Supplementary Material: Mutual-Complementing Framework for Nuclei Detection and Segmentation in Pathology Image

Zunlei Feng[1,5,6,#], Zhonghua Wang[1,#], Xinchao Wang[2], Yining Mao[1], Thomas Li[3], Jie Lei[4], Yuexuan Wang[1], Mingli Song[1,5,6*]
[1]Zhejiang University, [2]National University of Singapore, [3]The University of Hong Kong, [4]Zhejiang University of Technology, [5]Zhejiang Provincial Key Laboratory of Service Robot, [6]Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies

In the supplementary materials, we provide experiment details, visual results of different cut filters and different scale filters, and experiment on NuCLS [1] dataset.

A. Experiment Details

In the experiment, the size of the middle-size nucleus patch is 42×42. For handling different scale nuclei, the middle-size nucleus patch is resized into 22×22, 30×30, 50×50, and 62×62. Finally, we get five scale nucleus patches.

For each nucleus patch, the correlation filter template is calculated with Eqn. (1).

$$\hat{w}^1 = \frac{\hat{\mathbf{x}}^* \odot \hat{\mathbf{y}}}{\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}} + \lambda}, \tag{1}$$

where $\hat{\mathbf{x}} = DFT(\mathbf{x})$, DFT denotes discrete Fourier transform, $\mathbf{x}$ denotes the cyclic shift feature vector extracted from the image patch I^p, $\hat{}$ is a shorthand for the DFT of a vector, λ is a regularization parameter. $\mathbf{x}$ is composed of two kinds of features: Color Name (CN) features [12] and Histogram of Oriented Gridients (HOG) [2] features. For each nucleus patch and the original image, we extract 11 channel CN features and 9 channel HOG features. The size of extracted feature map are half of size of cut nucleus patch and the original image.

scale	filter size	corresponding cut filter size
1	(11×11)	(9×9), (7×7) , (5×5), (3×3)
2	(15×15)	(13×13), (11×11) , (9×9), (7×7)
3	(21×21)	(17×17), (13×13) , (9×9), (5×5)
4	(25×25)	(21×21), (17×17) , (13×13), (9×9)
5	(31×31)	(27×27), (23×23) , (19×19), (15×15)

Table 1. The size of correlation filter and corresponding cut filters.

*Corresponding author. #Equal contribution to this work.

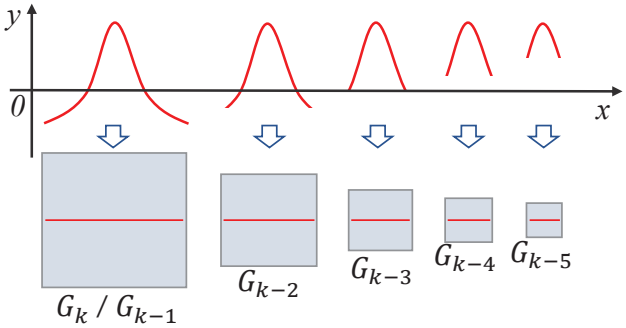

Figure 1. The discrete weight templates and corresponding 2D Gaussion surfaces (for simplicity, we only show the single axis values and corresponding 1D Gaussion curve, denoted by red lines). The values of templates are obtained from the 2D Gaussion surface by discrete sampling.

Next, the corresponding center patches with different sizes are cut from each correlation filter. The size of each correlation filter and the corresponding size of cut correlation filters are summarized in Table 1.

In the *CFF & MFF section* of the original paper, we define the discrete weight template G_k for the k-th scale filter and u-th cut patch template G_{k-u} from G_k, which are devised for summing the response maps with weights in Gaussian shape. The sum value of G_k equal to zero, the shape of the template G_k is the same as the 2D Gaussian shape, and the size of G_k is the same with the k-th scale filter. Fig. 1 gives the Gaussian shape and corresponding cut filter patch template G_{k-u}. For simplicity, we only show the single axis values and corresponding 1D Gaussion curve, denoted by red lines. The height value a and deviation σ of the 2D gaussian function $ae^{-\frac{(x-\mu_x)}{2\sigma^2}-\frac{(y-\mu_y)}{2\sigma^2}}$ are 1 and 5, respectively. More details about those weight templates, please refer to the source code in the "CodeAppendix" file.

