

Prior Information Guided Regularized Deep Learning for Cell Nucleus Detection

Mohammad Tofighi, *Student Member, IEEE*, Tiantong Guo, *Student Member, IEEE*, Jairam K.P. Vanamala, and Vishal Monga, *Senior Member, IEEE*

Abstract—Cell nuclei detection is a challenging research topic because of limitations in cellular image quality and diversity of nuclear morphology, i.e. varying nuclei shapes, sizes, and overlaps between multiple cell nuclei. This has been a topic of enduring interest with promising recent success shown by deep learning methods. These methods train Convolutional Neural Networks (CNNs) with a training set of input images and known, labeled nuclei locations. Many such methods are supplemented by spatial or morphological processing. Using a set of canonical cell nuclei shapes, prepared with the help of a domain expert, we develop a new approach that we call Shape Priors with Convolutional Neural Networks (SP-CNN). We further extend the network to introduce a shape prior (SP) layer and then allowing it to become trainable (i.e. optimizable). We call this network tunable SP-CNN (TSP-CNN). In summary, we present new network structures that can incorporate ‘expected behavior’ of nucleus shapes via two components: *learnable* layers that perform the nucleus detection and a *fixed* processing part that guides the learning with prior information. Analytically, we formulate two new regularization terms that are targeted at: 1) learning the shapes, 2) reducing false positives while simultaneously encouraging detection inside the cell nucleus boundary. Experimental results on two challenging datasets reveal that the proposed SP-CNN and TSP-CNN can outperform state-of-the-art alternatives.

Index Terms—nucleus detection, deep learning, convolutional neural networks, shape priors, learnable shapes

I. INTRODUCTION

Microscopic images often exhibit high degree of cellular heterogeneity [1], [2]. Cell nucleus detection methods locate the cell and annotate the center of the cell nuclei. Visual-based techniques in many medical imaging applications, e.g. manual detection and counting, are extremely time consuming and do not scale well as the task is to be performed for a large number of images [3], [4]. Therefore, automatic analysis of cellular imagery to determine nuclei locations is a centrally important problem in diagnosis of several medical conditions including tumor and cancer detection [5], [6].

Related work: Some of the earliest attempts at cell nuclei detection involved tailored feature extraction and morphological processing [7], [8], [9], [10]. One key limitation of these approaches is that the best features for nuclei detection are rarely readily apparent. Further, often the designed techniques are too specific to the choice of dataset and not versatile. Because of their ability to perform feature discovery and

M. Tofighi, T. Guo, and V. Monga are with the Department of Electrical Engineering, Pennsylvania State University, University Park, PA. e-mail: tofighi@psu.edu, tiantong@psu.edu, juv4@psu.edu, vmonga@engr.psu.edu.

J.K.P. Vanamala is with Center for Molecular Immunology and Infectious Disease, Pennsylvania State University, University Park, PA.

Research has been supported by an NSF CAREER award (to V. Monga).

generalizable inference, deep learning methods have recently become popular for this problem [11], [12], [13], [14]. For instance, Cruz-Roa *et al.* [15] showed that a deep learning architecture for nuclei detection outperforms methods based on different image representation strategies e.g. bag of features, canonical and wavelet transforms. Xie *et al.* [16], proposed a structural regression model for CNN, where a cell nuclei center is detected if it has the maximum value in the proximity map. In [17], Xu *et al.* proposed a cell detection method based on a stacked sparse autoencoder, where it learns high level features of cell centroids and then a softmax classifier is used to separate the nuclear and non-nuclear image patches. Sirinukunwattana *et al.* proposed SC-CNN [5] which uses a regression approach to find the likelihood of a pixel being the center of a nucleus. In SC-CNN, the probability values are topologically constrained in a way that in vicinity of nuclei center the probability is higher. Another recent approach [18] uses a combination of well-known traditional CNNs for cell nuclei segmentation and dictionary learning techniques for refining segmentations results. Combining CNNs with sparse coding, [19] uses a sparse convolutional autoencoder (CSP-CNN) for simultaneous nucleus detection and feature extraction. In CSP-CNN, the detection is based on a fairly deep network (> 15 layers) comprising of multiple CNN branches that perform detection, segmentation and image reconstruction.

