54]These authors jointly supervised this work: Charles Swanton, Yinyin Yuan. *A full list of authors and their affiliations appears at the end of the paper. ✉e-mail: jlq2@leicester.ac.uk; Charles.Swanton@crick.ac.uk; Yinyin.Yuan@icr.ac.uk

STROBE Statement—Checklist of items that should be included in reports of ***cohort studies***

	Item No	Recommendation
Title and abstract DONE	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found
Introduction		
Background/rationale DONE	2	Explain the scientific background and rationale for the investigation being reported
Objectives DONE	3	State specific objectives, including any prespecified hypotheses
Methods		
Study design DONE	4	Present key elements of study design early in the paper
Setting DONE – methods/reporting summary	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants DONE - reporting summary	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up
		(*b*) For matched studies, give matching criteria and number of exposed and unexposed
Variables DONE	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement DONE	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias DONE – results/Fig3/Ext Fig 6	9	Describe any efforts to address potential sources of bias
Study size DONE – methods/Ext Fig 1	10	Explain how the study size was arrived at
Quantitative variables DONE	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods DONE	12	(*a*) Describe all statistical methods, including those used to control for confounding
		(*b*) Describe any methods used to examine subgroups and interactions
		(*c*) Explain how missing data were addressed
		(*d*) If applicable, explain how loss to follow-up was addressed
		(*e*) Describe any sensitivity analyses
Results		
Participants DONE – CONSRT diagram in Ext Fig 1	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed
		(b) Give reasons for non-participation at each stage
		(c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social)

DONE		and information on exposures and potential confounders
		(b) Indicate number of participants with missing data for each variable of interest
		(c) Summarise follow-up time (eg, average and total amount)
Outcome data DONE	15*	Report numbers of outcome events or summary measures over time
Main results DONE	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included
		(*b*) Report category boundaries when continuous variables were categorized
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses DONE	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses
Discussion		
Key results DONE	18	Summarise key results with reference to study objectives
Limitations DONE	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation DONE	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability DONE	21	Discuss the generalisability (external validity) of the study results
Other information		
Funding DONE	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

Comparison of tissue segmentation performance between deep learning and classic machine learning

We implemented the Micro-Net [1] algorithm for tissue segmentation which has been shown to perform accurate segmentation compared to the state-of-the-art algorithms. Due to complex structures in histology slides or weak staining, classic machine learning algorithms often fail and are problematic to tune. This problem can be more complex while segmenting sections without Eosin or cytoplasmic staining such as IHC-stained images due to weak contrast between the tissue region and the background glass.

As an independent comparison, we experimented classic methods such as threshold, active contours [2], watershed segmentation [3] and Support Vector Machines (SVM) based method trained on local binary pattern features [4] on 10 different images randomly selected from the TRACERx histology cohort (Supplementary Figures 1-20). Deep learning (MicroNet) outperformed all classic methods using various accuracy metrics as shown in Supplementary Table 6.

Supplementary Figure 10 demonstrates the effect of weaker staining on threshold and active contours algorithms, whereas MicroNet consistently performed better on all the images. The watershed algorithm segmented multiple regions in all images, with such variation, it was very hard to fine-tune the algorithm in order to merge all relevant tissue regions into a single segment. For SVM, the local binary patterns (LBP) features were extracted to segment the tissue regions, however, a major limitation can be observed in the form of discarding Eosin only areas, as shown on Supplementary Figure 9 and Supplementary Figure 21.

Supplementary Table 6: Quantitative comparison of tissue segmentation results for proposed (Micro-Net) vs classic machine learning.

Method	Dice	Object Dice	Pixel Accuracy	F1-Score
Threshold	0.631500453	0.491553304	0.699176744	0.000949264
Watershed	0.38605082	0.370697822	0.447965797	0.176859504
Chan_Vese	0.598405502	0.469242029	0.667303606	0.000914946
SVM_LBP	0.917539751	0.915315159	0.910894913	0.749012437
MicroNet	0.961247872	0.955819821	0.964068566	0.773748864

1. Raza, S. E. A. *et al.* Micro-Net: A unified model for segmentation of various objects in microscopy images. *Med. Image Anal.* **52,** 160–173 (2019).

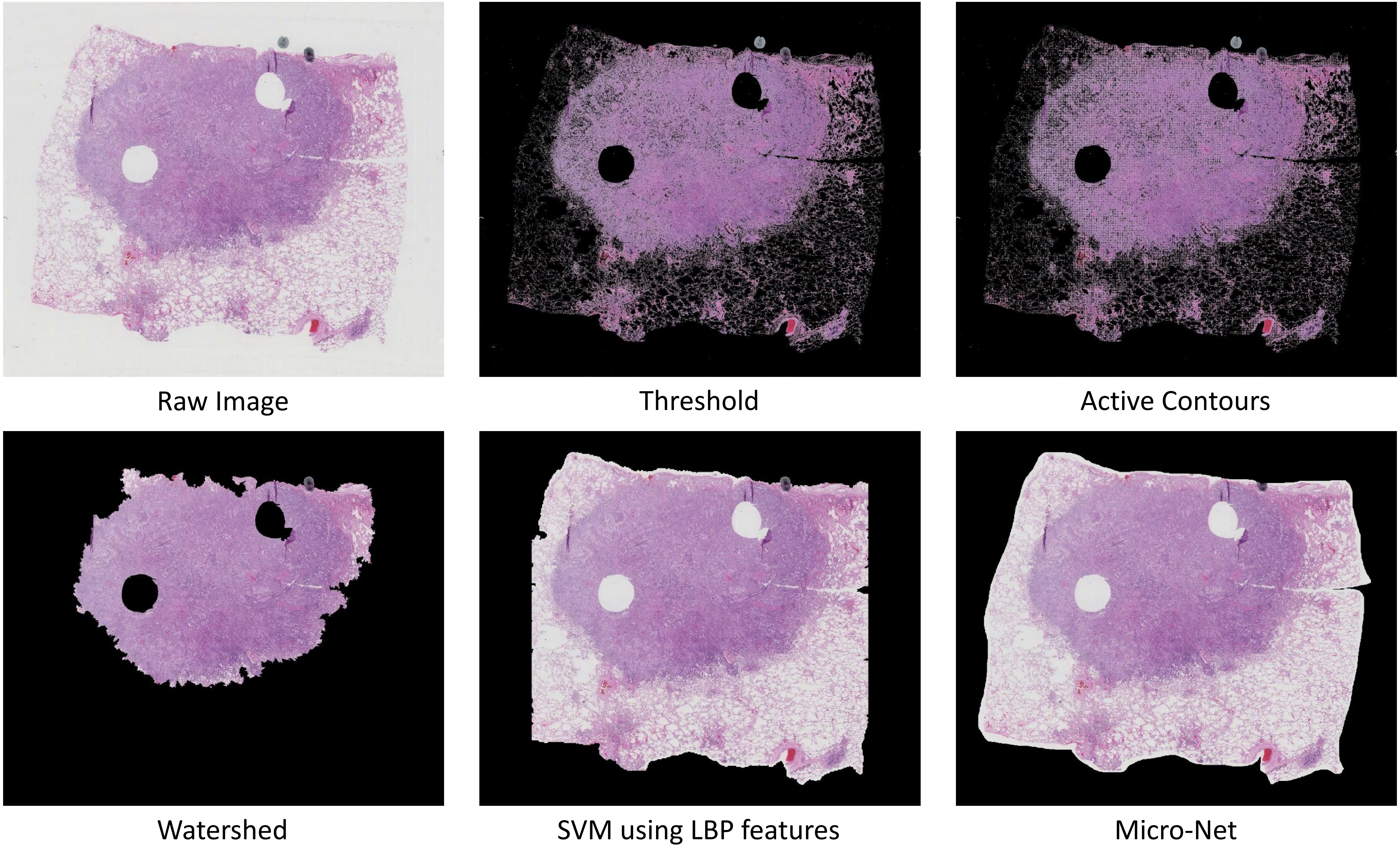
2. Chan, T. F., Sandberg, B. Y. & Vese, L. A. Active Contours without Edges for Vector-Valued Images. *J. Vis. Commun. Image Represent.* **11,** 130–141 (2000).

3. Meyer, F. Topographic distance and watershed lines. *Signal Processing* **38,** 113–125 (1994).

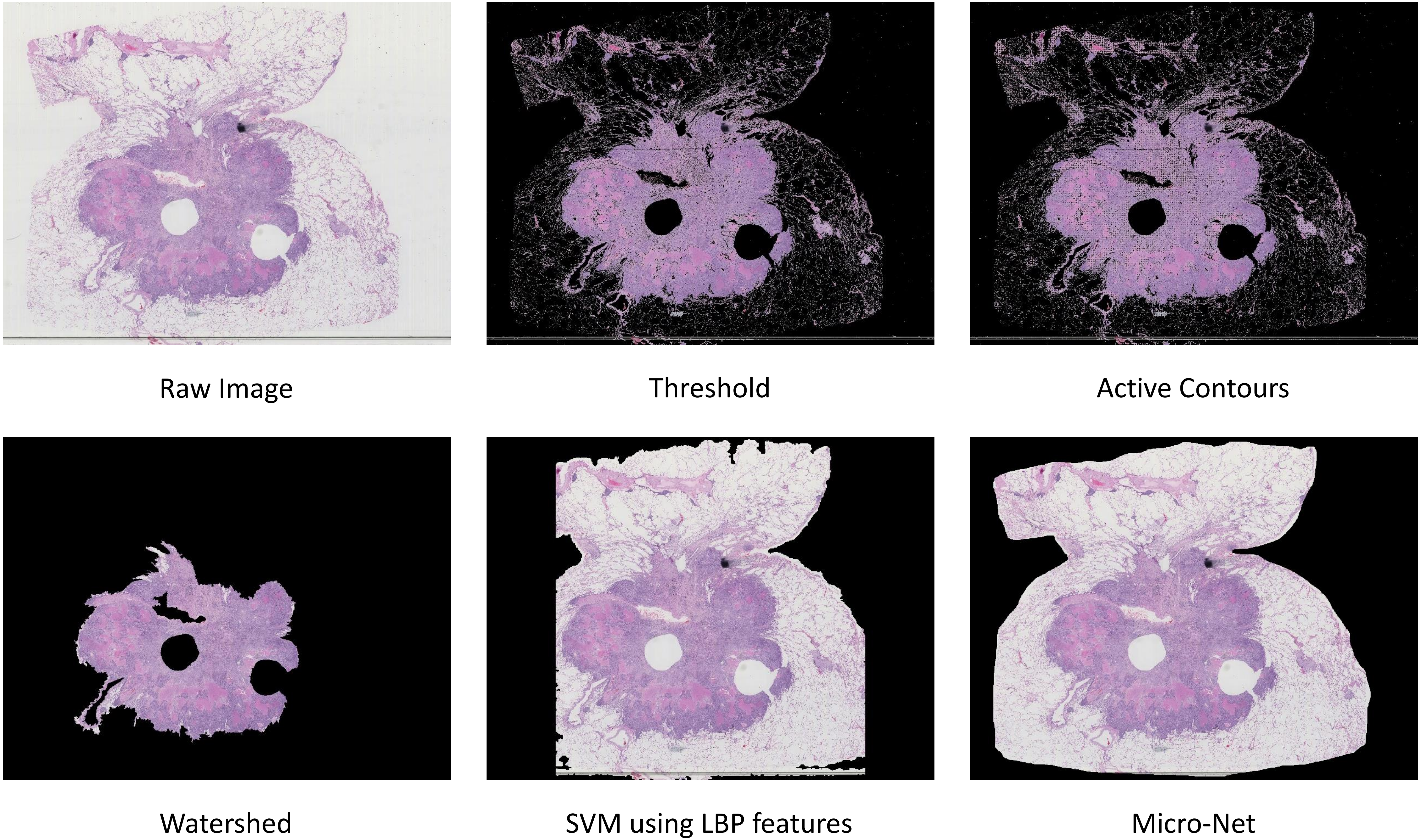
4. Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24,** 971–987 (2002).