For clear description, the definition of Eqn. (2) is de-

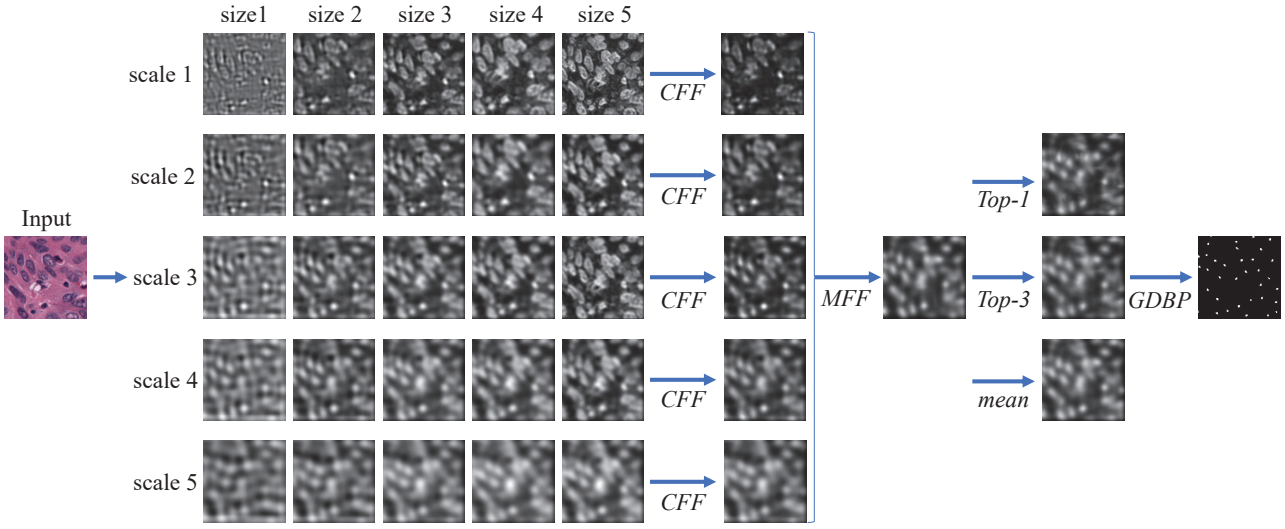

Figure 2. The progressive response maps after CFF, MFF, Top-H/mean, and GDBP.

scribed again as follows:

$$S^n(i,j) = \mathbb{A}^T\{G_1 \odot S_1^n[i,j], .., G_k \odot S_k^n[i,j]\}, \quad (2)$$
$$S_k^n(i,j) = \mathbb{A}^T\{G_{k-1} \odot S_{k-1}^n[i,j], .., G_{k-U} \odot S_{k-U}^n[i,j]\},$$

where $\mathbb{A}^T\{\}$ denotes accumulating the top T values, $S_{k-u}^n[i,j]$ and $S_k^n[i,j]$ denote the patch area of response map S_{k-u}^n and CFF map S_k^n centered at position (i,j), respectively. $S^n(i,j)$ and $S_k^n(i,j)$ denote the response value of MFF map S^n and CFF map S_k^n centered at position (i,j), respectively.

Sample Synthesis. The ambiguous boundary areas of nuclei are not involve the training of the segmentation network, which leads to poor performance on ambiguous boundary areas. So, we segment some high confidence nucleus areas and paste those areas back into the pure background images, which are generated with some high confidence background areas.