We develop a novel prior guided learning approach by exploiting a prior understanding of the shape of the nuclei, which can be obtained in consultation with a domain expert. The value of shape based information has been recognized for medical image segmentation [20], [21]. Leventon *et al.* [22] proposed a method for medical image segmentation by incorporating shape information into the geodesic active contour method. Ali *et al.* [23] incorporated prior shape information into boundary and region based active contours for accurate segmentation of cells. In [24], Oktay *et al.* propose to incorporate shape prior information into enhancement and segmentation of MR cardiac images. They propose a fidelity term between the output segmentation regions and the ground-truth region shapes. While the approach in [24] also exploits anatomical shape information, technically the proposed SP-CNN is fundamentally different. First, the image modality in our work is cellular as opposed to organ (MR) imagery in [24], which leads to vastly different quantitative formulations. Second, our proposed regularizer is not a fidelity term but guides cell nuclei detection in regions where the probability of presence of a cell nuclei is higher according to the shape prior knowledge. Finally, architecturally the proposed SP/TSP-

CNN and the neural network in [24] are entirely different.

Prior information with deep networks has emerged as a promising direction in imaging inverse problems such as super-resolution [25], [26]. These methods use output image priors to enhance the super-resolution task. Similarly in related work, a Deep Image Prior [27] network has been specifically designed for tasks such as image in-painting, de-noising, and super-resolution. Shape based priors in a deep learning based nuclei detection framework however remain elusive and form the focus of our work.

Contributions: While existing deep learning approaches for cell nuclei detection are promising, our goal is to fundamentally alter the learning of the network by enriching it with domain knowledge provided by a medical expert. Specifically, our key contributions¹ are as follows:

- **Shape Prior with a Convolutional Neural Network (SP-CNN):**

We propose a novel network structure that can incorporate ‘expected behavior’ of nuclei shapes via two components: *learnable* layers that perform the nucleus detection and a *fixed* processing part that guides the learning with prior information. Three sources contribute to generating the prior: network output, raw edge map from the input image, and a set of predefined shapes prepared by the medical expert. These shapes are binary images representing the boundary of cell nuclei. The *fixed* priors guide the *learnable* layers to perform detection consistent with a nucleus boundary.

- **Tunable Shape Priors with a Convolutional Neural Network (TSP-CNN):**

As a key extension, we develop a new approach where the set of shapes is represented as a convolutional layer, which is no-longer fixed but learned. That is, using the expert provided shapes as a starting point, we both refine and remove redundancy from the shape set (using shape similarity measures) to arrive at a new learned shape set that is not only more economical but also enhances detection accuracy.

- **Novel Regularization Terms:**

Analytically, we formulate two new regularization terms: a shape prior term and a shape learning term. The shape prior term is used in both SP-CNN and TSP-CNN and incorporates the effect of prior information by using a set of shapes to guide the learning of the network towards greater accuracy. The shape learning term is only used in TSP-CNN and achieves the refinement/learning of shapes in a constrained manner, i.e. by emphasizing similarity to a reference shape set. We carefully design these terms so they are differentiable w.r.t. the output and hence the network parameters, enabling tractable learning through standard back-propagation schemes.

- **A New Dataset, Broad Experimental Validation and Reproducibility:**

Experimental validation of SP-CNN and TSP-CNN is carried out on two diverse cancer tissue datasets to show its broad applicability. We also provide a new open source expert annotated dataset of microscopic colon tissue images for cell nuclei detection.

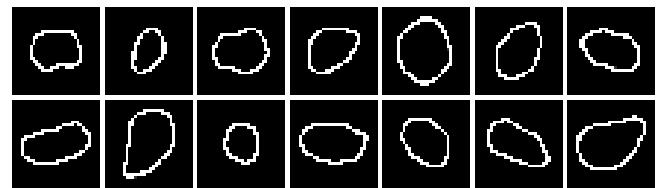


Fig. 1: Samples of handcrafted cell nuclei shapes from colon tissues in UW Dataset. For dataset details see Sec. IV.