Supplementary Figures 1-20: comparison of H&E tissue segmentation across five different methods: MicroNet, threshold, active contours, watershed segmentation and SVM based method trained on local binary pattern features. Supplementary Figures 1-10 show the results for segmenting entire diagnostic slides and Supplementary Figures 11-20 show various zoomed-in examples.

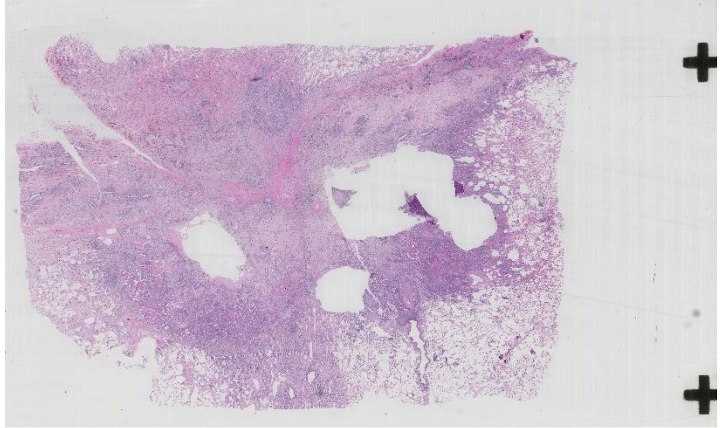
1 Supplementary Figure 1