NuCLS [1] dataset contains $1,716$ samples, the size of which range from (300×300) to (800×800). From Fig. 3, we can see that NuCLS [1] dataset contains mixed annotations, part of which are segmentation masks and rest part are bounding boxes. The train, validation and testing dataset contain 1000, 361 and 355 samples in the experiment. Due to the lack of segmentation masks, existing supervised segmentation methods (Mahmood [7], MedT [11], FCN-based [8], U-Net [10] and Mask R-CNN [3]) cannot be trained on NuCLS [1] dataset. So, we only give the detection results of existing methods on NuCLS [1] dataset in Section C.

For the segmentation network Unet [10], hyper-parameters are set as follows: batch size $= 8$, learning rate $= 0.0001$, weight decay $= 1e-8$, and momentum $= 0.9$.

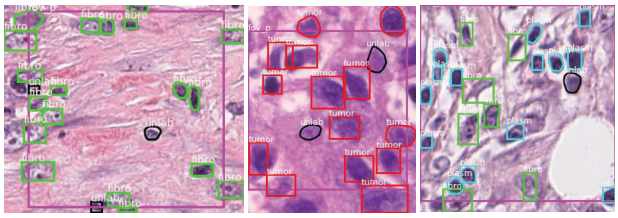

Figure 3. Samples and annotations of NuCLS [1] dataset. Annotations are mixed segmentation masks and bounding boxes.

B. Visual Results of Different Filters

In the detection branch, Cut Filter Fusion (CFF) and Multi-scale Filter Fusion (MFF) are devised for handling crowed nuclei and different scale nuclei. The fusion of top H response maps of multiple filters obtains the optimal responses, which eliminates the disturbance of low response values of some inappropriate correlation templates. Gradient Direction Based Postprocessing (GDBP) is proposed to erase the inaccurate response using gradients of each position in all directions, which can effectively locate the center position of nuclei.

For an input image, Fig. 2 shows the progressive detection response maps after CFF, MFF, Top-H, and GBDP. 'size 1' denotes the size of original filter and 'size 2-5' denote the size of cut filters. We can see that small and large scale filters have high response values for small and large nuclei, respectively. For the cut filters, the original filters ('size 1') have high response values in the center of nuclei. The cut filters ('size 3-5') have high response values for the indigo color of nuclei, which leads to the result that nuclei have high response values and background have low response values. CFF combines the center responses

	Type	*Unsupervised*	*fully-supervised*		*Ours*
	Method	**CAE [5]**	**SSAE [13]**	**Mask R-CNN [3]**	**MCF**
NuCLS	**Precision**	52.53 / 50.43	66.71 / 64.36	70.36 / 66.91	61.28 / 58.76
	Recall	68.45 / 65.99	64.02 / 60.27	75.80 / 71.52	74.75 / 69.08
	F-score	59.44 / 57.17	65.34 / 62.25	72.98 / 69.14	66.86 / 63.50

Table 2. The detection results of different methods. 'score 1 / score 2' denotes the segmentation results of original testing samples of NuCLS [1] and newly collected cancerous samples.

detected by large size filters ('size 1-2') and responses of nuclei detected by small size filters ('size 3-5'). The cut filters have high response for nuclei according nuclei color features, which can detect the crowed nuclei effectively. Meanwhile, MFF and Top H effectively combine the response values of appropriate scales and optimal response values of different filters. GDBP effectively locate the position of nucleus centers.

C. Results on NuCLS [1] Dataset

In this section, we give the detection results of different methods on NuCLS [1] dataset, which is shown in Table 2. Meanwhile, the detection results of cancerous samples are also given. 'score 1 / score 2' denotes the segmentation results of original testing samples and newly collected cancerous samples. In the experiment, there are 16 nucleus patch templates for NuCLS [1] dataset. From Table 2, the fully-supervised method Mask R-CNN [3] achieves the best performance, and the unsupervised method CAE [5] achieves the worst detection results. With only 16 nuclei patches as supervision, Recall score of MCF drops by about 1% on NuCLS [1] dataset compared with fully supervised methods, which demonstrates that MCF can effectively detect all the correct nuclei. What's more, other scores of MCF only drop by about 3% on the cancerous samples compared with results on normal samples, which verifies the excellent robustness of MCF again.

