

Sampling Based Tumor Recognition in Whole-slide Histology Image with Deep Learning Approaches

Yiqing Shen and Jing Ke 

Abstract—Histopathological identification of tumor tissue is one of the routine pathological diagnoses for pathologists. Recently, computational pathology has been successfully interpreted by a variety of deep learning-based applications. Nevertheless, the high-efficient and spatial-correlated processing of individual patches have always attracted attention in whole-slide image (WSI) analysis. In this paper, we propose a high-throughput system to detect tumor regions in colorectal cancer histology slides precisely. We train a deep convolutional neural network (CNN) model and design a Monte Carlo (MC) adaptive sampling method to estimate the most representative patches in a WSI. Two conditional random field (CRF) models are incorporated, namely the correction CRF and the prediction CRF are integrated for spatial dependencies of patches. We use three datasets of colorectal cancer from The Cancer Genome Atlas (TCGA) to evaluate the performance of the system. The overall diagnostic time can be reduced from 56.7% to 71.7% on different slides in the tumor location task, with an increase in classification accuracy.

Index Terms—Tumor location, Colorectal cancer, Performance acceleration, Deep learning, Conditional random field.

1 INTRODUCTION

C OLORECTAL cancer is one of the major causes of morbidity and mortality worldwide, however, when discovered at an early stage, it is highly treatable. As the number of specimens increases every year, there is an urgent demand for computer-aided approaches to assist the diagnoses [1]. Computational pathology utilizes mathematical models to generate diagnostic inferences and presents clinical knowledge to patients. In light of the development of GPU-based parallel computing and various deep learning algorithms over the last decade, computational pathologies are taking place of pathologists in the review of digital screening and provide prompt interpretation in cancer care [2], [3], [4], [5], [6], [7]. Although other medical disciplines, such as radiology, possess a long history of research and clinical applications, it is until very recent years that digital pathology [8], [9] emerged when glass slides could be scanned into digital whole slide images (WSIs) with high resolution. Computational pathology has to inevitably face additional challenges due to its characteristic of huge size, like the tumor tissue microarray samples. Currently, a WSI has to be split into numerous patches in the most competitive high performance computing (HPC) environments [10], where each and every patch is processed individually as an input image [11], [12], [13], [14], and the computational time increases when overlapping is present. Consequently, two major limitations of the patch-based computing method-

ologies, namely expensive computational cost and being unaware of spatial correlation, are widely noticed by researchers. Only a few pieces of research in previous work have been conducted to address the former issue, such as an adaptive gradient-based sampling method is proposed to separate invasive and non-invasive tissue of breast cancer based on patch-wise classification [15], a resolution adaptive deep hierarchical learning scheme is applied to nuclear segmentation to save computational time [16], a semantic segmentation method is proposed to adaptively use image features with different magnifications [17], a model is proposed to identify some of the diagnostically relevant regions of interest by following a parameterized policy [18] and a recurrent visual attention model is designed attending to the most discriminative regions in WSI [19]. For the latter, end-to-end network architectures are proposed to incorporate spatial correlations [11], [20], [21], yet with limited patterns of neighboring patches and higher computational cost. Without dimensionality reduction, it is short of efficient approaches to locate tumor regions, and in particular, for negative cases or slides with only a small proportion of tumor [22], [23]. The multi-gigabyte nature of whole slide images requires a different methodological approach than traditional computer vision problems.

In this paper, we design a novel tumor detection system that can locate regions of interest (ROIs) promptly and accurately with patch-based convolutional neural networks (CNNs). We propose a high-throughput system to accelerate the tumor detection task in colorectal cancer histopathology images. The acceleration is achieved via the reduction of image patches to be processed by CNN. In this system, a Monte Carlo (MC) sampling based tumor distribution approximation algorithm is designed to sample regions of interest. It generates a slide-level tumor likelihood distribu-

- Y. Shen is with the School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai, China, 200240. E-mail: shenyq@sjtu.edu.cn.
- J. Ke is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, 200240, and Biren Research, Shanghai, China, 201114. J. Ke is also with the School of Computer Science and Engineering, University of New South Wales, Sydney, Australia. Correspondence to J. Ke. E-mail: kejing@sjtu.edu.cn.

Manuscript received April 19, 2005; revised August 26, 2015.

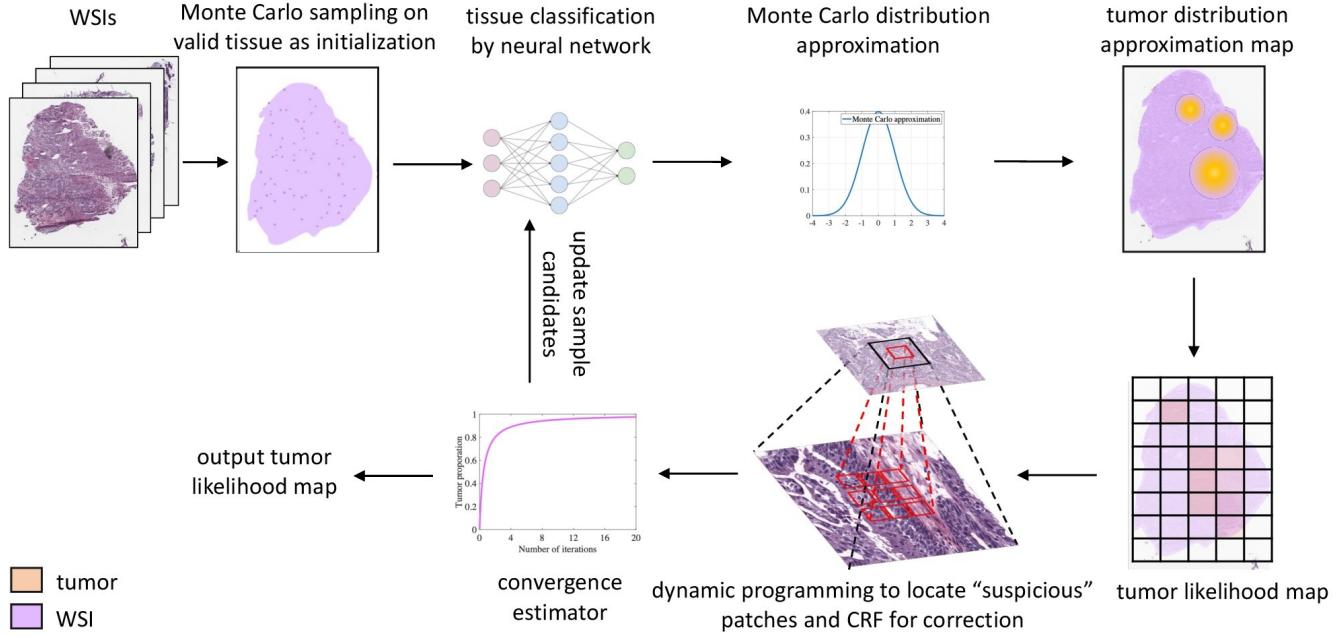


Fig. 1. An overall workflow of the proposed dynamic programming based fast tumor detector (DP-FTD).

tion map, by which only a small proportion of patches are required to be processed with CNNs. We propose two models to incorporate spatial context for a precise classification of patches, with deep neural networks and dynamic programming, namely the dynamic programming based fast tumor detector (DP-FTD) and deformable conditional random field-based fast tumor detector (DCRF-FTD). The former uses conditional random field neural network (NCRF) [11] for dynamic correlation with progressive MC sampling. The latter employs DCRF [24] based network as the classifier for an adaptively enlarged receptive field in morphological feature extraction. We achieve obvious performance speed-up with both approaches, while the second approach is higher efficient at the inference stage at the cost of longer training time. These proposed models can work with any CNN structures in histopathology images for performance improvement. The computation acceleration is even more obvious for WSIs which contain only a small proportion of tumor tissues.

In the proposed systems, after a patch is selected as a sampling candidate, it will undergo deep learning processing. The sampling strategy is to generate high-possibility tumor regions in a WSI. Empirically, we prove that not every patch in a WSI needs to be processed by a neural network for tumor identification, given the nature of cancer histology.

The main contributions are summarized as follows.