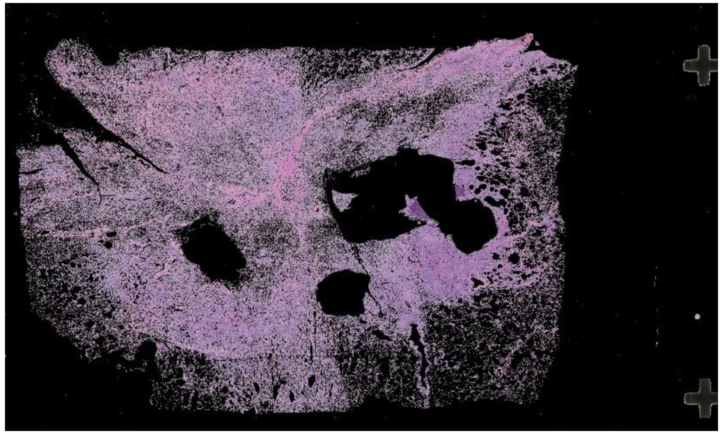
2 Supplementary Figure 2



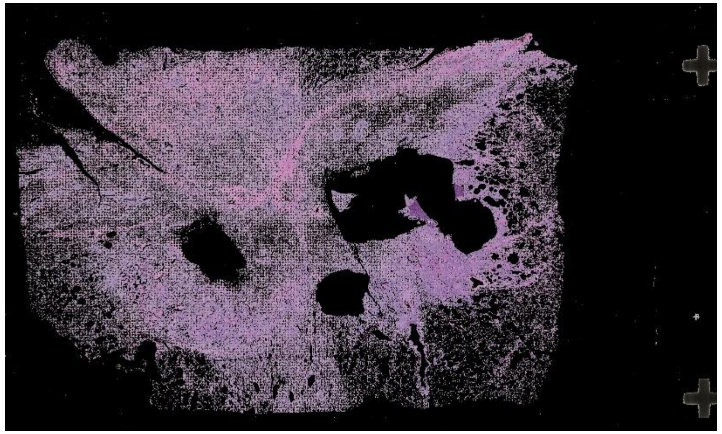
3 Supplementary Figure 3



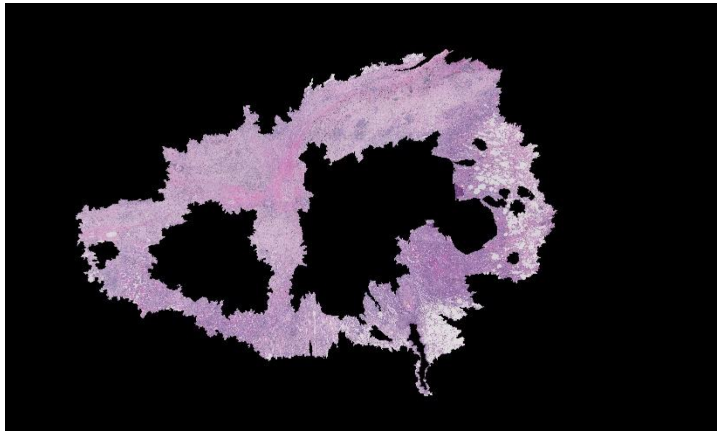
Raw Image



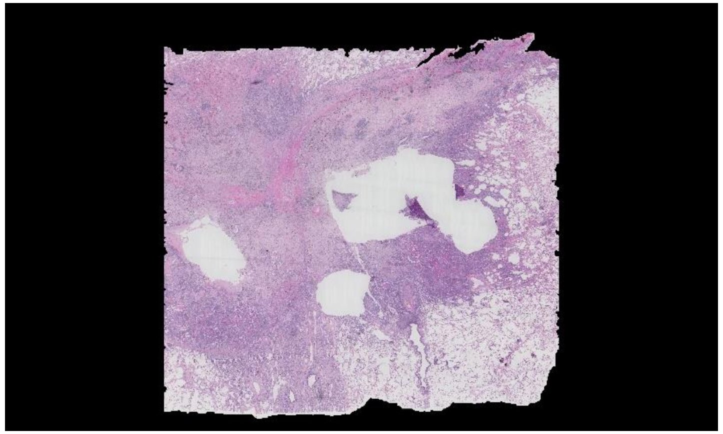
Threshold



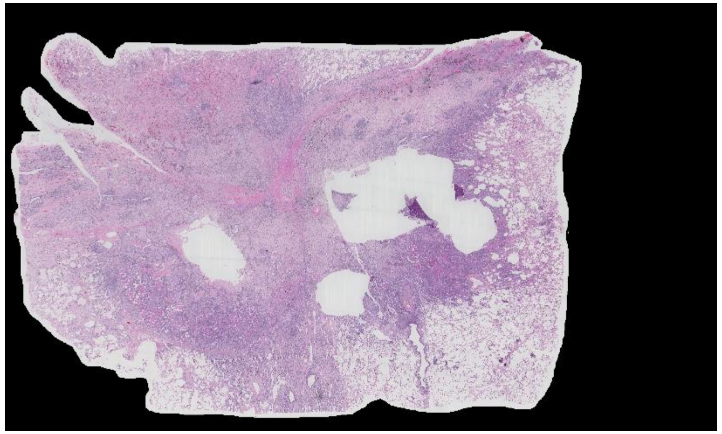
Active Contours



Watershed

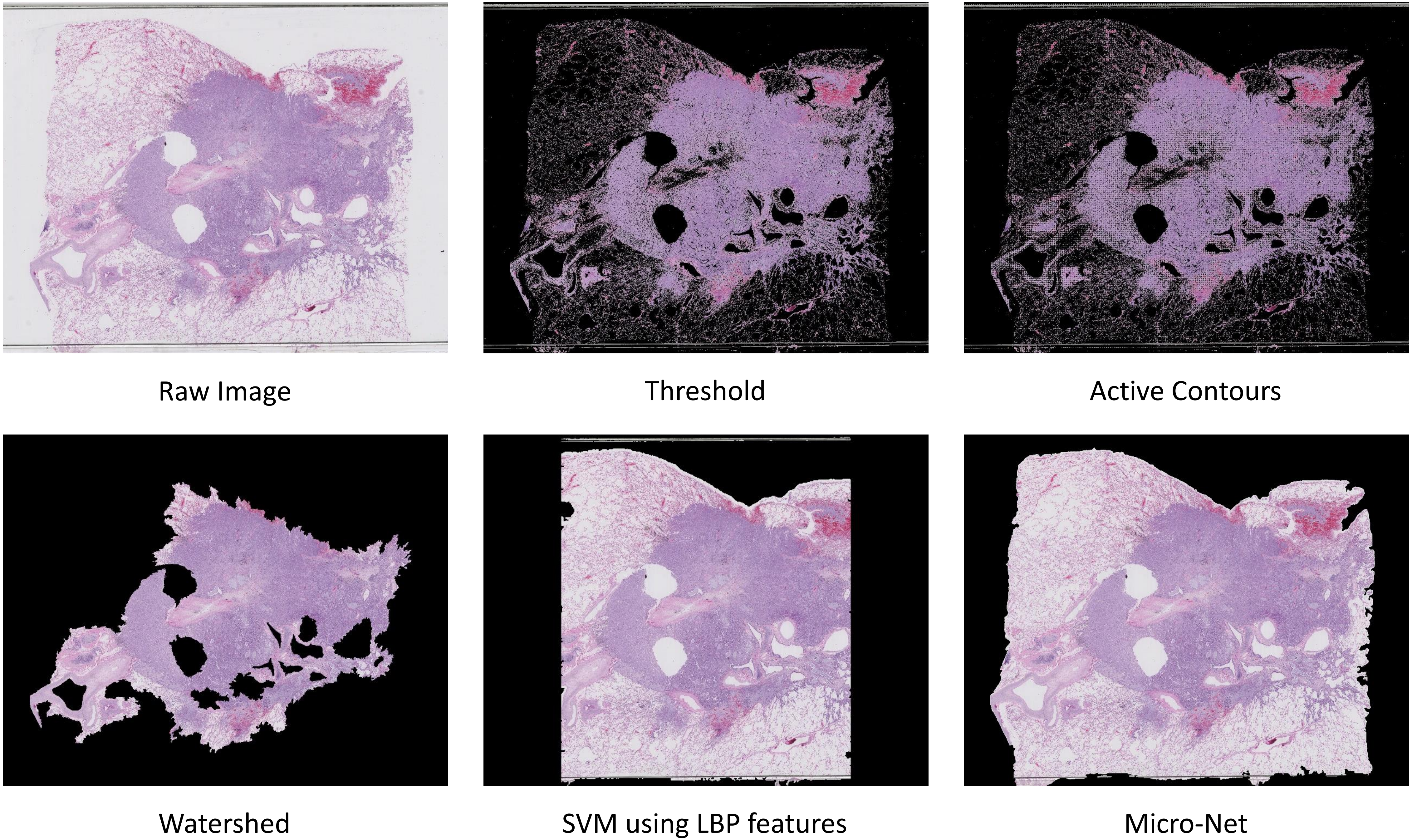


SVM using LBP features

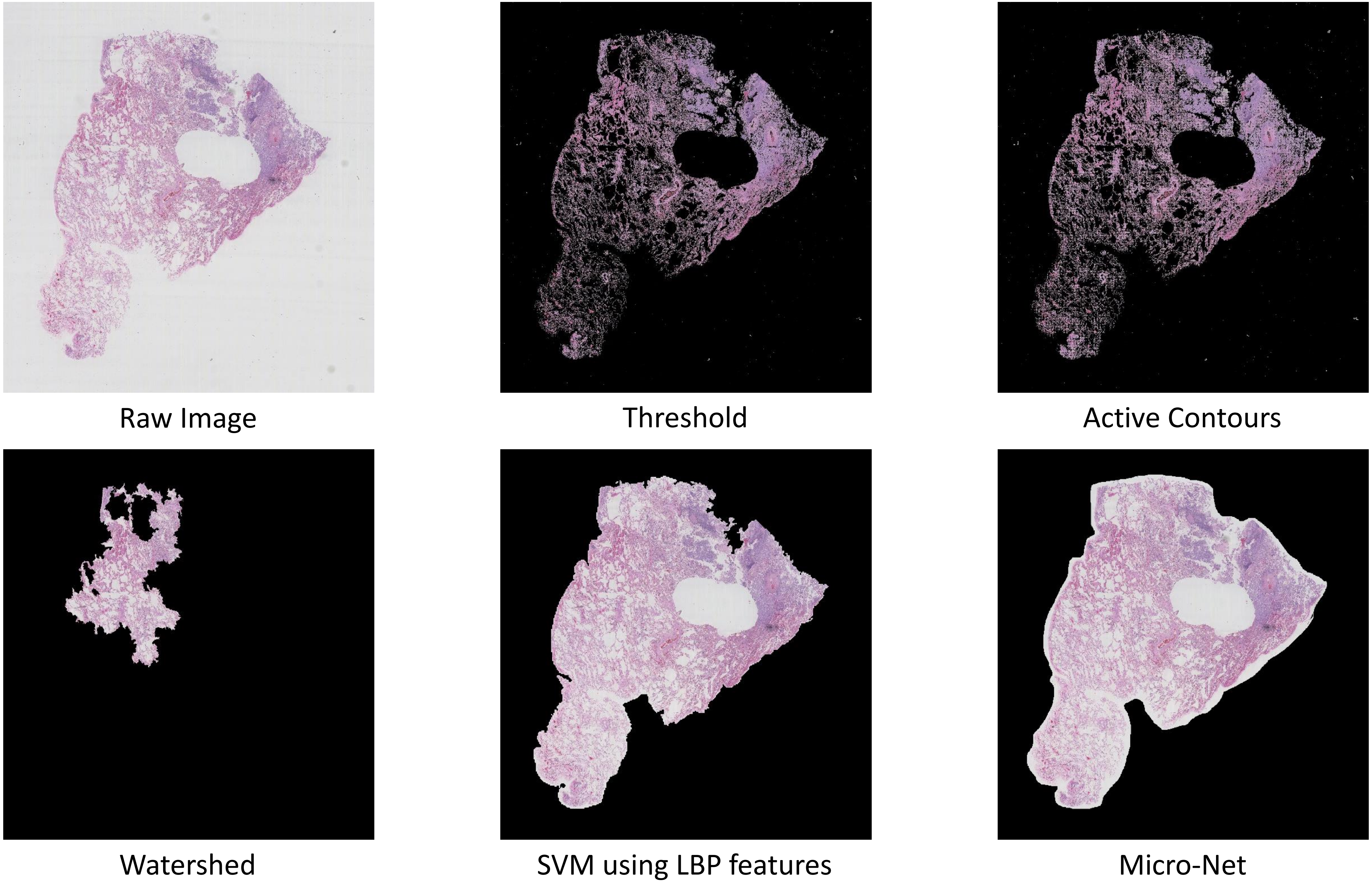


Micro-Net