We call it the PSU Dataset and it is prepared by the help of experts in Center for Molecular Immunology and Infectious Disease, Penn State University [29]. The second dataset is courtesy of Sirinukunwattana *et al.* at Department of Computer Science, University of Warwick (we call it UW Dataset) which includes colorectal adenocarcinoma images [5]. Extensive experimental results in the form of benchmark measures such as F1 scores and precision-recall curves show that SP-CNN and TSP-CNN outperform many recent methods [5], [16], [7], [17], [30], [19], particularly in very challenging scenarios. We also make our code and the PSU dataset freely available at our project webpage [29].

The rest of this paper is organized as follows. The proposed SP-CNN structure that incorporates shape priors into cell nuclei detection and formulation of the new regularized cost function are detailed in Section II. Section III presents the extension of SP-CNN to tunable shape priors. Detailed experimental comparisons against competing state-of-the-art methods are provided in Section IV. Section V summarizes the findings and concludes the paper.

II. SHAPE PRIORS WITH CONVOLUTIONAL NEURAL NETWORKS (SP-CNN)

A. CNN for Nucleus Detection

SP-CNN detects the cell nucleus using a regression CNN. In regression networks, the goal is to obtain a function that interprets the relationship between the input image \mathbf{x} and ground-truth labeled image \mathbf{y} . The network is modeled by parameters set $\Theta = \{\mathbf{W}_l, \mathbf{b}_l\}_{l=1}^D$, where \mathbf{W}_l and \mathbf{b}_l denote respectively the weights and bias of l -th layer of total D layers collectively. The CNN is learned by solving the well-known optimization problem [5], [31]:

$$\Theta = \arg \min_{\Theta} \|f(\mathbf{x}; \Theta) - \mathbf{y}\|_2^2 \quad (1)$$

where $f(\mathbf{x}; \Theta)$ represents the non-linear mapping of the CNN that generates the detection maps $\hat{\mathbf{y}}$. We work with soft labels \mathbf{y} and $\hat{\mathbf{y}}$, i.e. $\mathbf{y}, \hat{\mathbf{y}}$ can take values in the range $[0, 1]$. In practice, \mathbf{y} is obtained by processing the binary image (0 or 1 at each pixel) of ground-truth nuclei locations as in [5]; also see Section IV. Each of the D CNN layers comprises of a convolutional layer followed by an activation function, which is a Rectified Linear Unit (ReLU) [32] (except the last layer).

B. Deep Networks With Shape Priors

As discussed in Section I, we now incorporate the prior information about the shape of the cell nuclei into the training

¹A preliminary 4 page version of this work was published at IEEE ICIP 2018 in Athens, Greece held in October 2018 [28].