- For an accelerated and precise tumor region prediction for histopathology, we propose a Dynamic Programming based Fast Tumor Detector (DP-FTD). It incorporates a conditional random field neural network (NCRF) for dynamic correction for patch-level false positives and negatives. This novel model can learn spatial correlations automatically to effectively improve the WSI-level prediction accuracy. Compared with previous work [25], it effectively

avoids the adoption of domain knowledge and the heavy workload in hyper-parameters tuning. Empirically, it shows a higher prediction outcome.

- To further improve inference efficiency, we propose another Deformable Conditional Random Field based Fast Tumor Detector (DCRF-FTD), where a DCRF [24] framework is employed as the classification backbone. Instead of an additional procedure to revise false patch-level predictions, we avoid the false predictions in Monte Carlo sampling by the incorporation of spatial information. The adoption of DCRF targets both classification accuracy improvement and performance acceleration.

In the two proposed models, DCRF-FTD is more efficient in the inference stage, at the cost of more training time. Likewise, DP-FTD is faster in the training stage, but slower at the inference stage. Compared to existing methods, both models can achieve improvement in classification accuracy and performance acceleration.

The rest of this paper is structured as follows. In Sec. 2, we propose our algorithms and methodologies for the efficient tumor location model. Sec. 3 presents the empirical results based on the clinical datasets in large patient cohorts from The Cancer Genome Atlas (TCGA), and Sec. 4 will conclude with the advantages and limitations of the proposed models.

2 METHODOLOGY

In this section, we present the methodology in the design of the tumor detection system. In Sec. 2.1, we employ a Monte Carlo (MC) adaptively sampling to localize patches with the highest likelihood of tumor tissues. Dynamic programming (DP) algorithm is applied to correct the false predictions in the dynamic programming based fast tumor detector (DP-FTD). In Sec. 2.2, a deformable conditional random

field based fast tumor detector (DCRF-FTD) is proposed to further incorporate spatial correlation in the inference stage for an improved prediction performance. Patches with the highest degree of uncertainty in the MC approximation with being sampled and processed by CNN. In Sec. 2.3, we propose the convergence criterion to our models in the sampling procedure of the inference stage.

2.1 Dynamic Programming based Fast Tumor Detector

2.1.1 Monte Carlo Distribution Approximation

In this acceleration algorithm, we aim to sample a minimum proportion of patches, characterized with the highest likelihood of tumor tissues. The sampling distribution map is progressively updated with the iterative outcome of tumor predictions from a deep convolutional neural network.

In the estimation of tumor distribution, we denote the ground-truth tumor probability distribution as $P(L = l|\mathbf{x})$, where $l \in \mathcal{L}$ is a label from the set \mathcal{L} and \mathbf{x} is a patch in WSI. In the tumor detection task, $\mathcal{L} = \{\text{tumor}, \text{nontumor}\}$, and we only focus on the tumor distribution, i.e., the case of $L = \text{tumor}$. We utilize a Monte Carlo sampling method to efficiently sort out tumor areas i.e. $P(L = \text{tumor}|\mathbf{x})$ by estimating patch-wise tumor possibility distribution $p(\mathbf{x})$. With the prediction outcome from a neural network $\{p(\mathbf{x}_i)\}_{i=1}^N$ of N patches $\{\mathbf{x}_i\}_{i=1}^N$, which are selected randomly to as the initializations. Then, target distribution $p(\mathbf{x})$ can be estimated with the Gaussian as follows [26],

$$\hat{p}(\mathbf{x}) = \sum_{i=1}^N \delta_{\mathbf{x}_i}^\sigma(\mathbf{x}) \cdot w_i. \quad (1)$$

We denote the Gaussian function centered at \mathbf{x}_i as $\delta_{\mathbf{x}_i}^\sigma(\mathbf{x})$, with a restriction condition that $\sigma \ll 1$. It functions as a tumor possibility factor, which is formulated by,

$$\delta_{\mathbf{x}_i}^\sigma(\mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{-\|\mathbf{x}-\mathbf{x}_i\|^2}{2\sigma^2}}. \quad (2)$$

Given the possibility $\{p(\mathbf{x}_i)\}_{i=1}^N$ at $\{\mathbf{x}_i\}_{i=1}^N$, then weights $\{w_i\}_{i=1}^N$ for the j -th factors are valued by the following linear equations,

$$\sum_{i=1}^N \delta_{\mathbf{x}_i}^\sigma(\mathbf{x}_j) \cdot w_i = p(\mathbf{x}_j), \quad (j = 1, \dots, N). \quad (3)$$

As $(\delta_{\mathbf{x}_i}^\sigma(\mathbf{x}_j))_{ij}$ is a symmetric matrix and its determinant $|\delta_{\mathbf{x}_i}^\sigma(\mathbf{x}_j)| \neq 0$, we come to an unique solution to Eq. (3) for weight $\{w_i\}_{i=1}^N$ in Eq. (1). However, a large value N will add to the computational complexity to the solving equations. Therefore, given $N > C_M$, where C_M is the threshold for efficiently solving high dimensional linear equations, we will estimate the tumor distribution with an optimized approximation, denoted as \hat{p}_k , and the subscript indicates the approximation is updated in the k -th iteration:

$$\hat{p}_k(\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \int_{\Omega} r(\mathbf{x} - \mathbf{t}) \cdot \mathbb{I}_{\{\mathbf{x}_i\}}^{\{p(\mathbf{x}_i)\}}(\mathbf{t}) dt, \quad (4)$$

where Ω denotes the entire region in a WSI, and $\mathbb{I}_{\{\mathbf{x}_i\}}^{\{p(\mathbf{x}_i)\}}(\mathbf{t})$ is an indicator function, defined as:

$$\mathbb{I}_{\{\mathbf{x}_i\}}^{\{p(\mathbf{x}_i)\}}(\mathbf{t}) = \begin{cases} p(\mathbf{x}_i), \mathbf{t} \in \{\mathbf{x}_i | i = 1, 2, \dots, N\}, \\ 0, \mathbf{t} \notin \{\mathbf{x}_i | i = 1, 2, \dots, N\}, \end{cases} \quad (5)$$

and $r(\mathbf{x} - \mathbf{t}) = \frac{1}{\|\mathbf{x} - \mathbf{t}\|}$ is the spatial distance kernel, $Z(\mathbf{x}) = \int_{\Omega} r(\mathbf{x} - \mathbf{t}) \cdot \mathbb{I}_{\{\mathbf{x}_i\}}^{\{p(\mathbf{x}_i)\}}(\mathbf{t}) dt$ is a normalization constant with respect to \mathbf{p} to make $\hat{p}_k(\mathbf{x})$ a valid distribution function. In particular, we set $\hat{p}(\mathbf{x}_i) = p(\mathbf{x}_i)$ for patches that have already been sampled in the previous iterations. We exclude the sampled patch \mathbf{x}_i when it falls out the valid area of tissues within a WSI, and this patch is discarded from processing with CNN. In particular, we set $\hat{p}(\mathbf{x}_i) = 0$.

2.1.2 Monte Carlo Adaptive Sampling Strategy

We design the MC adaptive sampling strategy as follows. First, a number of n_0 patches from a slide are randomly sampled as initialization, denoted as $\{\mathbf{x}_i^{(0)}\}_{i=1}^{n_0}$. Tumor possibility of these patches are subsequently predicted by a fine-tuned neural network where the outcome $\{p(\mathbf{x}_i^{(0)})\}$ is the possibility of tumor at the sampled patches. Afterwards, applying the Monte Carlo distribution approximation in Eq. (1) or (4), which is determined by the range of N , we reach an initialized approximated tumor distribution, denoted as $\hat{p}_0(\mathbf{x})$. When $k - 1$ iterations have been completed, with tumor distribution approximation $\hat{p}_{k-1}(\mathbf{x})$ in the $k - 1$ iteration, a total number of n_k patches $\{\mathbf{x}_i^{(k)}\}_{i=1}^{n_k}$ have been sampled progressively with the highest likelihood to obtain tumor possibilities $\{p(\mathbf{x}_i^{(k)})\}_{i=1}^{n_k}$ from neural network.

Iteratively, the approximation distribution $\hat{p}_k(\mathbf{x})$ will be updated via Eq. (1) or (4) to coincide with the restriction that $\hat{p}_k(\mathbf{x}_i^{(j)}) = p(\mathbf{x}_i^{(j)})$ for all the valid i and j , where the superscript denotes that the patch is sampled in the j -th iteration. As a result, patches with highest likelihood remained in the WSI in the current iteration are sampled, and then followed by the processing of the neural network.