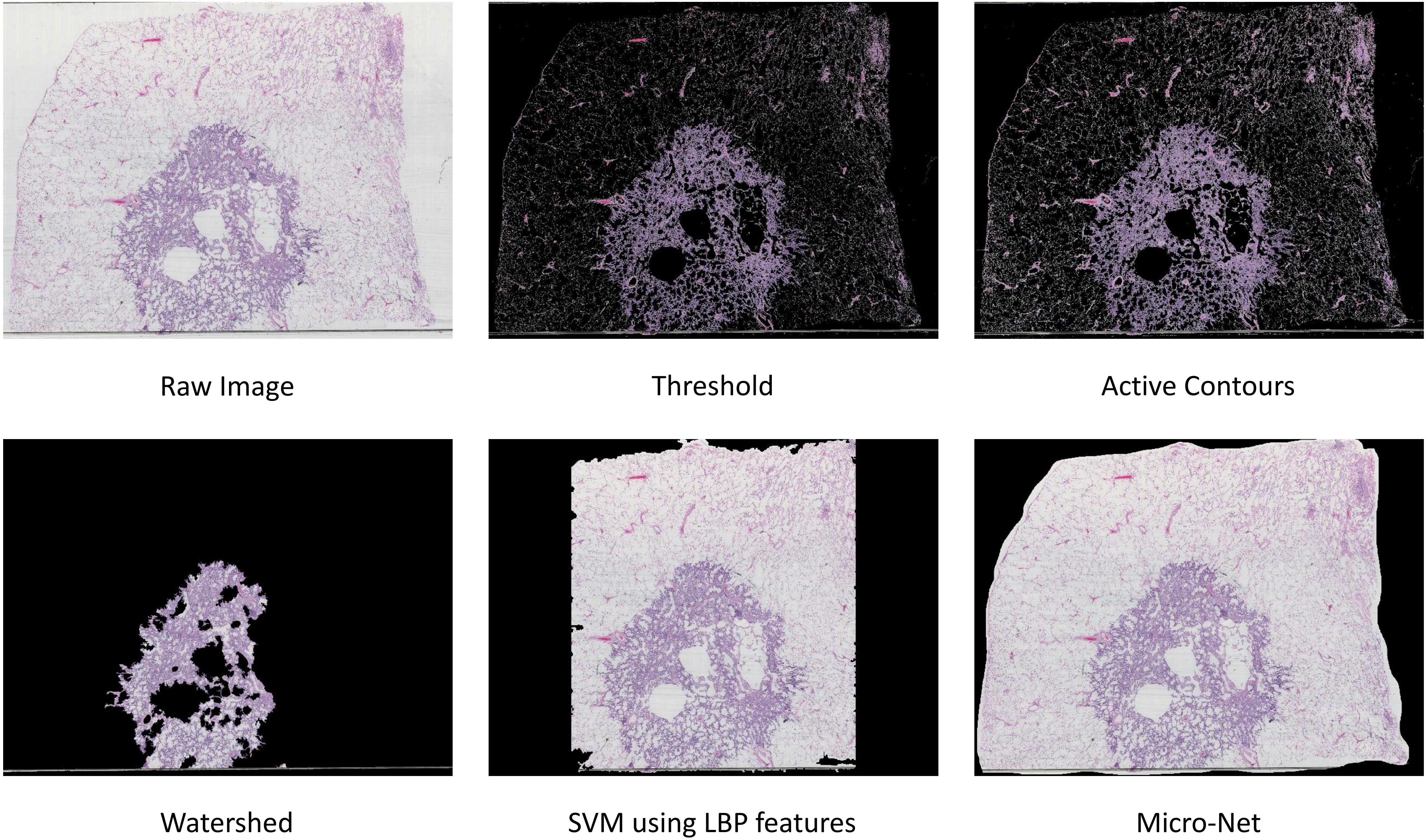
Supplementary Figure 4



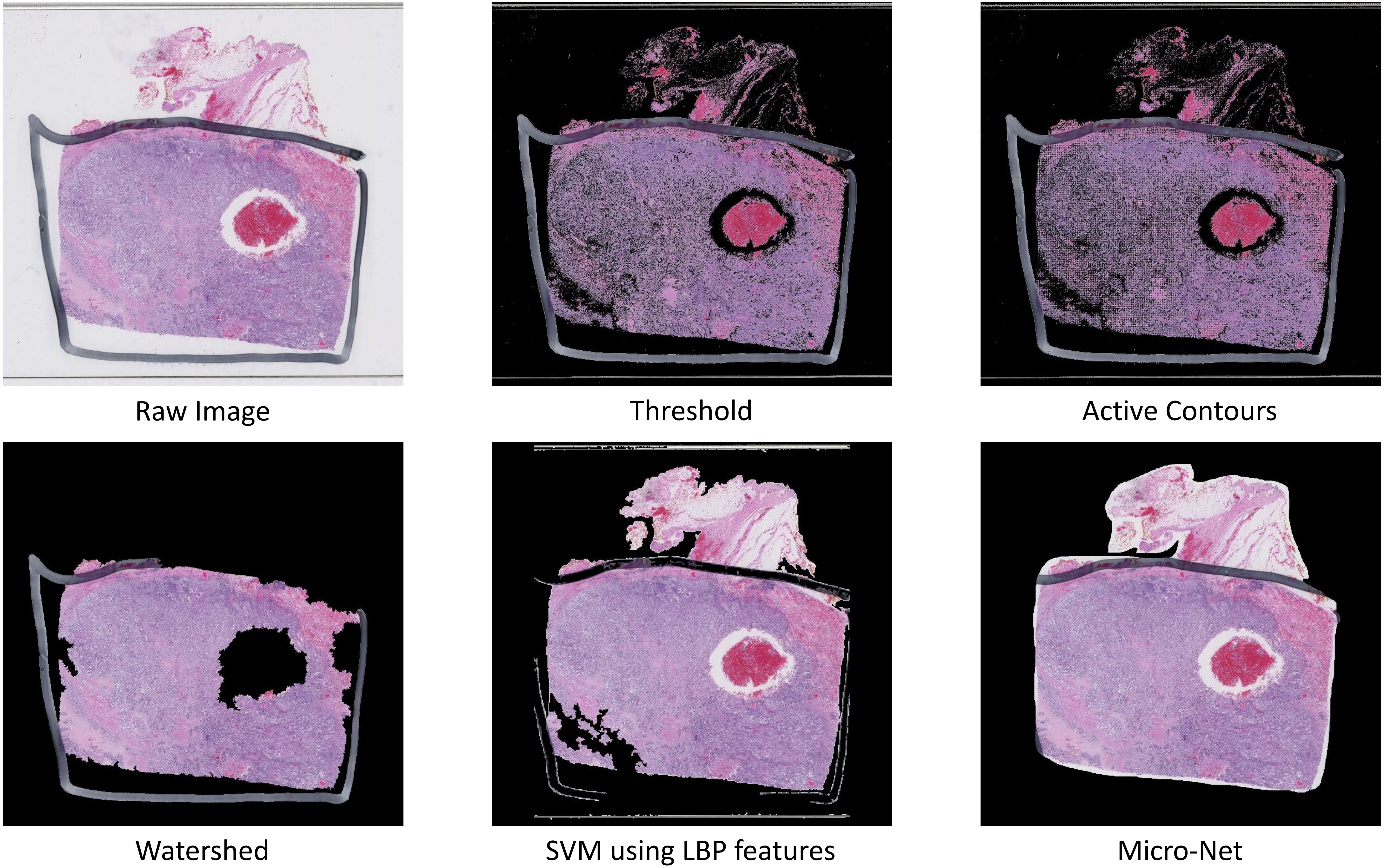
5 Supplementary Figure 5



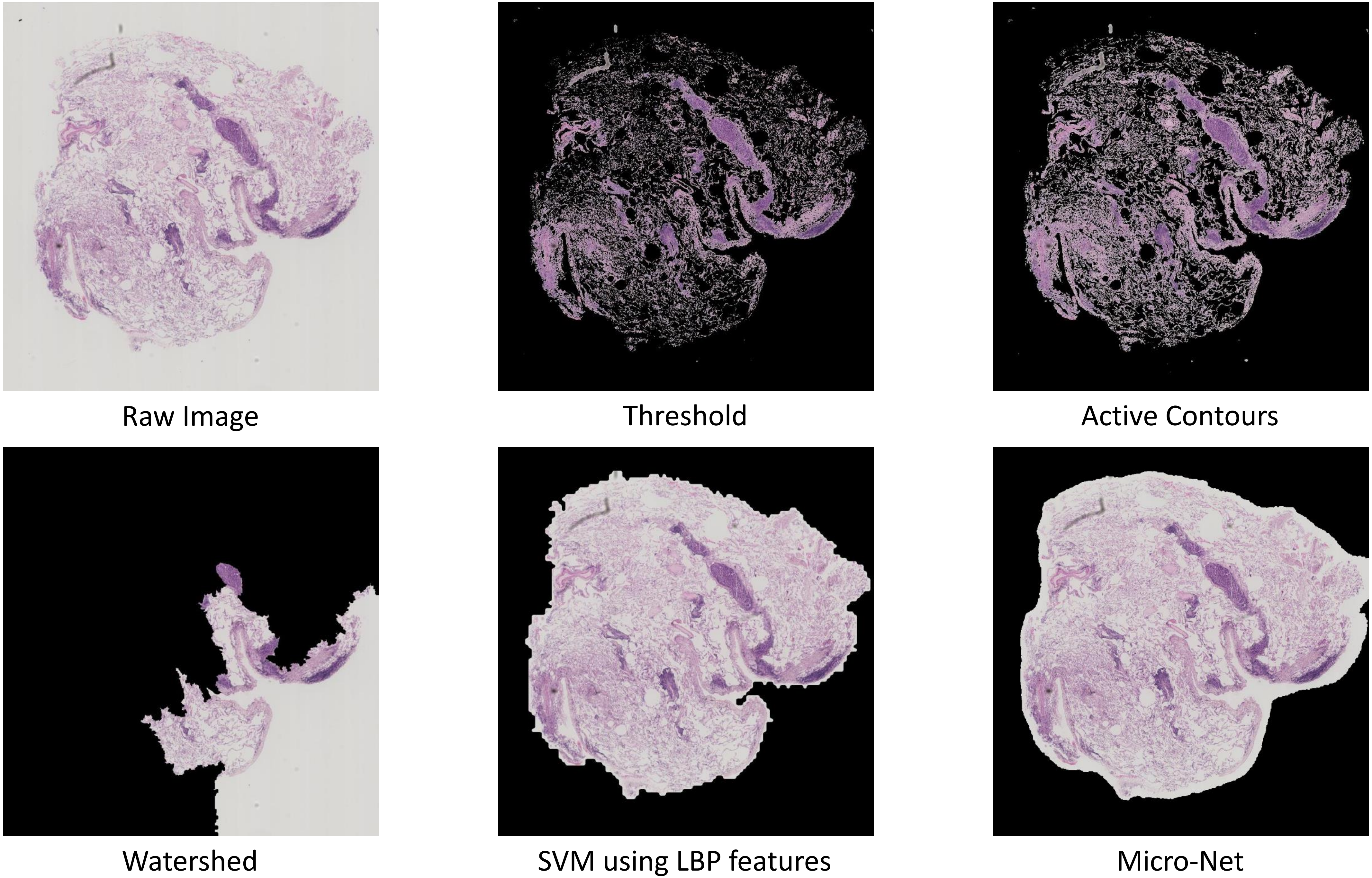
Supplementary Figure 6



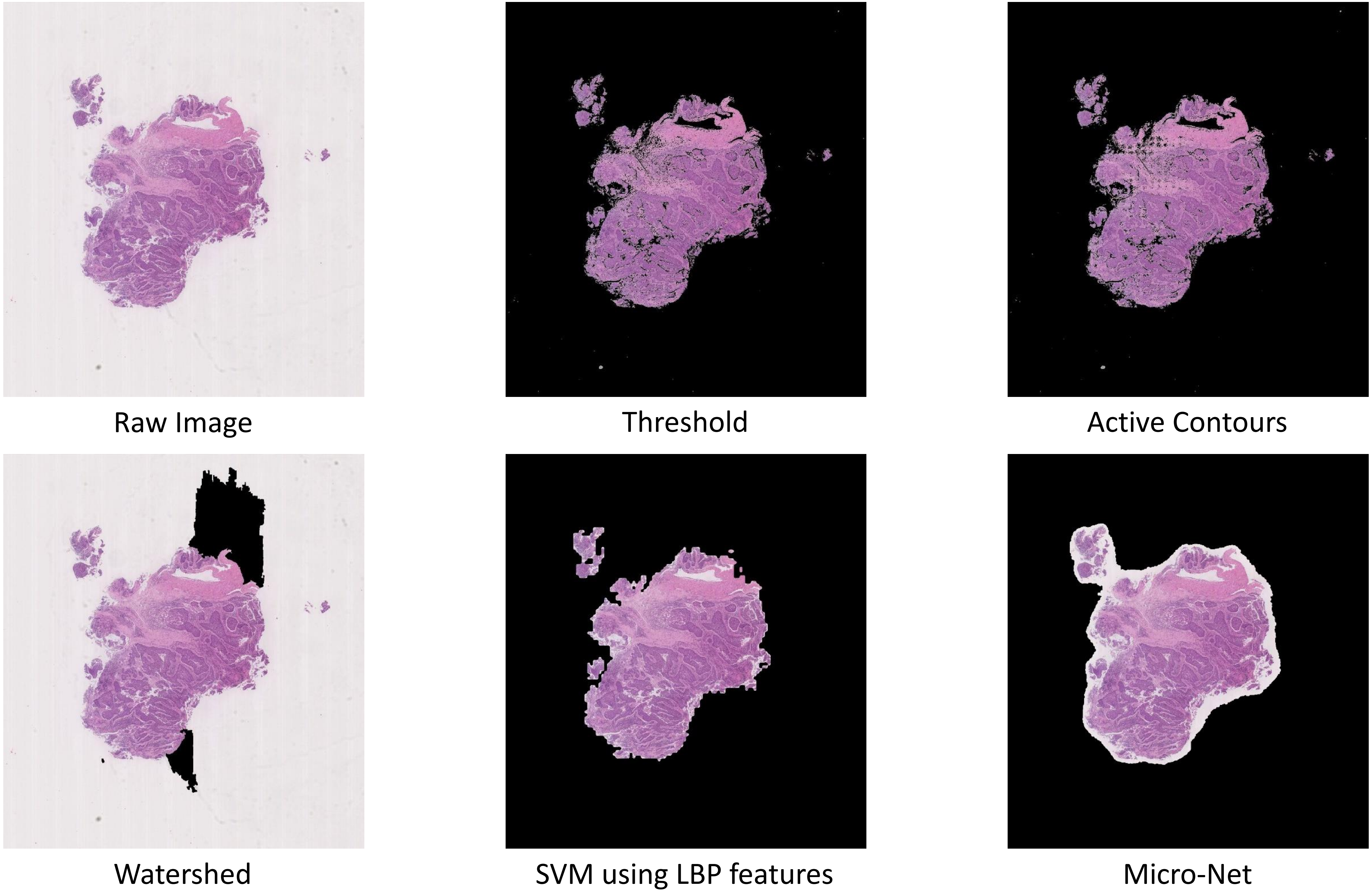
Supplementary Figure 7



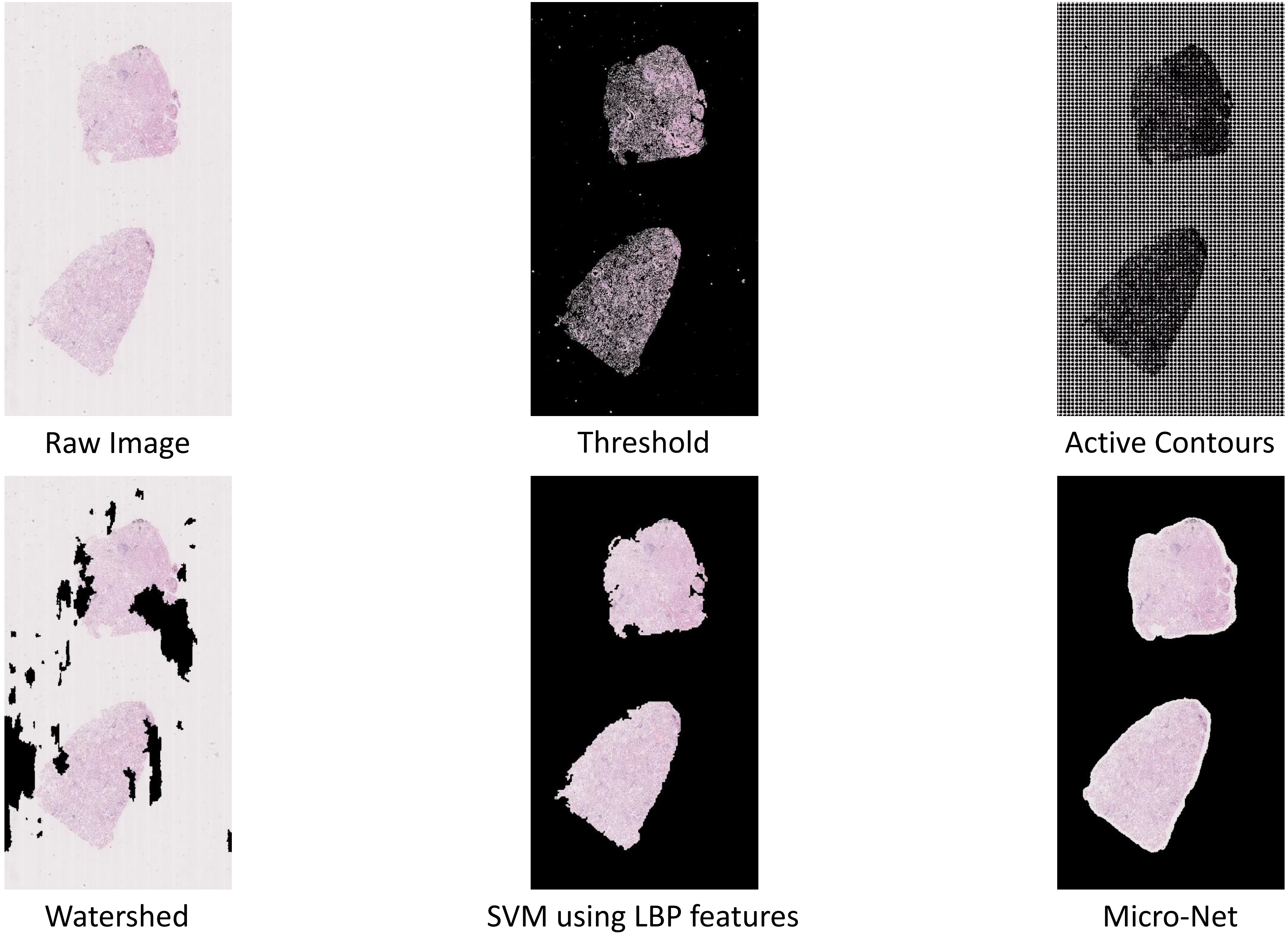
Supplementary Figure 8



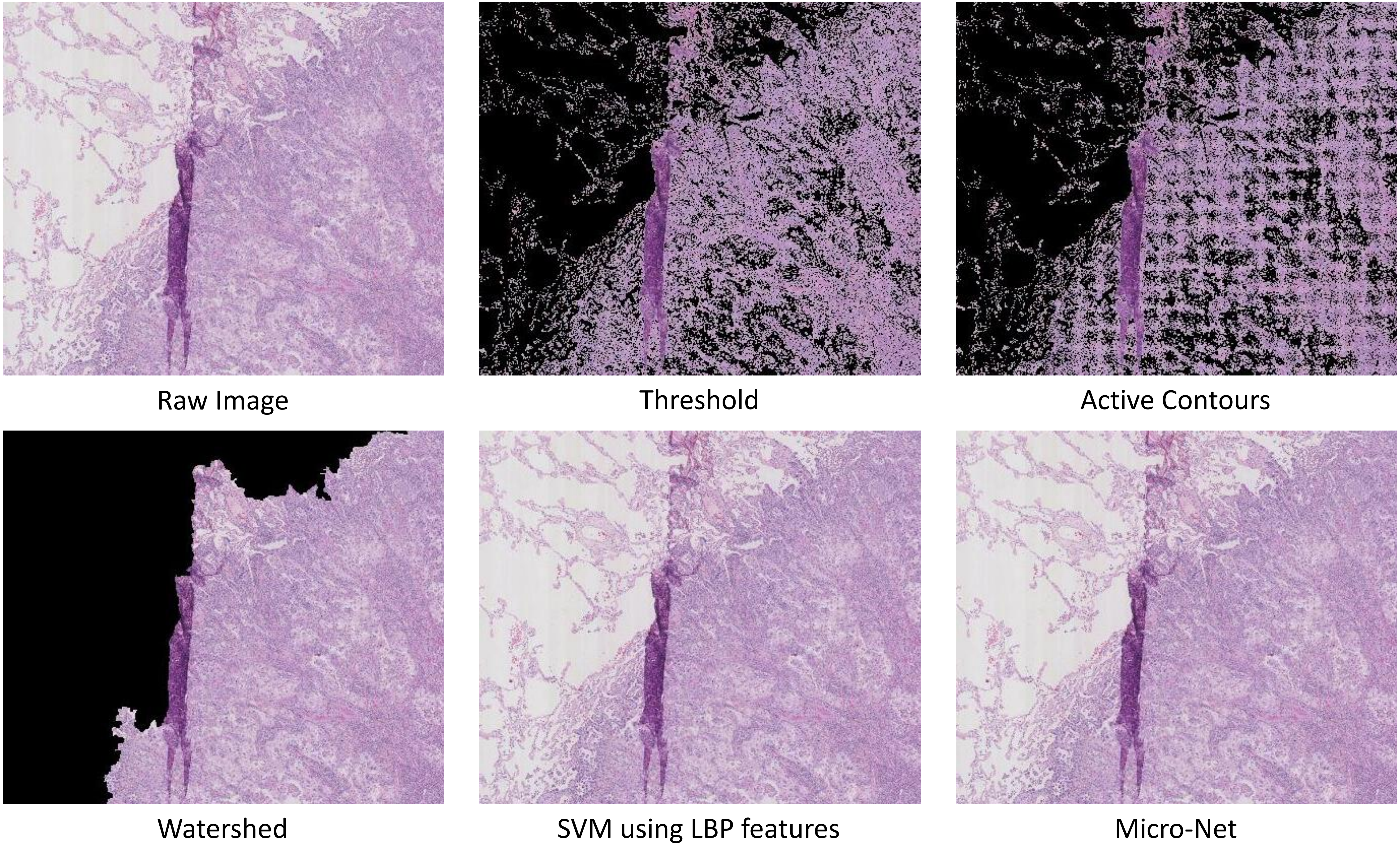
Supplementary Figure 9



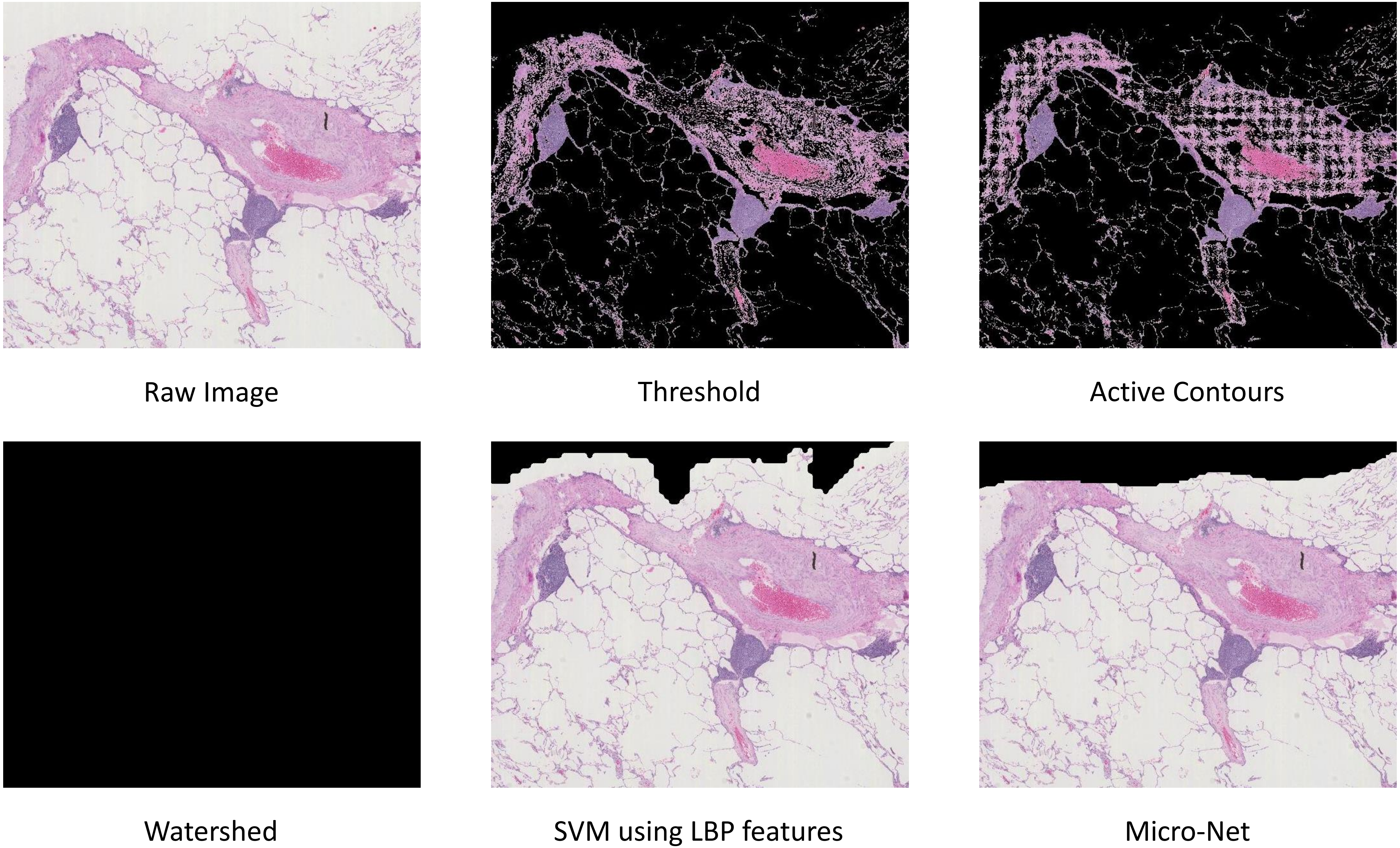