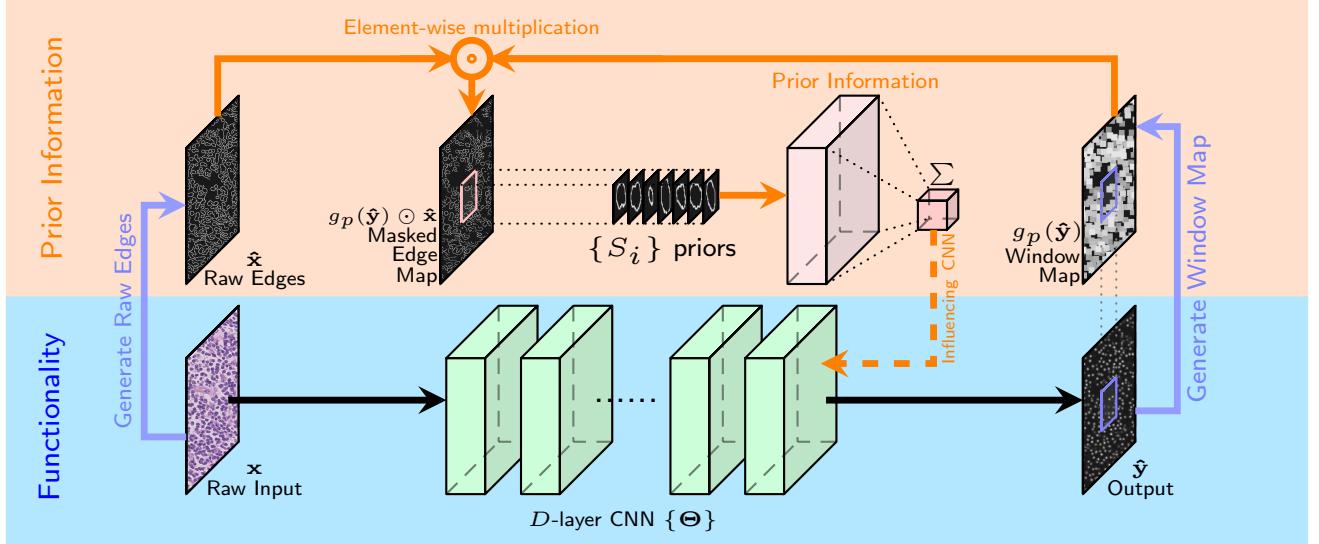


Fig. 2: SP-CNN illustration. There are two parts of the SP-CNN: Functionality part (blue) and Prior Information part (orange). The Functionality part consists of one D -layer CNN that takes input raw image x and generates the detected labels for cell nuclei \hat{y} . The prior information part computes the prior cost term as in Eq. (2) and feeds the information into the (learning of) CNN through back-propagation to guide it towards enhanced nuclei detection. Note the prior information part generates a regularization term used in training the D learnable layers of the CNN and *only* the functionality part of the network is applied to a test image. Hence, while the black, orange, and blue arrows are used during training, in testing stage only the black arrows are used.

of the CNN. Ideally, the labels produced by the CNN should lie inside of the nuclei boundaries. We set up a regularization term to explicitly encourage the learned network to achieve detection inside the nucleus boundary while simultaneously reducing false positives. The regularizer is based on a set of shape priors developed with the help of domain expert and given by: $S = \{S_i | i = 1, 2, \dots, N\}$.

For each dataset, multiple training images are analyzed by a medical expert to hand label the nuclei boundaries. A set of N representative shapes is then hand crafted by the medical expert to form the set S . Some examples of the nucleus shape priors are shown in Fig. 1. These are corresponding to colon tissue images – detailed explanation is provided in Section IV. To construct a meaningful regularization term emphasizing shape priors, we need the nucleus boundary information of the input raw image x . We employ the widely used Canny edge detection² filter [33] to generate the raw edge image \hat{x} with edges labeled as 1 and background as 0, as shown in Fig. 3-(b). Note that the raw edge image \hat{x} is only used during the training process. We now define the regularization term that captures shape priors:

$$\mathcal{L}^{\text{SP}} = -\lambda \sum_{i=1}^N \| (g_p(\hat{y}) \odot \hat{x}) * S_i \|_2^2 \quad (2)$$

where the term $g_p(\cdot)$ denotes the max pooling operation on \hat{y} with window size p , \odot represents element-wise multiplication, and $*$ is the 2D convolution operation. Based on Eq. (2), the computation of the shape priors cost term consists of three

²Note we select Canny edge detection filter because of its simplicity and efficiency. Although its performance satisfies our intentions, other edge detection methods can also be used.