2.1.3 Dynamic Programming Based False Correction

In the analysis of tumor distribution, we are aware that neighboring patches often have spatial correlations, where an inconsistent CNN prediction may imply a potential false positive or false negative of tumor tissue. To improve the prediction and guide the adaptive sampling, we progressively correct the outcome from the tumor classifier. The spatial context from neighboring patches is integrated and a conditional random field (CRF) model is incorporated into the CNN design.

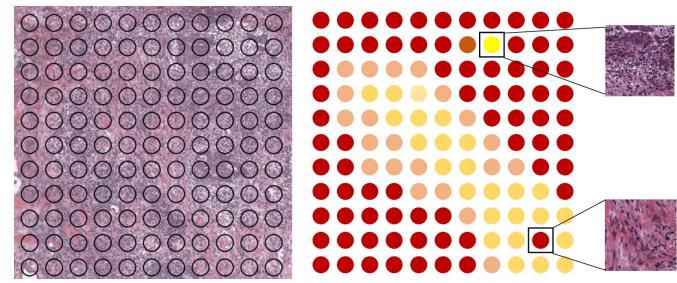


Fig. 2. The classification of tissue patches with "suspicious" patches marked by black boxes. They have higher likelihood of being false predictions.

Let $\mathcal{S} = \{\mathbf{x}_i^{(j)}\}$ denote all the sampled regions, $N_1(\mathbf{x})$ denotes the neighborhood of \mathbf{x} within a radius value of 1.

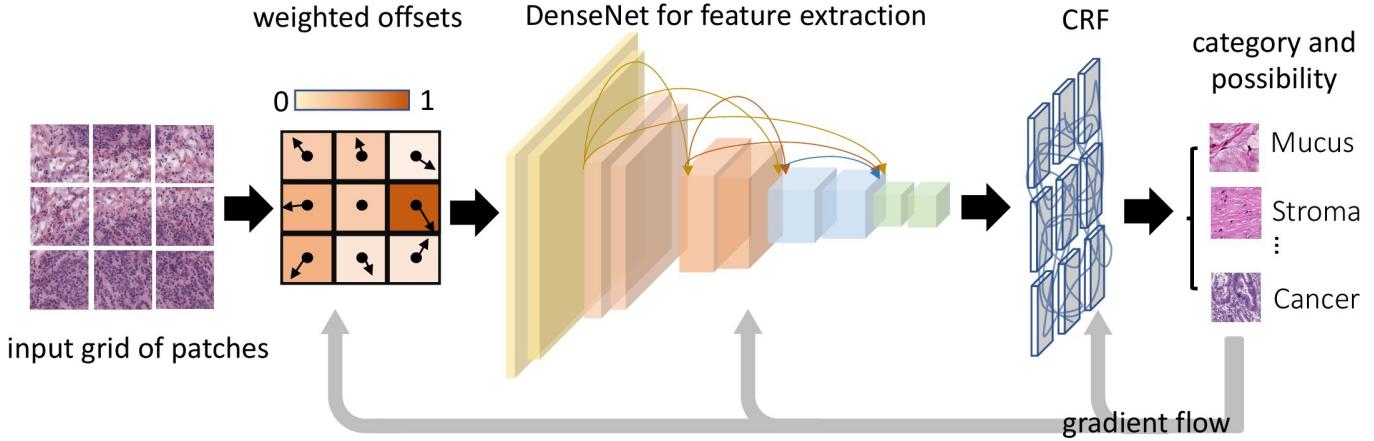


Fig. 3. Overall architecture of the deformable conditional random field [24] based network for the second model. It consists of three parts i.e., learnt weighted offsets, CNN feature extractor and conditional random field [27] block for spatial context. The model employs DenseNet [28] as the feature extractor.

$N_1^*(\mathbf{x}) = N_1(\mathbf{x}) \cap \mathcal{S}$ is the set of sampled patches in $N_1(\mathbf{x})$. In the k -th iteration, sampled patches $\mathbf{x}_0 \in \mathcal{S}$ satisfying the following equation is voted to be a potential false negative or false positive of a tumor patch:

$$|p(\mathbf{x}_0) - \sum_{\mathbf{x}_0 \in N_1^*(\mathbf{x}_0)} p(\mathbf{x}_0^*) / \#(N_1^*(\mathbf{x}_0))| > \chi(k), \quad (6)$$

where $\#(S)$ is the number of elements in the set S . $\chi(k) \in [0, 1]$ is the criterion function to estimate the likelihood of a wrong prediction by neural network. A NCRF is designed to further identify these “suspicious” patches (shown in Fig. 2), which is pre-trained on the grids of patches to learn how to extract spatial features from neighboring patches. However, the NCRF is not used as the tumor detector in our system, due to its extra $5 \times$ to $10 \times$ inference time more than a regular CNN. Consequently, we only apply NCRF to a small number of patches voted to be spatial inconsistent. Its prediction outcome, denoted as $p'(\mathbf{x}_0)$, is added to the weighted average for the false correction procedure:

$$p^*(\mathbf{x}_0) = \theta_0 \cdot p(\mathbf{x}_0) + \theta_1 \cdot \sum_{\mathbf{x}_0 \in N_1^*(\mathbf{x}_0)} p(\mathbf{x}_0^*) / \#(N_1^*(\mathbf{x}_0)) + \theta_2 \cdot p'(\mathbf{x}_0)$$

To follow the convergence criterion designed in Eq. (10), we set the weights θ_i as follows.

$$\begin{cases} \theta_0 = \max\{d(\partial\mathfrak{D}_{k-1}, \partial\mathfrak{D}_{k-2}), d(\partial\mathfrak{D}_k, \partial\mathfrak{D}_{k-1})\} / \varepsilon(\partial\mathfrak{D}_k) \\ \theta_1 = \theta_2 = \frac{1-\theta_0}{2}. \end{cases}$$

where $d(\cdot, \cdot)$ is a metric of closed curves in WSIs, and $\varepsilon(\mathfrak{D})$ determines the convergence speed. All these notations will be elaborated in Sec. 2.3. We will reach $\theta_0 < 1$ within a couple of iterations. The overall inference processing of the proposed DP-FTD is summarized in Alg. 1. We will discuss the implementation convergence criterion in Sec. 2.3.

The overall architecture of DP-FTD is depicted in Fig. 1.

2.2 Deformable Conditional Random Field Network Based Fast Tumor Detector

2.2.1 Deformable Conditional Random Filed Network

To take account of spatial dependencies in the inference stage of deep learning, instead of post-processing, we propose a deformable conditional random field network based

Algorithm 1 Dynamic Programming Based Fast Tumor Detector (DP-FTD)

Input: Ω whole slide image
Output: $\tilde{p}(\mathbf{x})$ tumor distribution estimation
 $\{\mathbf{x}_i^{(0)}\}_{i=1}^{n_0} \leftarrow$ randomly sampled n_0 patches from Ω
 $\hat{p}_0(\mathbf{x}) \leftarrow$ Monte Carlo distribution approximation
Set $\mathcal{S}_0 = \{\mathbf{x}_i^{(0)}\}_{i=1}^{n_0} \%$ \mathcal{S} denotes already sampled patches
while not convergent **do**
 $\{\mathbf{x}_i^{(k)}\}_{i=1}^{n_k} \leftarrow$ sample n_k from $\Omega / \mathcal{S}_{k-1}$ based on $\hat{p}_{k-1}(\mathbf{x})$
 $\{p(\mathbf{x}_i^{(k)})\}_{i=1}^{n_k} \leftarrow$ attain predictions on samples
Set $\mathcal{S}_k = \mathcal{S}_{k-1} \cup \{\mathbf{x}_i^{(k)}\}_{i=1}^{n_k} \%$ update sampled points
 $\hat{p}_k(\mathbf{x}) \leftarrow$ updated Monte Carlo approximation
CRF based dynamic correction for $\hat{p}_k(\mathbf{x})$
 $k \leftarrow k + 1 \%$ count the number of iterations
end while
 $\tilde{p}(\mathbf{x}) = \hat{p}_k(\mathbf{x})$

Algorithm 2 Deformable Conditional Random Field Based Fast Tumor Detector (DCRF-FTD)