10 Supplementary Figure 10



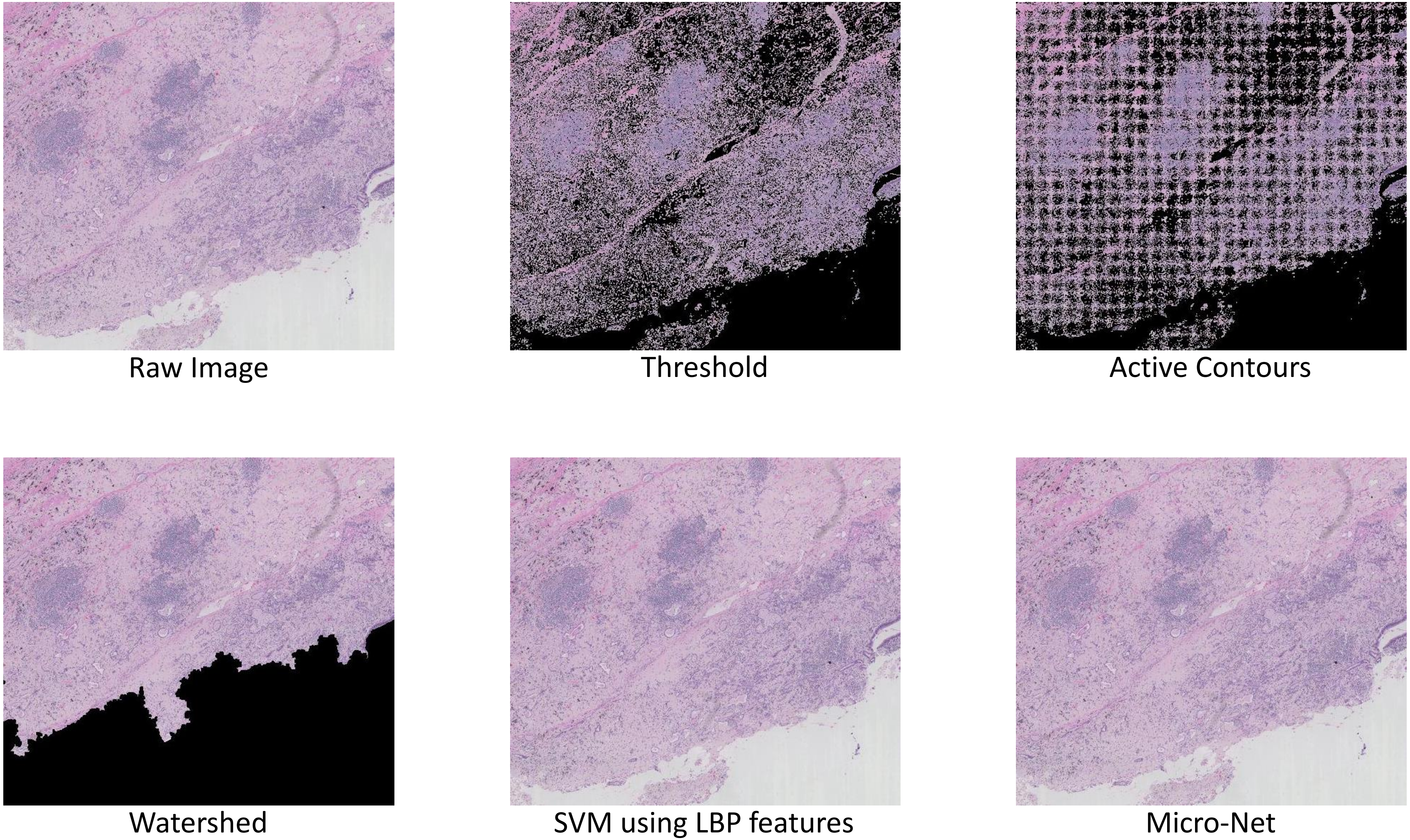
11 Supplementary Figure 11



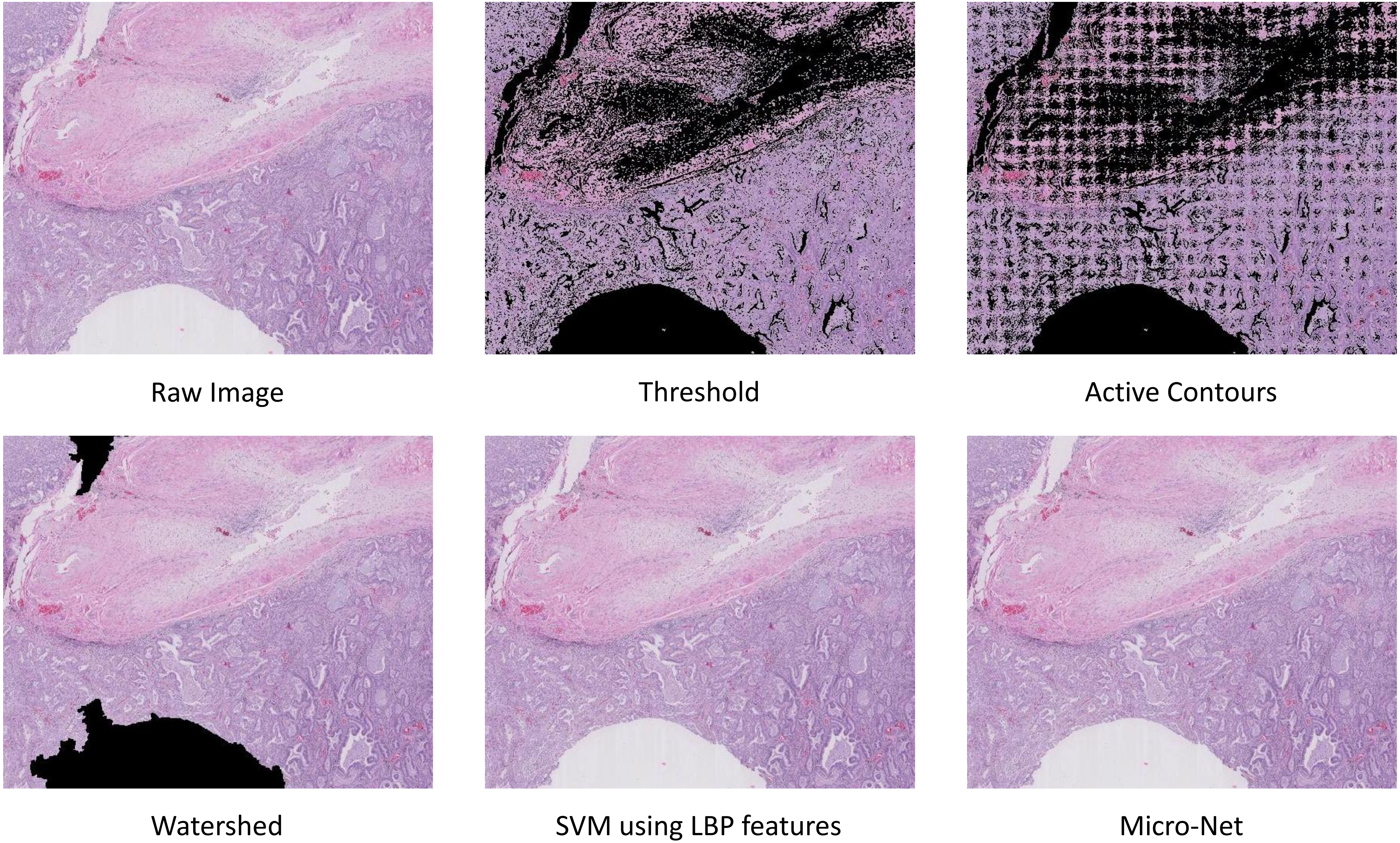
Supplementary Figure 12



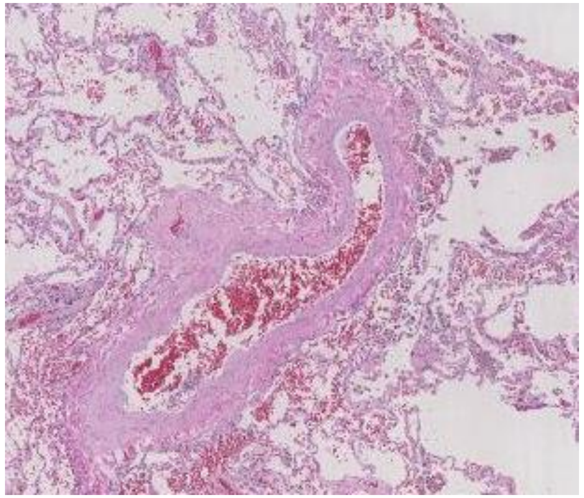
Supplementary Figure 13



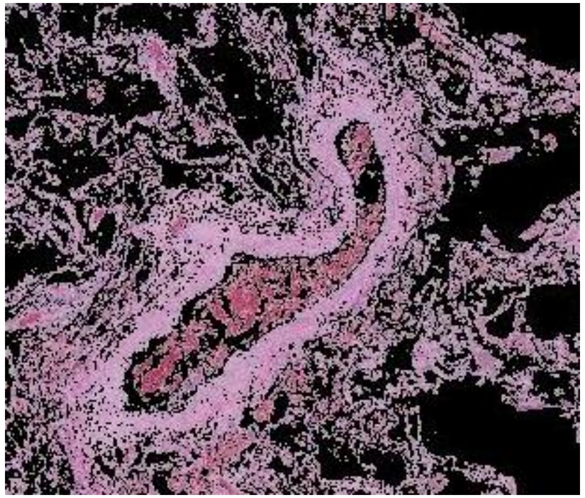
 Supplementary Figure 14



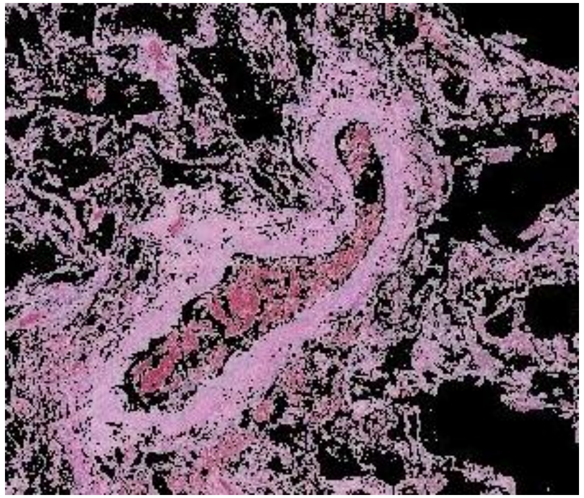
Supplementary Figure 15



Raw Image

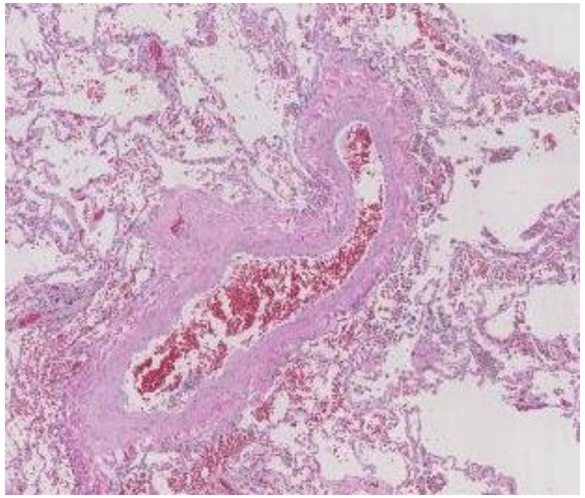