D. Segmentation and Detection Visual Results

More segmentation and detection visual results of different methods on TNBC [9], MICCAI18 [6] and cancerous samples are given in Fig. 4 & 5. The first six rows are the normal samples from TNBC [9] and MICCAI18 [6], respectively. The last six rows are cancerous samples from the newly collected cancerous samples. From Fig. 4, we can see that most of the fully supervised methods (MedT [11], FCN-based [8], U-Net [10] and Mask R-CNN [3]) achieve similar and promising results. The unsupervised method Hou [4] also achieves promising segmentation results on the normal samples but poor segmentation results on cancerous samples. On the contrary, MCF still achieves promising results on cancerous samples, which verifies the excellent robustness of MCF on cancerous samples. Meanwhile, more visual detection results are given in Fig. 5. Mask R-CNN [3] achieves more accurate results than other methods. Compared with the unsupervised method CAE [5], MCF has better detection peformance. For cancerous samples, MCF still achieves not bad detection results on cancerous samples, which demonstrates the excellent robustness of MCF on degenerated samples again.

References

[1] Mohamed Amgad et al. Nucls: A scalable crowdsourcing, deep learning approach and dataset for nucleus classification, localization and segmentation. *Journal of Medical Internet Research*, 2021. 1, 2, 3

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. 1

[3] Kaiming He, Georgia Gkioxari, Piotr Dollr, and Ross Girshick. Mask r-cnn. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017. 2, 3, 4, 5

[4] Le Hou, Ayush Agarwal, Dimitris Samaras, Tahsin M. Kurc, and Joel H. Saltz. Robust histopathology image analysis: To label or to synthesize? In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019. 3, 4

[5] Le Hou, Vu Nguyen, Ariel B. Kanevsky, Dimitris Samaras, Tahsin M. Kurc, Tianhao Zhao, Rajarsi R. Gupta, Yi Gao, Wenjin Chen, and David and Foran. Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images. *Pattern recognition*, 2019. 3, 5

[6] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, pages 1–1, 2017. 3

[7] Faisal Mahmood, Daniel Borders, Richard J. Chen, Gregory N. Mckay, and Nicholas J. Durr. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Transactions on Medical Imaging*, PP(99):1–1, 2019. 2, 4

[8] Peter Naylor, Marick La, Fabien Reyal, and Thomas Walter. Nuclei segmentation in histopathology images using deep neural networks. In *IEEE International Symposium on Biomedical Imaging*, 2017. 2, 3, 4

[9] Peter, Naylor, Marick, La, Fabien, Reyal, Thomas, and Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE Transactions on Medical Imaging*, 2018. 3

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015. 2, 3, 4

[11] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer: Gated axial-attention for medical image segmentation. *arXiv preprint arXiv:2102.10662*, 2021. 2, 3, 4

[12] J. Van, de Weijer, C. Schmid, J. Verbeek, and D. Larlus. Learning color names for real-world applications. *IEEE*