Input: Ω whole slide image
Output: $\tilde{p}(\mathbf{x})$ tumor distribution estimation
 $\{\mathbf{x}_i^{(0)}\}_{i=1}^{n_0} \leftarrow$ randomly sampled n_0 patches from Ω
 $\hat{p}_0(\mathbf{x}) \leftarrow$ Monte Carlo distribution approximation
Set $\mathcal{S}_0 = \{\mathbf{x}_i^{(0)}\}_{i=1}^{n_0} \%$ \mathcal{S} denotes already sampled patches
 $q_0(\mathbf{x}) \leftarrow$ attain query map on $\{p(\mathbf{x}_i^{(0)})\}_{i=1}^{n_0}$
while not convergent **do**
 $\{\mathbf{x}_i^{(k)}\}_{i=1}^{n_k} \leftarrow$ randomly sampled n_k patches from $\arg \max_{\mathbf{x}} q_{k-1}(\mathbf{x}) \}_{i=1}^{2n_k} \%$ quasi MC sampling
 $\{p(\mathbf{x}_i^{(k)})\}_{i=1}^{n_k} \leftarrow$ attain predictions on samples
Set $\mathcal{S}_k = \mathcal{S}_{k-1} \cup \{\mathbf{x}_i^{(k)}\}_{i=1}^{n_k} \%$ update sampled points
 $q_k(\mathbf{x}) \leftarrow$ update query map with $\{p(\mathbf{x}_i^{(k)})\}_{i=1}^{n_k}$
 $k \leftarrow k + 1 \%$ count the number of iterations
end while
 $\tilde{p}(\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \int_{\Omega} \hat{p}(\mathbf{t}) \cdot k(\mathbf{x}, \mathbf{t}) \cdot \mathbb{I}_S(\mathbf{t}) d\mathbf{t}$

fast tumor detector (DCRF-FTD). In this approach, we define regions of interest to be those where the classification was uncertain. We employ a state-of-the-art deformable conditional random field neural network (DCRF) [24] as the classifier to obtain data dependencies between neighboring patches. The incorporation of spatial context makes a better prediction by CNN. DCRF is a state-of-the-art end-to-end tumor classification model without hand-crafted features, which can learn the morphological features adaptively and automatically with a larger range of field [24].

As shown in previous work [24], DCRF has demonstrated its stronger capability for feature extraction from spatial-correlated patches. The DCRF is made up of three blocks, including an offset-adjust block to learn offset of the most representative patches, a CNN feature extractor of DenseNet [28], and a CRF block to take into account spatial patch dependencies. As DenseNet outperforms other prototypical CNN architecture when implemented with DCRF shown in previous work [24]. More experimental results are elaborated in Sec. 3.1.5. The model incorporates a trainable deformable offset block with a neural conditional random field deep learning framework [11] as the patch-level prediction model. The overall architecture of DCRF is depicted in Fig. 3.

2.2.2 Quasi Monte Carlo Sampling Strategy

As the DCRF algorithm [24] can effectively incorporate spatial information as a classifier, we abandon the dynamic correction block as shown in Fig. 1. We also alter the query map for the sampling strategy to fit this case better. The strategy of DCRF-FTD is depicted in Alg. 2, with the quasi Monte Carlo sampling to locate the region of interests.

We set the query map for quasi MC sampling in the k -th iteration as follows.

$$\begin{aligned} q_k(\mathbf{x}) &= I_{\mathcal{S}_k}(\mathbf{x}) \cdot \int_{WSI} H(\mathbf{t}) \cdot k(\mathbf{x}, \mathbf{t}) \cdot \mathbb{I}_{\mathcal{S}_k}(\mathbf{t}) d\mathbf{t} - \\ &\quad \epsilon \cdot I_{\mathcal{S}_k}(\mathbf{x}) \cdot \int_{WSI} \hat{p}_k(\mathbf{x}) \cdot k(\mathbf{x}, \mathbf{t}) \cdot \mathbb{I}_{\mathcal{S}_k}(\mathbf{t}) d\mathbf{t}, \end{aligned} \quad (7)$$

where ϵ is the penalty term, $\mathbb{I}(\cdot)$ is the indicator function, $\hat{p}_k(\cdot)$ is the Monte Carlo distribution approximation as described in Sec. 2.1, $k(\cdot, \cdot)$ is the spatial kernel, $\mathcal{S}_k = \cup_{j=0}^k \{\mathbf{x}_i^{(j)}\}_{i=1}^{n_j}$ denotes the already sampled patches up to the k -th iteration. $H(\mathbf{t})$ is the entropy at $\mathbf{t} \in \mathcal{S}_k$, defined by,

$$H(\mathbf{t}) = \sum_{l \in \mathcal{L}} \hat{p}(L = l | \mathbf{t}) \cdot \ln[\hat{p}(L = l | \mathbf{t})]. \quad (8)$$

Note that the value of $\hat{P}(L = l | \mathbf{t})$ has already been assigned at $\mathbf{t} \in \mathcal{S}_k$.

When DCRF-FTD converges, we set the output estimated tumor distribution as $\tilde{p}(\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \int_{\Omega} \hat{p}(\mathbf{t}) \cdot k(\mathbf{x}, \mathbf{t}) \cdot \mathbb{I}_{\mathcal{S}}(\mathbf{t}) d\mathbf{t}$, where $Z(\mathbf{x})$ is a normalization constant.

2.3 Convergence Criterion

As tumor proportion and distribution may vary a lot on different histopathology images, we propose a convergence criterion to effectively terminate the sampling processing. In the k iteration, we describe the high possibility tumor region \mathfrak{D}_k from WSI as follows:

$$\mathfrak{D}_k = \text{supp}\{\mathbb{I}_{\{\hat{p}_k(\mathbf{x}) \geq \mu_k\}}[\hat{p}_k(\mathbf{x})]\}, \quad (9)$$

where $\text{supp}(f)$ denotes the support of f . μ_k functions as a threshold of "high possibility" tumor region. The arrival of the convergence in the estimation of tumor boundary will stop the iterative adaptive sampling. Let $\partial\mathfrak{D}_k$ denote the estimated boundary of tumor regions, and the convergence criterion is given by:

$$\max\{d(\partial\mathfrak{D}_{k-1}, \partial\mathfrak{D}_{k-2}), d(\partial\mathfrak{D}_k, \partial\mathfrak{D}_{k-1})\} < \varepsilon(\partial\mathfrak{D}_k), \quad (10)$$

where $d(\cdot, \cdot)$ is a metric of closed curves in WSIs. Let Φ_k be a one-to-one mapping from $[0, 1]$ to the boundary of each connected components in tumor region $\partial\mathfrak{D}_k$ i.e., $\Phi_k : [0, 1] \rightarrow \partial\mathfrak{D}_k$. By our definition, we have $\lim_{x \rightarrow 1^{-}} \Phi_k(x) = \Phi_k(0)$, hence we extend Φ_k from $[0, 1)$ to $[0, 1]$ by setting $\Phi_k(0) = \Phi_k(1)$ for a clear notation, then we can define the metric based on the same definition of Φ_{k-1} as:

$$d(\partial\mathfrak{D}_k, \partial\mathfrak{D}_{k-1}) = \int_{[0, 1]} \|\Phi_k(t) - \Phi_{k-1}(t)\|_2 dt. \quad (11)$$

We write $\|\cdot\|_2$ for the discrete l_2 norm for measuring the distances between pixels. Since $\Phi_k(t)$ is the image pixel, it can be formulated as (t^i, t^j) for integers t^i, t^j , then the discrete l_2 norm can be formulated as:

$$\|(t^i, t^j)\|_2 = \sqrt{(t^i)^2 + (t^j)^2}. \quad (12)$$

We employ $\varepsilon(\mathfrak{D})$ to determine the convergence speed of the proposed sampling algorithm, which is tunable on the area of set \mathfrak{D} up to the performance requirement. For example, we value $\varepsilon(\mathfrak{D})$ with $C_1 \text{Area}(\mathfrak{D})^a$ for a fast convergence or conversely, $C_2 \exp(-b \text{Area}(\mathfrak{D}))$ for a gradual convergence. C_1 and C_2 are normalizing constants, and a, b are the threshold ratios in the growth of sample areas to the boundary of the tumor region and $\text{Area}(D)$ is the area of D i.e., counting the number of pixels in D .