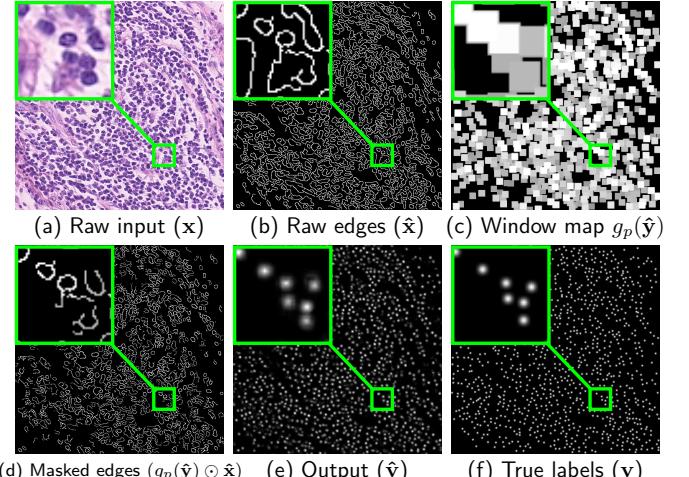


Fig. 3: Images in each step of SP-CNN.

steps as shown in Fig. 2's prior information (orange) part:

- 1) The CNN output \hat{y} is first thresholded by $T_p = 0.2$ to eliminate the background noise (Fig. 3-(e)) and then max pooled by $g_p(\cdot)$ with stride of 1 and the ‘SAME’ padding scheme³. This results in a window map $g_p(\hat{y})$ that has $p \times p$ window centered at each location within the soft detected region (Fig. 3-(c)). As expected, a window with higher numerical value will result if the detected label values (in \hat{y}) are correspondingly higher (closer to 1).
- 2) The window map $g_p(\hat{y})$ is then multiplied with the raw edge image \hat{x} element-wise. This step serves to mask

³Because there is no change in the image grid, this is similar to the box-filtering operation in image processing.

out the edges from $\hat{\mathbf{x}}$ that surround the detected location in $\hat{\mathbf{y}}$, as shown in Fig. 3-(d).

- 3) The masked edge image ($g_p(\hat{\mathbf{y}}) \odot \hat{\mathbf{x}}$) is convolved with the shape priors in set \mathbf{S} to generate a measurement of how well does the detection fit inside the nucleus shape. If $\hat{\mathbf{y}}$ has more labels predicted inside the nucleus boundary, Eq. (2) will achieve a smaller value (more negative).

Note that the effect of the shape prior is captured by a negative regularization term since the goal is to maximize (and not minimize) correlation with ‘expected shapes’. Overall, the cost function of the SP-CNN is given by:

$$\begin{aligned} \Theta &= \arg \min_{\Theta} \mathcal{L}^{\text{Loss}} + \mathcal{L}^{\text{SP}} \\ &= \arg \min_{\Theta} \|f(\mathbf{x}; \Theta) - \mathbf{y}\|_2^2 - \lambda \sum_{i=1}^N \|(g_p(\hat{\mathbf{y}}) \odot \hat{\mathbf{x}}) * \mathcal{S}_i\|_2^2 \end{aligned} \quad (3)$$

where λ is the trade-off parameter between the squared loss term and the regularizer representing the effect of the shape prior. The term \mathcal{L}^{SP} is carefully designed to simultaneously accomplish two tasks: 1.) a high detection rate of nuclei is encouraged since the element-wise Hadamard product highlights edge boundaries, and 2.) the subsequent convolution with the expert provided shape set reduces false positives. Note that $\hat{\mathbf{y}} := f(\mathbf{x}; \Theta)$, thus the shape prior cost term is effected by the network parameters and also introduces gradient terms that updates the network parameters during the training process using back-propagation [34]. This is indicated in Fig. 2 by dashed line under “Influencing CNN”. Note that while the black, orange, and blue arrows are used during training, the black arrows are just used in testing stage.

Let $\mathcal{L} = \mathcal{L}^{\text{Loss}} + \mathcal{L}^{\text{SP}}$ be the cost function. At iteration t of the back-propagation, the filters of CNN are updated as:

$$\Theta^{t+1} = \Theta^t - \eta \nabla_{\Theta} \mathcal{L} \quad (4)$$

where η denotes the learning rate, and $\Theta = \{\mathbf{W}_k, \mathbf{b}_k\}_{k=1}^D$. The following gradients are to be computed⁴:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_k}, \frac{\partial \mathcal{L}}{\partial \mathbf{b}_k}$$