Threshold

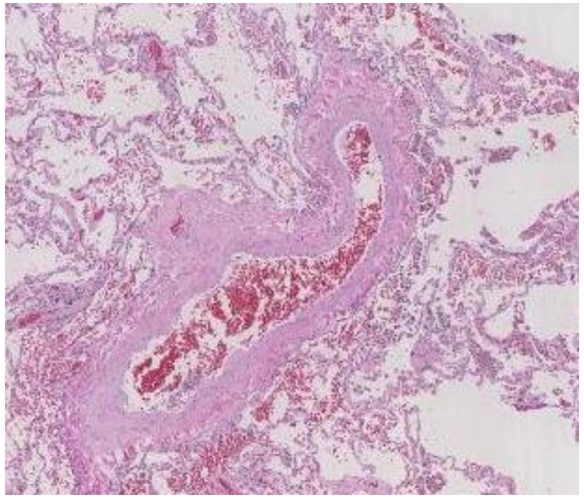


Active Contours

Watershed

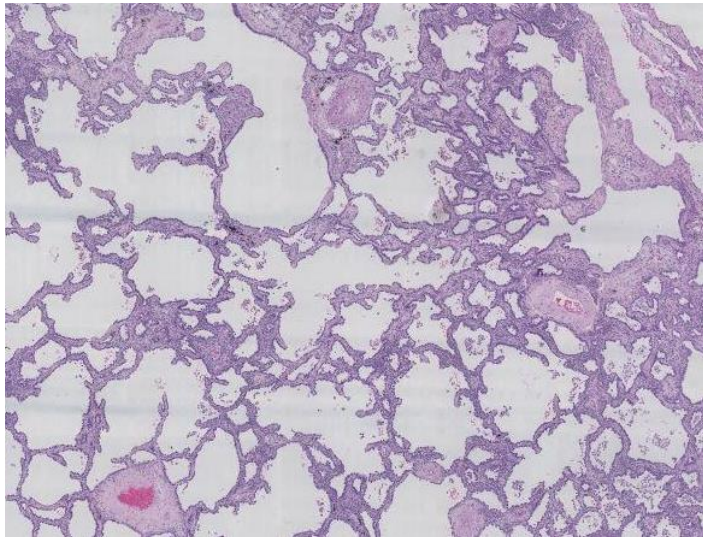


SVM using LBP features

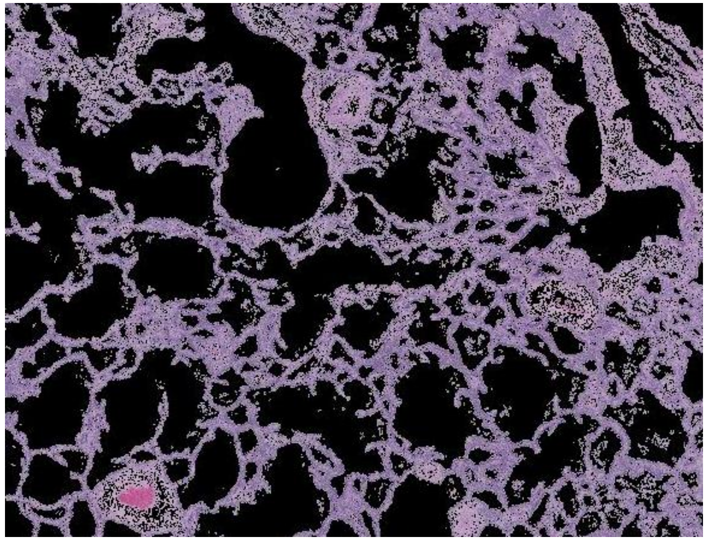


Micro-Net

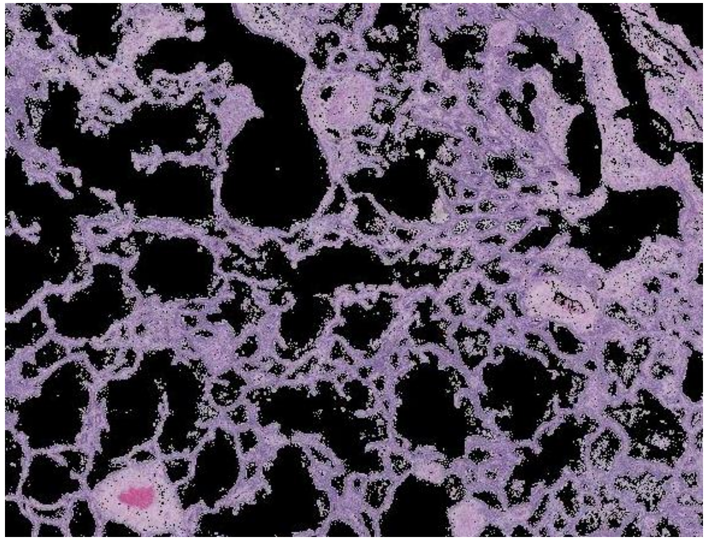
Supplementary Figure 16



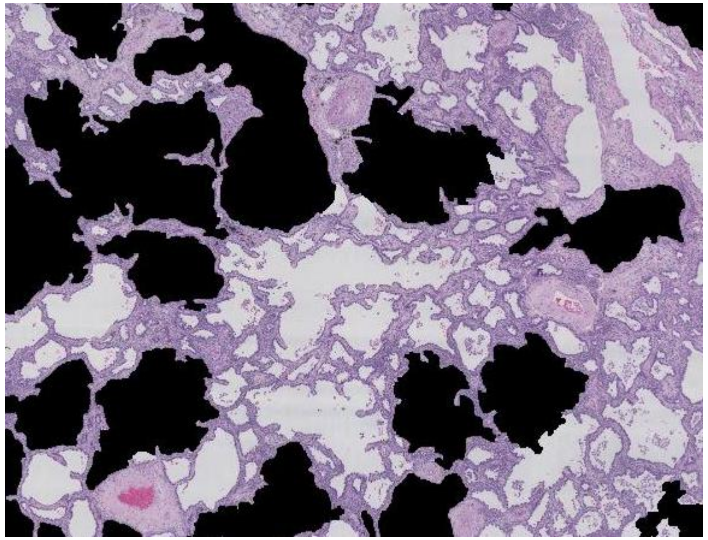
Raw Image



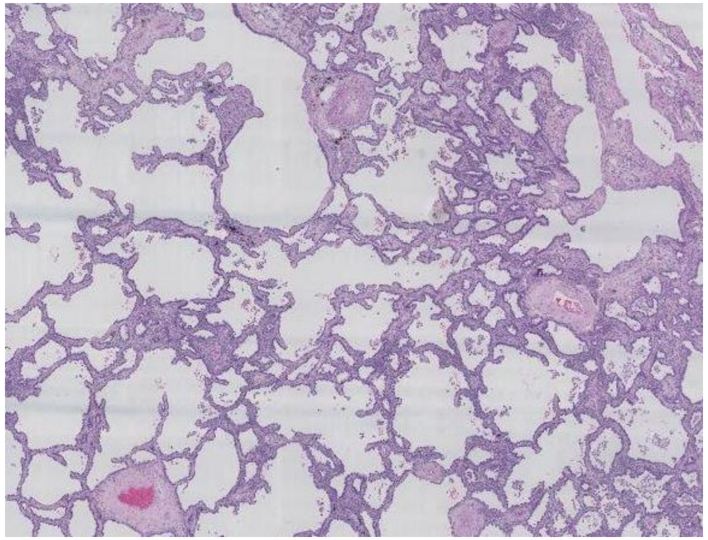
Threshold



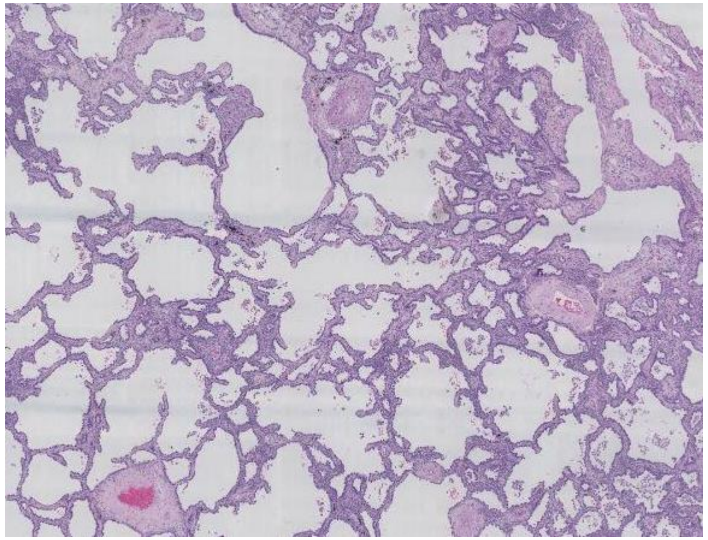
Active Contours