We claim that in the proposed system, the convergence criterion defined in Eq. (10) is guaranteed to be fulfilled after some iterations. We sketch out the proof as follows. By our definition, for any fixed coordinate \mathbf{x} is a WSI, the high possibility tumor region $\mathfrak{D}_k(\mathbf{x})$ is monotonically increasing and bounded between 0 and 1, i.e. $\mathfrak{D}_k(\mathbf{x}) \in [0, 1]$ for $\forall k \in \mathbb{N}$. It yields the existence of a point-wise limit $\mathfrak{D}(\mathbf{x}) = \lim_{k \rightarrow \infty} \mathfrak{D}_k(\mathbf{x})$. Since the support of $\mathfrak{D}(\mathbf{x})$ is universally compact. Hence the limit \mathfrak{D} is universal. It yields that $d_k = d(\partial\mathfrak{D}_k, \partial\mathfrak{D}_{k+1})$ is a *Cauchy Sequence* i.e., for any $\eta > 0$ there must exists integer K such that for any $k \geq K$, $0 < d_k < \eta$, which completes the proof.

3 RESULTS

We evaluate the performance of the proposed methods with two indicators, namely tumor prediction accuracy and WSI processing acceleration. Both proposed methods outperform previous work in three aspects of the high-throughput detection of tumor region in WSI with a significant reduction in auto-diagnostic time, the prediction accuracy, the training time. Although DCRF-FTD takes more training time and possesses more trainable parameters, it shows a better acceleration capability in the inference stage comparing with DP-FTD. The experiment tests are performed on the state-of-the-art deep learning computing platform NVIDIA Tesla™ V100 GPU device.

3.1 Dataset Pre-processing and Classifier Training

3.1.1 Dataset and Annotations

The experiments are performed on hematoxylin and eosin (H&E) stained whole-slide images from The Cancer Genome Atlas (TCGA) for gastric (stomach) adenocarcinoma and colorectal cancer detection, including three subsets of: i) TCGA-STAD (n=432 samples), ii) TCGA-COAD (n=460 samples) iii) TCGA-READ (n=171 samples). WSIs are stored in SVS format, most of which have a resolution of $40\times$ (0.25 microns/pixel). We classified patches into 3-classes, namely i) loose non-tumor tissue, ii) dense non-tumor tissue, and iii) gastrointestinal cancer tissues. The ground truth and patch-level annotations are provided by previous works and are freely available [29], [30], [31]. A few patch examples are demonstrated in Fig. 4-A. The ground truth tumor region percentages of each WSI are profiled and shown in Fig. 4-B.

3.1.2 Evaluation Metrics

We utilize three metrics for the classification performance evaluation, namely the 1) classification accuracy, 2) area under the receiver-operating characteristic curve (AUC) and free-response receiver-operating characteristic curve (FROC) [32].

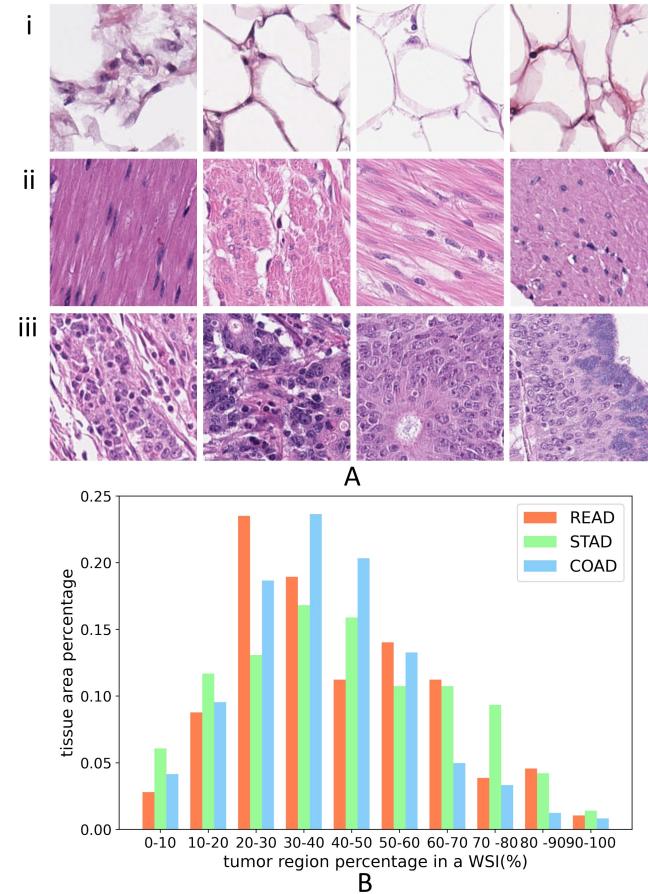


Fig. 4. A) A few patch examples of 224×224 pixels in benchmark test, i) loose non-tumor tissue ii) dense non-tumor tissue iii) gastrointestinal cancer tissues. B) Tumor region proportion marked by pathologist in a WSI from three datasets i.e., TCGA-COAD, TCGA-READ, TCGA-STAD.

- **AUC.** The receiver-operating characteristic curve (ROC) is generated by plotting true positive rate (TPR) against the false positive rate (FPR), where

$$\text{TPR} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}, \quad (13)$$

$$\text{FPR} = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}. \quad (14)$$

Then we can formulated the AUC by as follows,

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR}^{-1}(x))dx. \quad (15)$$

AUC is valued from 0 to 1. A higher value of AUC is associated with a better performance of classifier.

- **FROC.** Free-response receiver-operating characteristic curve is generated by plotting metastasis detection sensitivity against false positives. Higher FROC score is associated with better classification system.

3.1.3 Experimental Settings

In each dataset, we set 70% samples for training, 15% for validation and 15% for evaluation. We use a fixed size of 224×224 pixels as the input size of the patches in the pre-trained DCRF [24] with DenseNet [28] in DCRF-FTD and ResNet-18 [33] in DP-FTD, at the magnification of $20\times$ to retain the high resolutions. All the CNN are pre-trained on the ImageNet with 200 epochs.

We use 70% of the whole dataset as the training set, in case of overfitting in the training procedure. As shown in Fig. 5, we demonstrate the test accuracy of different scales of the training set. Each CNN is trained for 50 epochs on our training set.

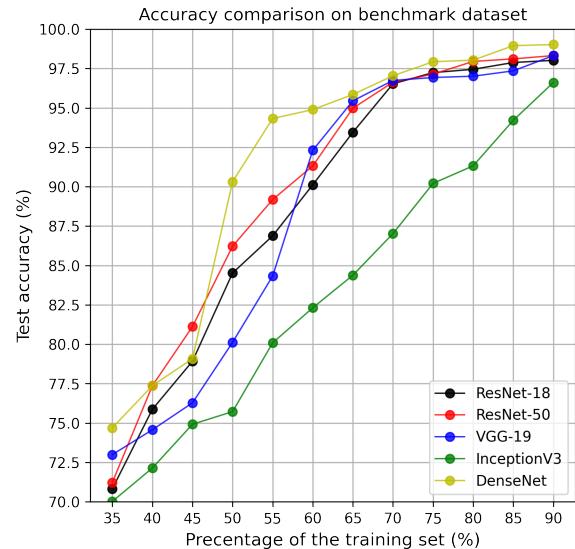


Fig. 5. Test accuracy on different scale of the training set.

To further improve the accuracy, we pre-process the images to reserve the magnification rate of $20\times$. The pre-processing procedure of splitting WSIs into patches and color normalization is conducted with Matlab R2020a, while