where \mathbf{W}_k and \mathbf{b}_k denote one of the filters and biases at k -th layer of the CNN, representatively. The equation for computing the gradient w.r.t. an arbitrary entry within filter \mathbf{W}_k in layer $k \in \{1, \dots, D\}$ is given by:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{W}_k^a} &= - \langle (\hat{\mathbf{y}} - \mathbf{y}), \frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{W}_k^a} \rangle_F \\ &\quad - \lambda \sum_{i=1}^N \langle (g_p(\hat{\mathbf{y}}) \odot \hat{\mathbf{x}}) * \mathcal{S}_i, g_p^{-1} \left(\frac{\partial \hat{\mathbf{y}}}{\partial \mathbf{W}_k^a} \right) \odot \hat{\mathbf{x}} * \mathcal{S}_i \rangle_F \end{aligned} \quad (5)$$

where \mathbf{W}_k^a denotes an arbitrary scalar entry within the representative filter \mathbf{W}_k , $\langle \cdot, \cdot \rangle_F$ denotes the real value Frobenius inner product⁵, and $g_p^{-1}(\cdot)$ assigns the derivative to where

⁴Note the update rules and the gradients for the bias terms are similar and are included with more details in the supplementary document [29].

⁵For two real valued matrices \mathbf{A} and \mathbf{B} with same dimension, $\langle \mathbf{A}, \mathbf{B} \rangle_F := \sum_{i,j} A_{i,j} B_{i,j}$ where i, j are the indexes of the entries.

activations come from, the “winning unit”, because other units in the previous layer’s pooling blocks did not contribute to it. Hence derivative of all the other units is assigned to zero. In Eq. (5), $\frac{\partial \mathbf{y}}{\partial \mathbf{W}_k^a}$ is computed by following the standard backpropagation rule for each layer k [35].

III. TUNABLE SHAPE PRIORS WITH CONVOLUTIONAL NEURAL NETWORKS (TSP-CNN)

While exploiting a shape set provided by a domain expert is meritorious, a fundamental open question is: What is the best, most compact shape set to employ and can it be adapted based on the image characteristics? This section provides an answer to that question in the following manner: the set of shapes can be seen as a collection of filters, which we now add to the network as a learnable component and call it the learnable shape layer – see Fig. 4 for an illustration of this idea. We call this new network structure as Tunable Shape Priors with CNNs (TSP-CNN).

We observe that the domain expert provided shape set in practice is quite redundant, i.e. it may contain pairs of hand-crafted shapes that are quite close to each other and do not add value to nucleus detection. We therefore propose a two stage process to arrive at a compact *Basis Shape Set* which is adapted based on training imagery: 1.) shape elimination stage to obtain a reduced/smaller reference shape set, and 2.) shape refinement or learning via TSP-CNN. This two stage process is illustrated in Fig. 5. For both stages, meaningful shape similarity measures are needed; the exact shape elimination and refinement procedure is detailed next. We next introduce shape elimination and refinement procedures, and then describe the training procedure to learn shapes that are adapted to the underlying dataset.

A. Shape Elimination

Many shape similarity measures have been developed in image processing and vision for a variety of different tasks [36], [37], [38], [39], [40], [41], [42]. Most of these methods measure the distance between shapes A and B , $d(A, B)$ in different ways, for example bottleneck distance, Hausdroff distance, turning function distance, etc [36]. Attalla *et al.* used a multi-resolutional polygonal shape descriptor that is invariant to scale and is used to match and recognize 2D shapes. In [39], Latecki *et al.* proposed a cognitively motivated similarity measure which divides a shape into best possible correspondences of visual parts and measures the similarity between each of them and then aggregates them to find the overall similarity measure.