Watershed

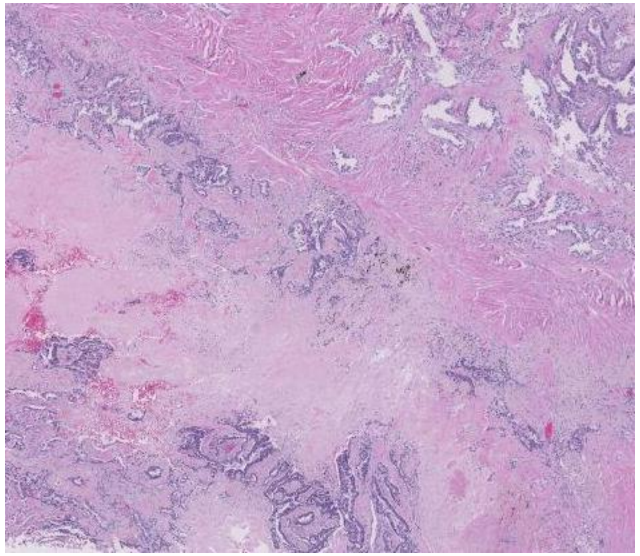


SVM using LBP features

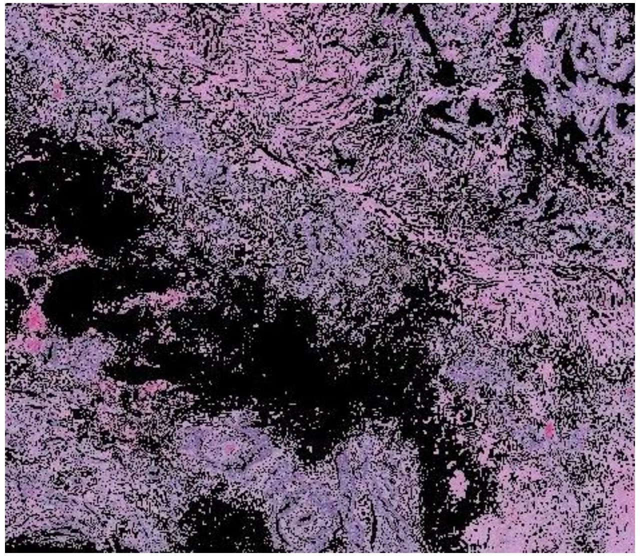


Micro-Net

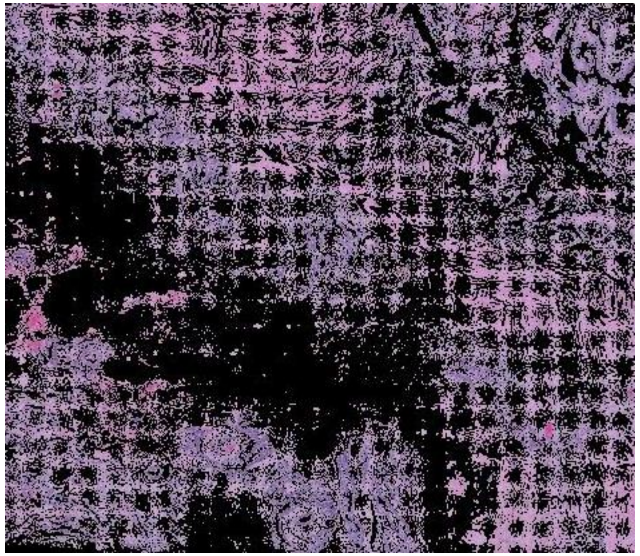
Supplementary Figure 17



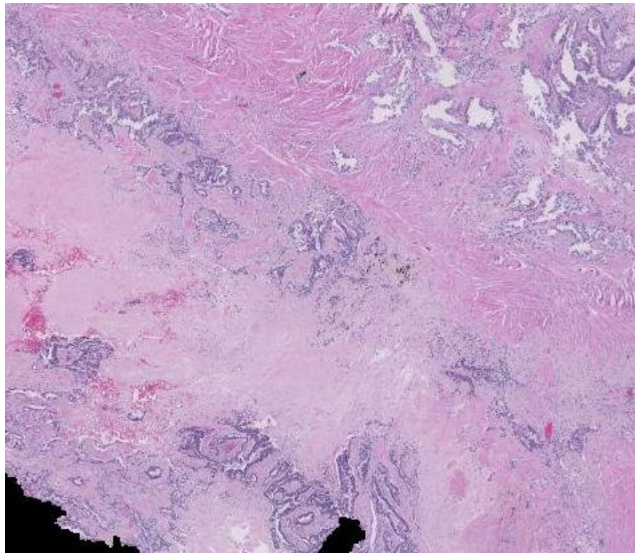
Raw Image



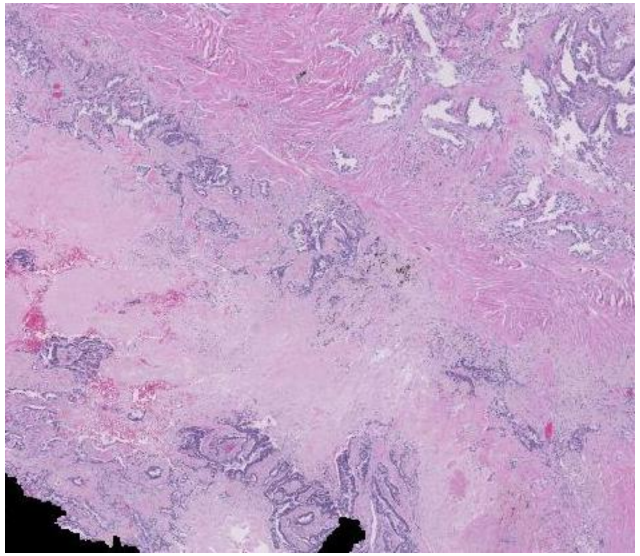
Threshold



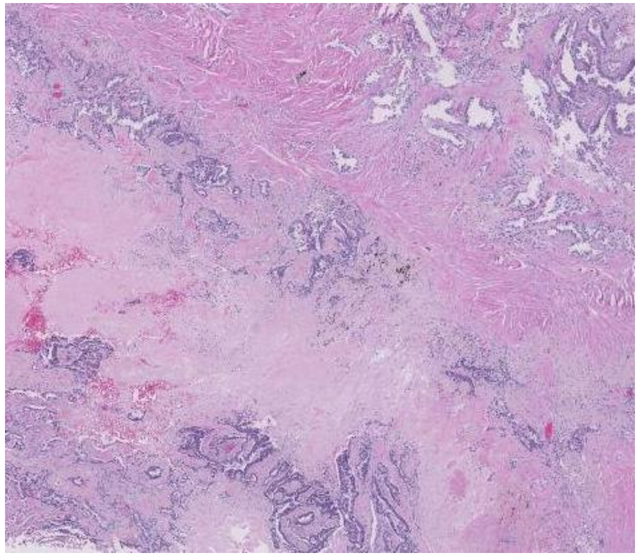
Active Contours



Watershed

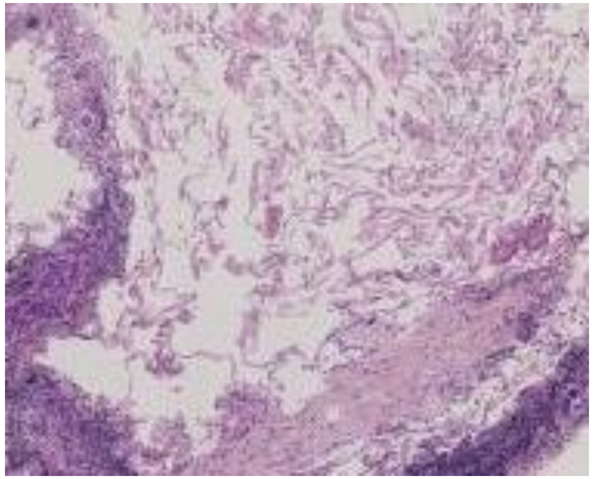


SVM using LBP features