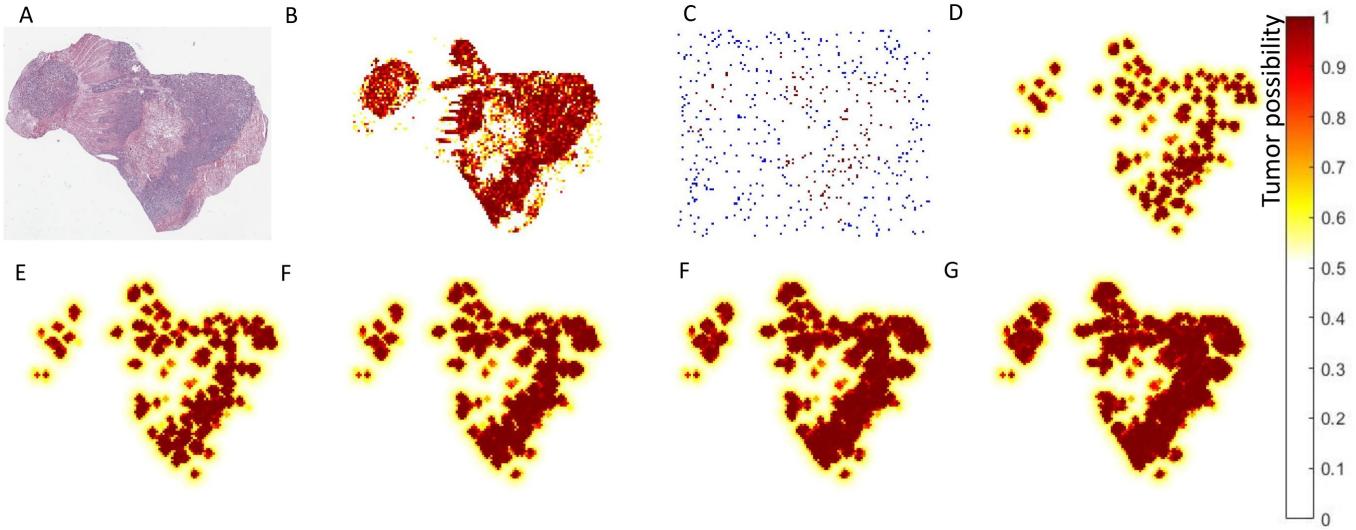


Fig. 6. An example of the proposed iterative quasi MC sampling in a WSI with DCRF-FTD. A) the original WSI, B) ground truth, C) randomly initialized patches, and D-G) the distribution approximation maps within increasing iteration by the proposed DCRF-FTD. In this example, we have achieved an accuracy of 98.90% with only 18.9% sampled patches. In comparison, regular sampling of the WSI achieves an accuracy of 97.12%.

other computation tasks are written with Python version 3.7.0. All the proposed model is implemented on Pytorch.

The neural networks for the DP-FTD and DCRF-FTD are pre-trained with a large number of images from the ImageNet dataset for 200 epochs. We fine-tune the classifier on the last couple of layers and train with 200 epochs with the transfer learning. The hyper-parameters to each neural network architecture are set as follows: *learning rate* = 1×10^{-5} for the Adam optimizer. We reduce the *learning rate* by 10% every epoch when improvement is absent on the validation set.

The hyper-parameters for the proposed methods used in this experiment are listed in Table 1.

TABLE 1
Hyper-parameters in our experiments.

parameters	model	value
σ in Eq. (2)	DP-FTD	1×10^{-2}
threshold C_M	DP-FTD	5×10^2
penalty ϵ in Eq. (7)	DCRF-FTD	2×10^{-1}
threshold μ_k in (9)	Convergence Criterion	$1 - \frac{2}{k+3}$
$\varepsilon(\mathfrak{D})$ in Eq. (10)	Convergence Criterion	$0.08 \cdot \text{Area}(\mathfrak{D})$

3.1.4 CNN Architecture for DP-FTD

In a previous works [25], [29], [34], [35], typical architectures of ResNet-18 [33], VGG-19 [36] and DenseNet [28] have demonstrated stronger capability in tumor classification. ResNet-18 [29] is outstanding in short training time and comparable classification performance, which is employed as the backbone CNN architecture in the proposed DP-FTD. The performance of our fine-tuned network is demonstrated in Table 2.

3.1.5 CNN Architecture for DCRF-FTD

DCRF is implemented on Pytorch [38]. The hyper-parameters for the DCRF were set as follows: $\beta_1 = 0.9$,

TABLE 2
The accuracy comparison between the baseline tumor classifier [29] (patch size 512×512 resized to 224×224) and our fine-tuned network (pre-processing of patch size 224×224).

Network	classifier in [29]	fine-tuned classifier
ResNet-18 [33]	96.53%	97.01%
ResNet-50 [33]	96.63%	97.09%
VGG-19 [36]	96.74%	97.23%
InceptionV3 [37]	87.02%	88.33%
DenseNet [28]	97.04%	97.41%

$\beta_2 = 0.999$, *learning rate* = 10^{-5} for the Adam optimizer, scheduler which reduces the *learning rate* by 5% each 5 epochs if the validation does not improve, $M = 2048$ as dimension of the embedding feature vectors, the grid size $N = 3 \times 3 = 9$, $\sigma = 56$ and spatial restriction = 784×784 for training each grid of nine patches.

We evaluate the performance of the proposed DCRF by the two indicators namely classification accuracy and free-response receiver operating characteristic (FROC). As shown in Table 3 the DenseNet based DCRF outperforms other CNN structures based one.

TABLE 3
The accuracy and FROC score comparison between DCRF implemented on different CNN structures in tumor classification on test sets.

CNN feature extractor	accuracy	FROC
ResNet-18 [33]	98.04%	0.801
ResNet-50 [33]	98.12%	0.797
VGG-19 [36]	97.04%	0.781
InceptionV3 [37]	92.12%	0.743
DenseNet [28]	98.87%	0.799

TABLE 4
The performance acceleration of the proposed methods. Both of our models outperform other literary works in terms of computational time and AUC. The DCRF-FTD achieves a higher performance speedup at the cost of more training time.

method	sampled patches	time cost	speed-up	training time per epoch	patch-level AUC	FROC
regular sampling [33]	100%	4m39s	1.00×	45m30s	0.801	0.801
Monte Carlo sampling [39], [40]	93.2%	3m52s	1.20×	45m12s		
quasi Monte Carlo sampling [39]	89.2%	3m33s	1.12×	45m22s		
HASHI [41]	63.2%	2m51s	1.63×	52m44s		
DMFTD [25]	56.2%	2m32s	1.84×	50m12s	0.785	0.810
DCRF based regular sampling [24]	100%	5m40s	0.82×	92m38s	0.776	0.813
DP-FTD without NCRF	34.8%	2m23s	1.95×	43m02s	0.776	0.811
our proposed DP-FTD	33.4%	2m1s	2.31×	43m02s	0.779	0.817
our proposed DCRF-FTD	17.4%	1m19s	3.53×	92m43s	0.786	0.821

3.2 Performance Evaluation

3.2.1 Iterative Sampling Results

The number of patches in the initial sampling and the iterative sampling is set to be 1% of valid patches in WSI (background excluded) in both two models. We show a generation of tumor likelihood maps via the iterative quasi MC sampling in DCRF-FTD in Fig. 6. We demonstrate an example of WSI with a dynamic correction model in the proposed DP-FTD in Fig. 7.

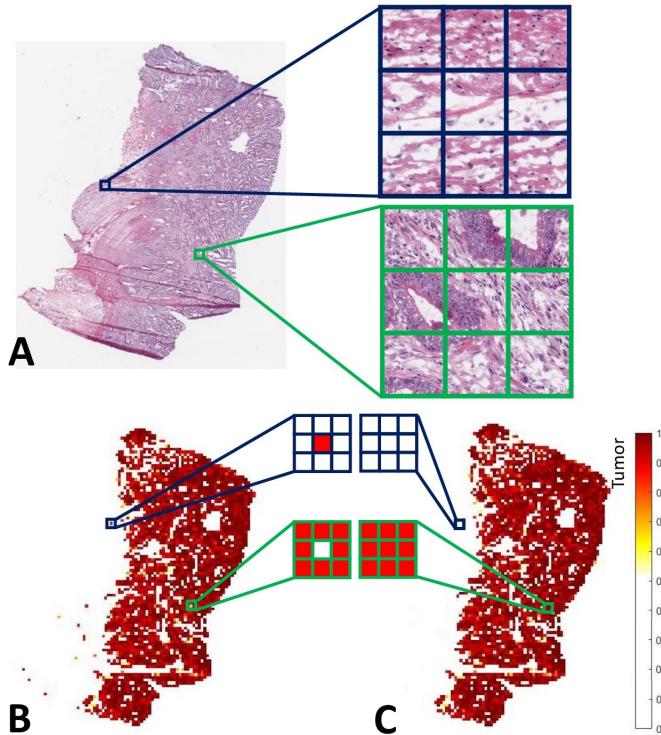


Fig. 7. WSI tumor distribution map from TCGA-COAD dataset. A) Scanned digital WSI, B) Prediction result of tumor distribution from regular sampling. C) The corrected tumor possibility map with spatial context by the proposed DP-FTD. Blue color cube highlights the false positive and green color for the false negative in tumor identification.

A couple of final prediction results by the proposed DCRF-FTD along with their original WSIs, are presented in Fig. 8. There is no obverse discrimination by eye between the classification outcome of DP-FTD, although there is a slight difference in accuracy, shown in Table 4.