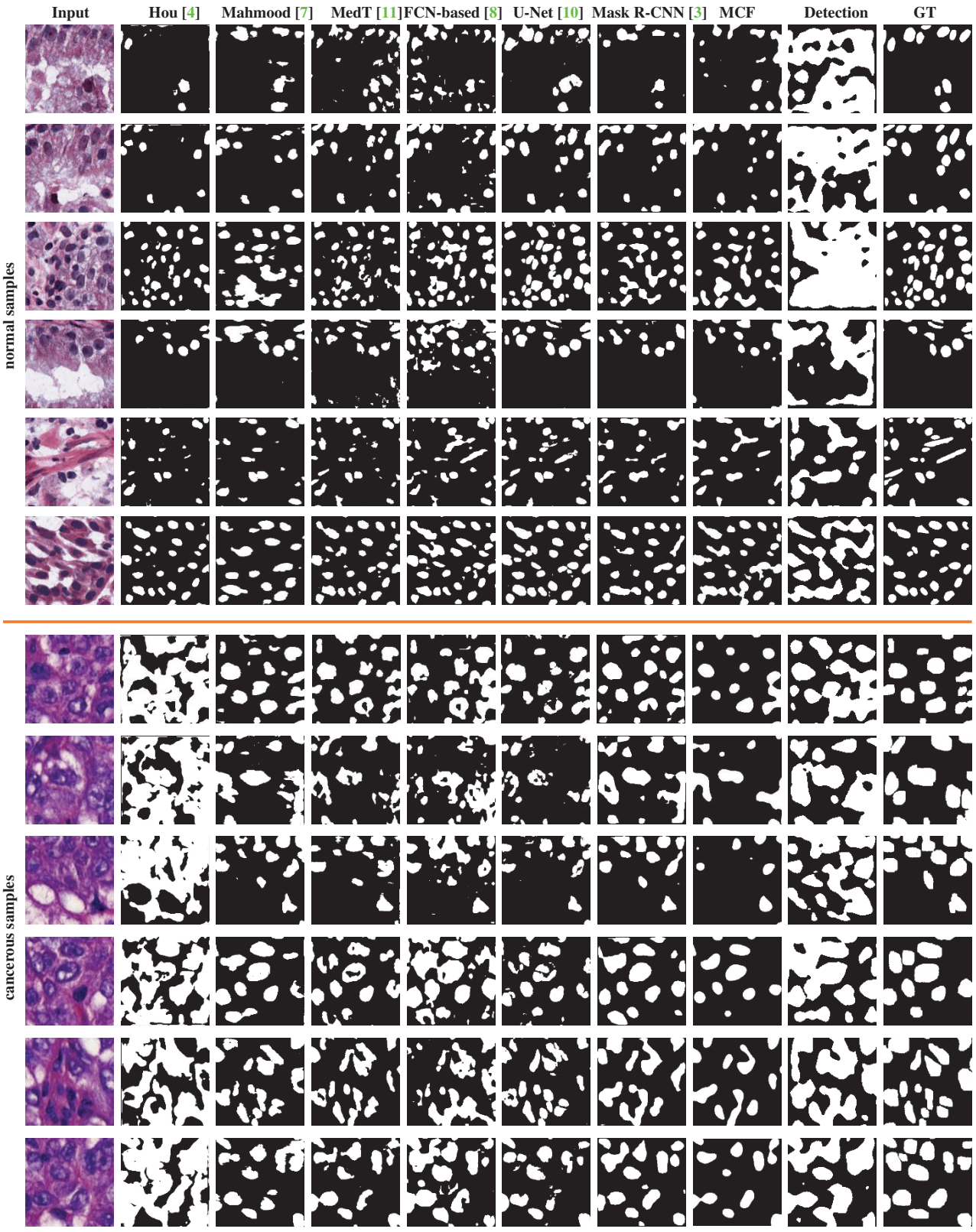

Figure 4. The visualized segmentation results of normal testing samples and cancerous samples.