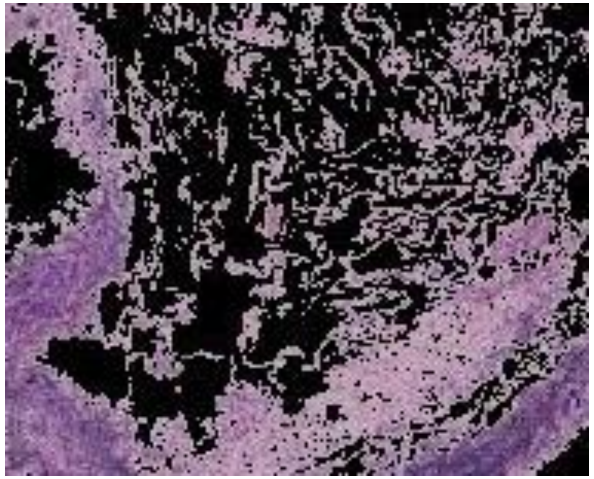


Micro-Net

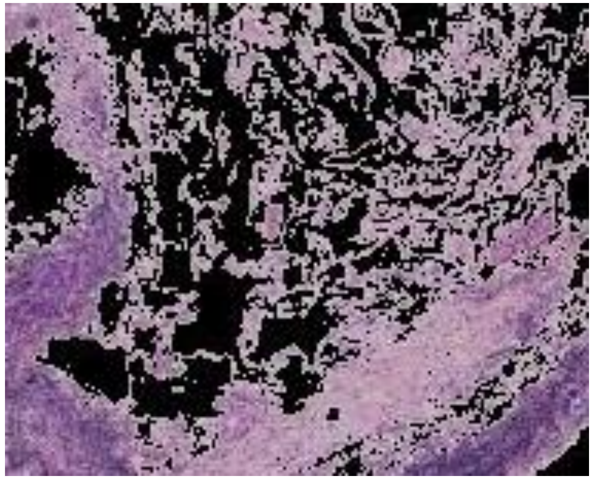
Supplementary Figure 18



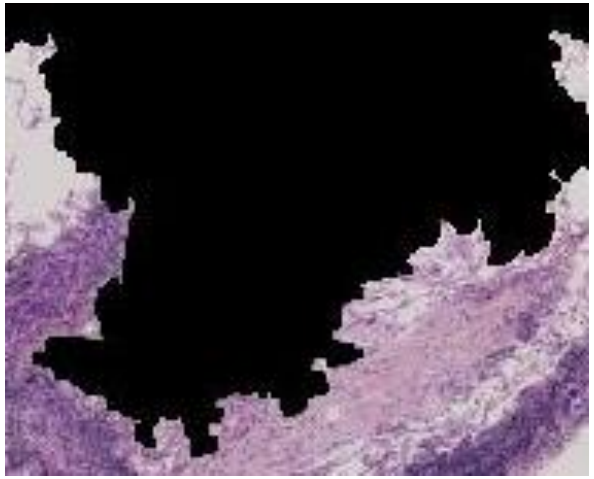
Raw Image



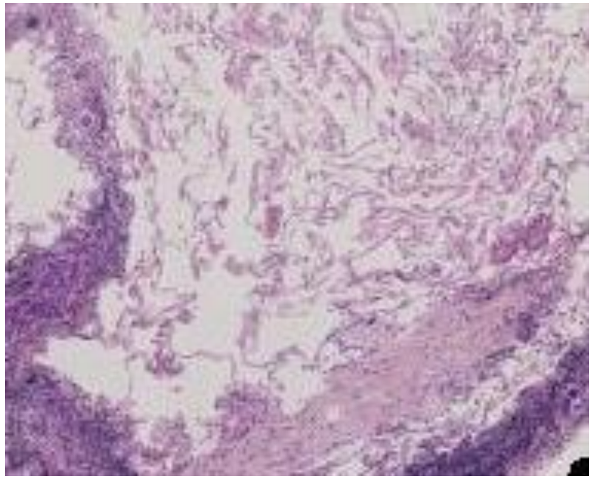
Threshold



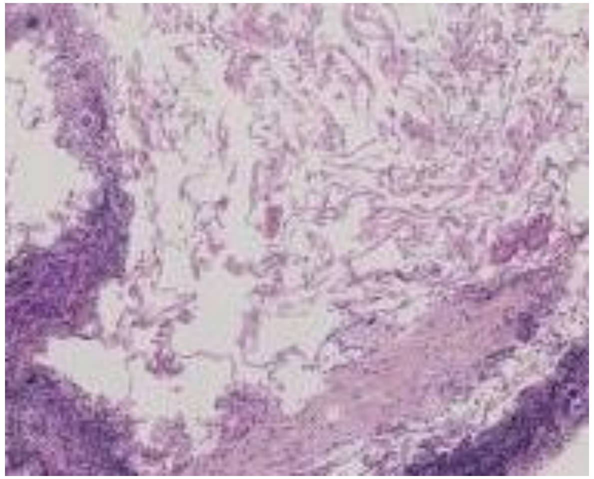
Active Contours



Watershed

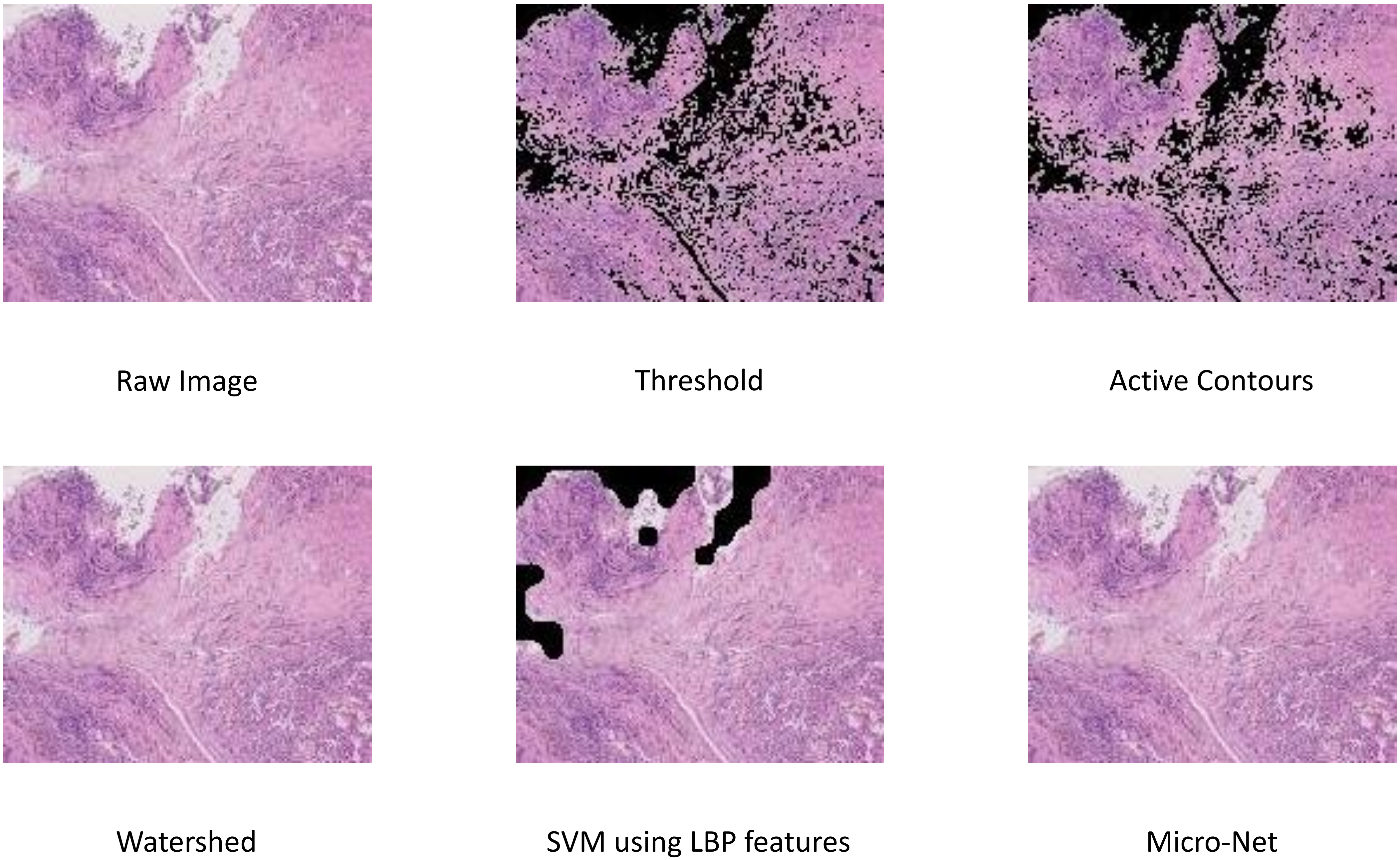


SVM using LBP features



Micro-Net

Supplementary Figure 19



Supplementary Figure 20