3.2.2 Performance Speedup and Accuracy Improvement

The processing time of examples in Fig. 8 is presented in Fig. 9. Our methods obviously outperform the regular sampling by a large margin of speedup in the recognition of tumor tissue in histopathology.

Given the variance of tumor proportion and distribution, we profile the performance of all the datasets in Fig. 10, where both proposed models outperform other approaches at their convergence. DCRF-FTD achieves an even faster convergence by sampling fewer image patches. We also have models compared with other baseline methods [33], [39], [40], [41]. Our models achieve a more stable performance, which is independent of the proportion or distribution of tumor tissue in a WSI. We present the performance improvement with the proposed DP-FTD and DCRF-FTD in Table 4. The absence of NCRF for dynamic corrections on suspicious patches leads to a slight decline from 0.779 to 0.776 in AUC, and 0.789 to 0.786 in FROC score. It indicates that utilizing NCRF for correcting the suspicious can improve the classification accuracy. The averaged percentages of the suspicious patches identified by our proposed DP-FTD is 4.32%.

As shown, we achieve an acceleration of 3.53× and 2.31× speed-up on WSI-level processing respectively with the DCRF-FTD and DP-FTD. The area under the receiver-operating characteristic curve (ROC) of two proposed models and the baseline regular sampling is shown in Fig. 11.

We empirically show that both proposed models, the DP-FTD and DCRF-FTD, show their advantages in WSI-level performance acceleration and a slight accuracy improvement in the comparison with the methods in literature [33], [39], [40], [41]. That is an empirical proof that our methodology can be valid to WSI analysis. DCRF-FTD is differentiated for its less time-costly in the inference stage, yet more time-consuming in the training stage.

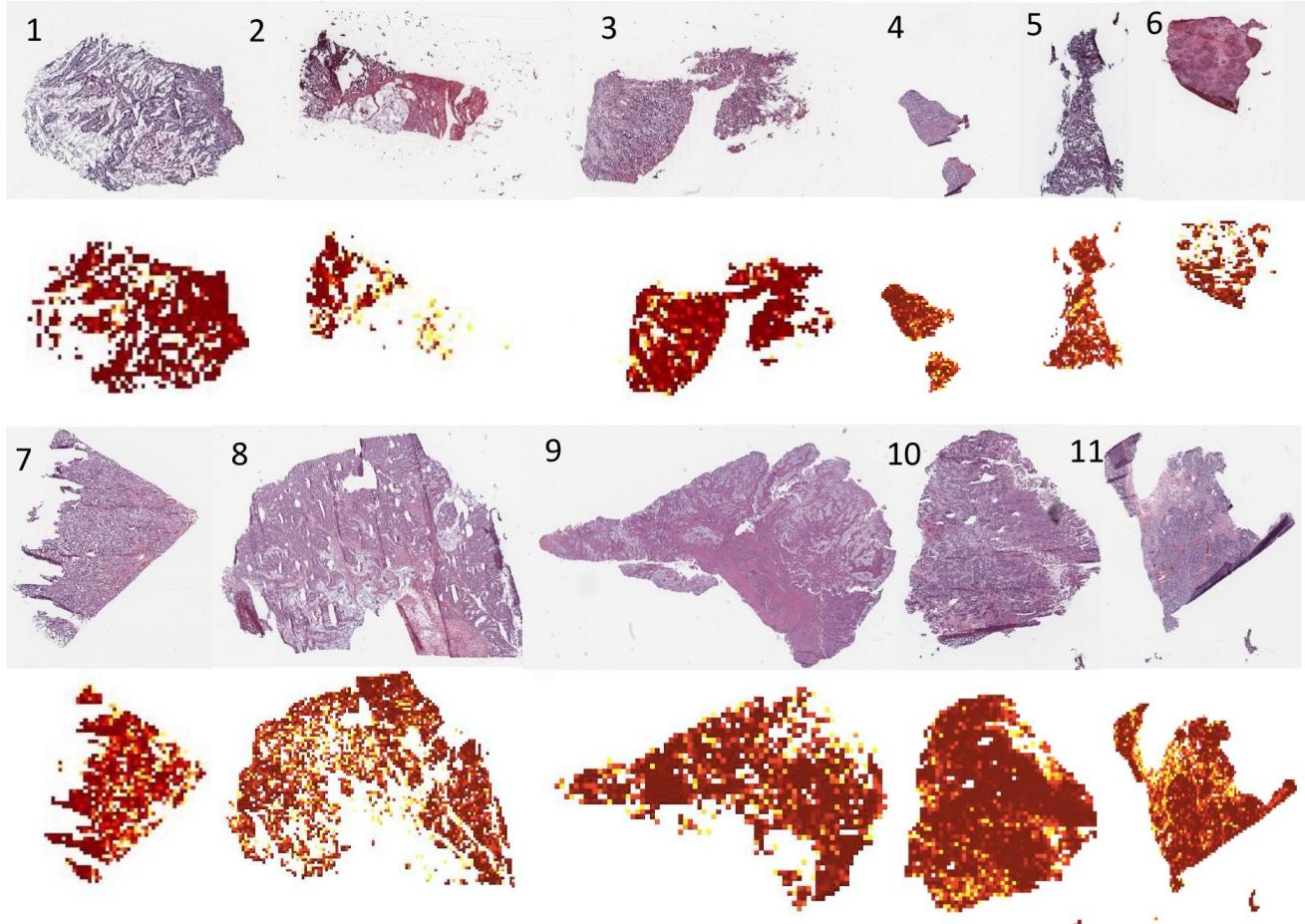


Fig. 8. A few examples of the predicted tumor possibility map along with their WSIs by the proposed DCRF-FTD.

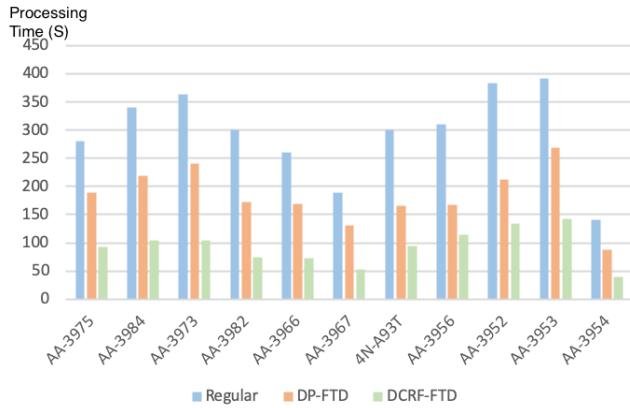


Fig. 9. A few examples of the processing time on WSI-level tumor detection from three methods i.e., the baseline regular sampling, the proposed DP-FTD and DCRF-FTD. Their original WSIs and the output tumor distribution maps are shown in Fig. 8.

4 CONCLUSION

In this paper, we propose a high-throughput computer-aided diagnostic system to efficiently locate tumor regions in colorectal cancer histopathology images. The performance speedup is achieved by a concentration of regions of interest in patch sampling. We significantly reduce the number of patches to be processed by convolutional neural

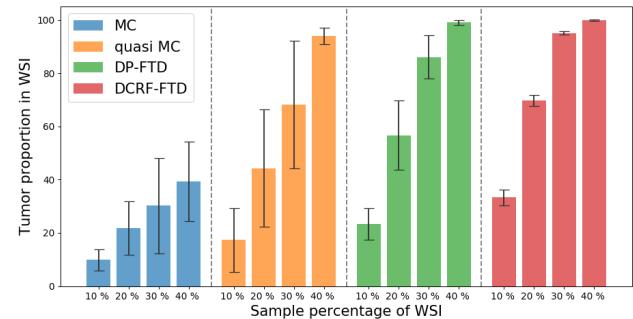


Fig. 10. The performance of WSI-level tumor detection evaluated on TCGA datasets. Our models achieve a faster convergence than the baseline models, compared with other baseline methods [39], [40], [41].

networks. Additionally, both proposed methods acquire a more precious prediction by incorporating spatial context from the neighboring patches. This high-efficient diagnostic system significantly decreases the heavy computational cost and can be applied in clinical practice to save more waiting time from patients. Furthermore, the performance acceleration is even more obvious when only a small proportion of tumor tissues are contained in a WSI, hence the model is particularly efficient at the identification of negative cases.