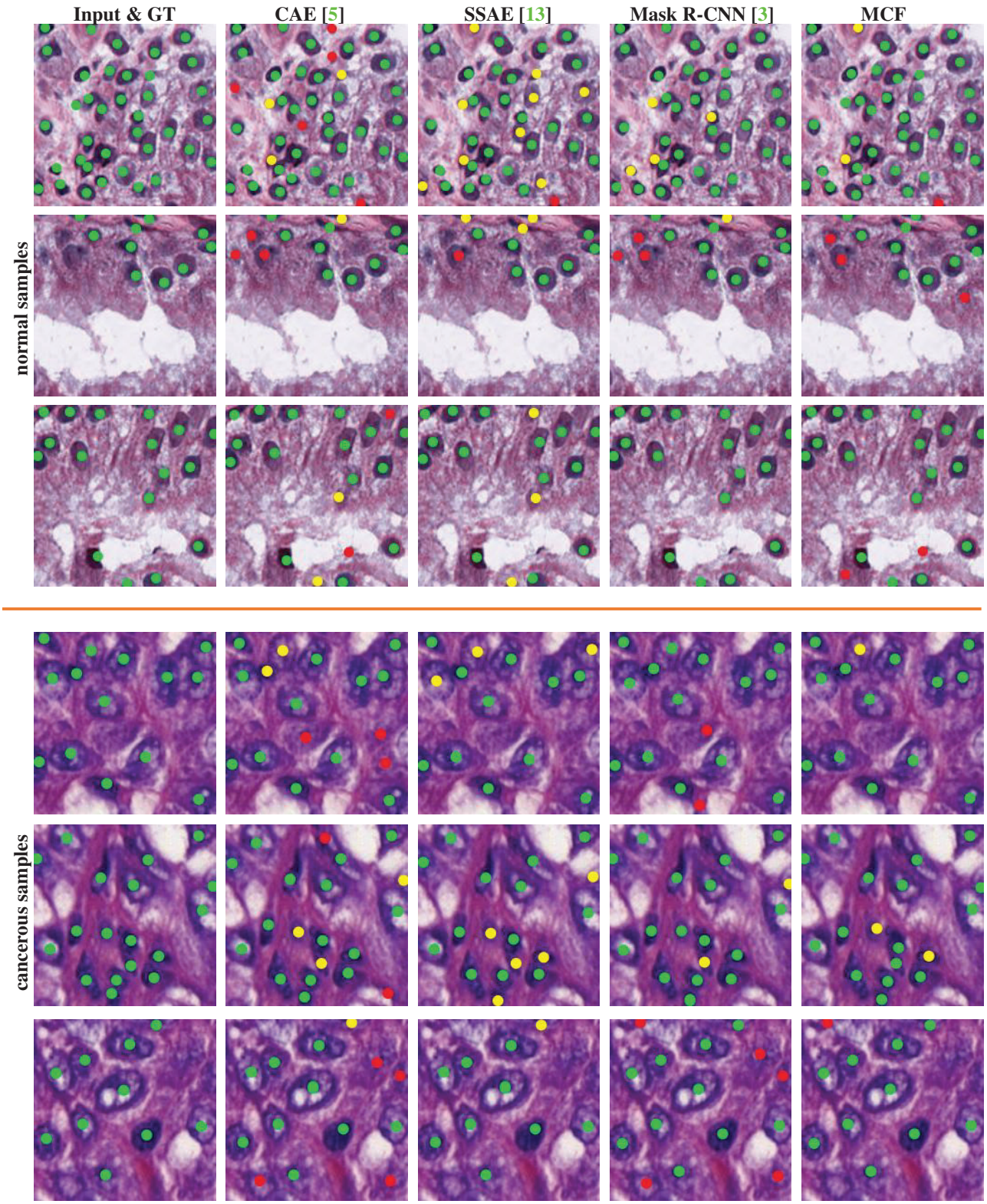

Figure 5. The detection visual results of different methods on normal samples and cancerous samples. True Positives (TP), False Positives (FP), and False Negatives (FN) are marked with green, yellow, and red color points.

transactions on image processing : a publication of the IEEE Signal Processing Society, 18(7):1512–23, 2009. 1

[13] Jun Xu, Lei Xiang, Renlong Hang, and Jianzhong Wu. Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images. In *IEEE International Symposium on Biomedical Imaging*, 2014. 3, 5