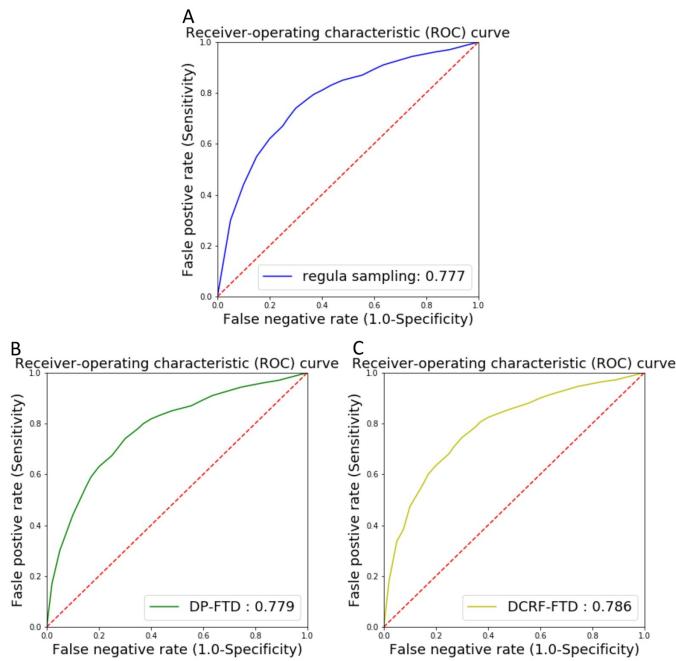


Fig. 11. Tumor detection evaluated by the area under the curve (AUC) of receiver-operating characteristic (ROC). A) Baseline regular sampling method, B) proposed DP-FTD, C) proposed DCRF-FTD. The WSI-level AUCs are 0.777, 0.779, 0.786 respectively.

REFERENCES

- [1] X. Yue, N. Dimitriou, and O. Arandjelovic, "Colorectal cancer outcome prediction from h&e whole slide images using machine learning and automatically inferred phenotype profiles," *arXiv preprint arXiv:1902.03582*, 2019.
- [2] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.
- [3] M. Alsharid, H. Sharma, L. Drukker, P. Chatelain, A. T. Papageorgiou, and J. A. Noble, "Captioning ultrasound images automatically," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 338–346.
- [4] K. M. van Wijnen, F. Dubost, P. Yilmaz, M. A. Ikram, W. J. Niessen, H. Adams, M. W. Vernooy, and M. de Brujinne, "Automated lesion detection by regressing intensity-based distance with a neural network," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 234–242.
- [5] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenholt *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.
- [6] H. Yoshida, T. Shimazu, T. Kiyuna, A. Marugame, Y. Yamashita, E. Cosatto, H. Taniguchi, S. Sekine, and A. Ochiai, "Automated histological classification of whole-slide images of gastric biopsy specimens," *Gastric Cancer*, vol. 21, no. 2, pp. 249–257, 2018.
- [7] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.
- [8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [9] D. C. Wilbur, "Digital cytology: current state of the art and prospects for the future," *Acta cytologica*, vol. 55, no. 3, pp. 227–238, 2011.
- [10] A. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," 2016.
- [11] Y. Li and W. Ping, "Cancer metastasis detection with neural conditional random field," *arXiv preprint arXiv:1806.07064*, 2018.
- [12] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz, "Patch-based convolutional neural network for whole slide tissue image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2424–2433.
- [13] D. Bychkov, N. Linder, R. Turkki, S. Nordling, P. E. Kovánen, C. Verrill, M. Walliander, M. Lundin, C. Haglund, and J. Lundin, "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Scientific reports*, vol. 8, no. 1, p. 3395, 2018.
- [14] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS medicine*, vol. 16, no. 1, p. e1002730, 2019.
- [15] J. N. Kather, Krisam *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS medicine*, vol. 16, no. 1, p. e1002730, 2019.
- [16] A. Janowczyk, S. Doyle, H. Gilmore, and A. Madabhushi, "A resolution adaptive deep hierarchical (radhical) learning scheme applied to nuclear segmentation of digital pathology images," *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 270–276, 2018.
- [17] H. Tokunaga, Y. Teramoto, A. Yoshizawa, and R. Bise, "Adaptive weighting multi-field-of-view cnn for semantic segmentation in pathology," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12597–12606.
- [18] T. Qaiser and N. M. Rajpoot, "Learning where to see: A novel attention model for automated immunohistochemical scoring," *IEEE transactions on medical imaging*, vol. 38, no. 11, pp. 2620–2631, 2019.
- [19] A. BenTaieb and G. Hamarneh, "Predicting cancer with a recurrent visual attention model for histopathology images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 129–137.
- [20] F. G. Zanjani, S. Zinger *et al.*, "Cancer detection in histopathology whole-slide images using conditional random fields on deep embedded spaces," in *Medical imaging 2018: Digital pathology*, vol. 10581. International Society for Optics and Photonics, 2018, p. 105810I.
- [21] B. Kong, X. Wang, Z. Li, Q. Song, and S. Zhang, "Cancer metastasis detection via spatially structured deep network," in *International Conference on Information Processing in Medical Imaging*. Springer, 2017, pp. 236–248.
- [22] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P. Heng, "Fast scannet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1948–1958, 2019.
- [23] Z. Guo, H. Liu, H. Ni, X. Wang, and Y. Qian, "Publisher correction: A fast and refined cancer regions segmentation framework in whole-slide breast pathological images," *Scientific Reports*, vol. 10, no. 1, p. 8591, 2020.
- [24] Y. Shen and J. Ke, "A deformable crf model for histopathology whole-slide image classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 500–508.
- [25] J. Ke, Y. Shen, Y. Guo, and X. Liang, "Fast tumor detector in whole-slide image with dynamic programming based monte carlo sampling," in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 2471–2475.
- [26] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan, "An introduction to mcmc for machine learning," *Machine Learning*, vol. 50, no. 1-2, pp. 5–43, 2003.
- [27] P. Krähenbühl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in *Advances in neural information processing systems*, 2011, pp. 109–117.
- [28] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "Densenet: Implementing efficient convnet descriptor pyramids," *arXiv preprint arXiv:1404.1869*, 2014.
- [29] J. N. Kather, A. T. Pearson, N. Halama, D. Jäger, J. Krause, S. H. Loosen, A. Marx, P. Boor, F. Tacke, U. P. Neumann *et al.*, "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature medicine*, p. 1, 2019.
- [30] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber *et al.*, "Predicting survival from colorectal cancer histology slides using

- deep learning: A retrospective multicenter study," *PLoS medicine*, vol. 16, no. 1, p. e1002730, 2019.
- [31] J. N. Kather, N. Halama, and A. Marx, "100,000 histological images of human colorectal cancer and healthy tissue," Apr. 2018. [Online]. Available: <https://doi.org/10.5281/zenodo.1214456>
- [32] A. I. Bandos, H. E. Rockette, T. Song, and D. Gur, "Area under the free-response roc curve (froc) and a related summary index," *other*, vol. 65, no. 1, 2009.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [34] J. Ke, Y. Shen, Y. Guo, J. D. Wright, and X. Liang, "A prediction model of microsatellite status from histology images," in *Proceedings of the 2020 10th International Conference on Biomedical Engineering and Technology*, 2020, pp. 334–338.
- [35] J. Ke, Y. Shen, Y. Guo, J. D. Wright, N. Jing, and X. Liang, "A high-throughput tumor location system with deep learning for colorectal cancer histopathology image," in *Artificial Intelligence in Medicine*. Springer International Publishing, 2020, pp. 260–269. [Online]. Available: https://doi.org/10.1007%2F978-3-030-59137-3_24
- [36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [38] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.
- [39] L. Martino, D. Luengo, and J. Míguez, *Independent random sampling methods*. Springer, 2018.
- [40] M. Shapcott, K. J. Hewitt, and N. Rajpoot, "Deep learning with sampling in colon cancer histology," *Frontiers in Bioengineering and Biotechnology*, vol. 7, p. 52, 2019.
- [41] A. Cruz-Roa, H. Gilmore, A. Basavanhally, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, A. Madabhushi, and F. González, "High-throughput adaptive sampling for whole-slide histopathology image analysis (hashi) via convolutional neural networks: Application to invasive breast cancer detection," *PloS one*, vol. 13, no. 5, p. e0196828, 2018.