In [37], Sampat *et al.* proposed Complex Wavelet SSIM (CW-SSIM) which is an extension of the well-known Structural Similarity Index Measure (SSIM) [43]. CW-SSIM has been shown to be insensitive to nonstructural geometric distortions and an excellent measure to compare binary shape similarity [37] because of its ability to capture phase information. CW-SSIM index [37] for two sets of coefficients in the complex wavelet transform domain as $\mathbf{c}_x = \{c_{x,i} | i = 1, \dots, M\}$ and $\mathbf{c}_y = \{c_{y,i} | i = 1, \dots, M\}$, extracted from

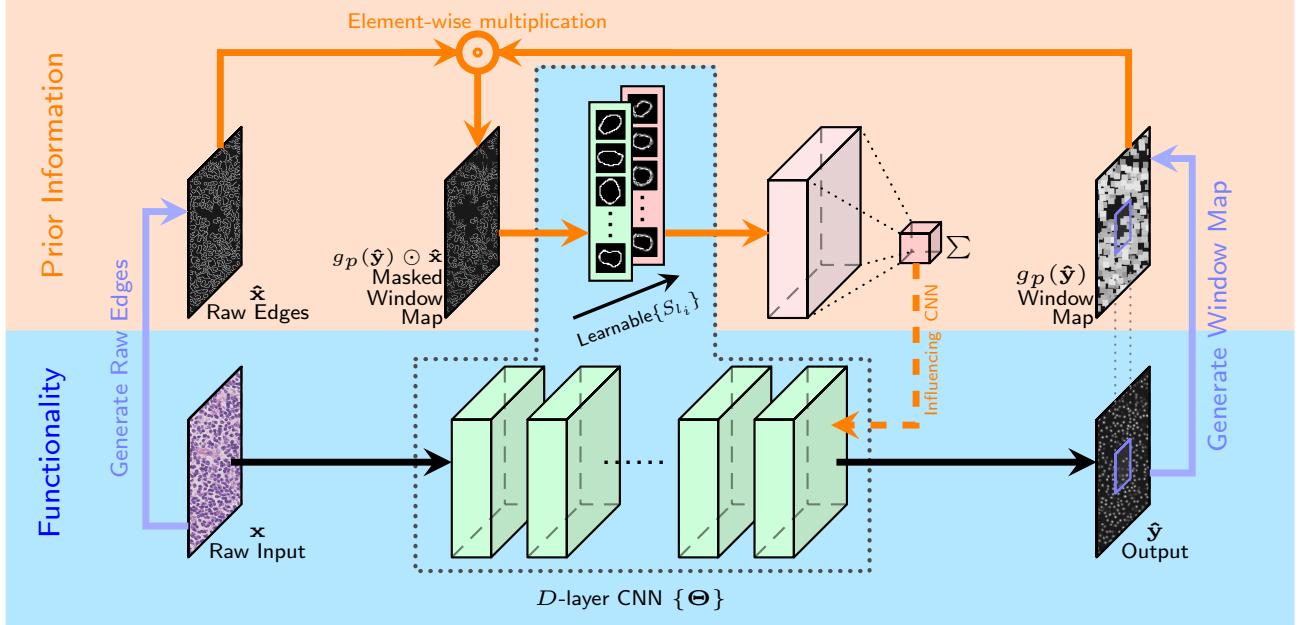


Fig. 4: TSP-CNN illustration. There are two parts of TSP-CNN: Functionality part and Prior Information part. Compared to SP-CNN, the shape set S_l is no longer fixed. It is now also a learnable component of the network. The shapes in the shape set are updated according to the shape learning regularization term as in Eq. (8). As with SP-CNN, the prior information part generates regularization terms used in training the D learnable layers of the CNN (through the blue and orange arrows) but *only* the functionality part of the network, i.e. the D layer CNN (excluding the learned shapes) is applied to a test image to predict nuclei centers (through the black arrows).

same location of same wavelet subbands of two shape images being compared is defined as:

$$\text{CW-SSIM}(\mathbf{c}_x, \mathbf{c}_y) = \frac{2|\sum_{i=1}^M c_{x,i} c_{y,i}^*| + K}{\sum_{i=1}^M |c_{x,i}|^2 + \sum_{i=1}^M |c_{y,i}|^2 + K}, \quad (6)$$

where c^* is the complex conjugate of c and K is a small positive constant. After computing CW-SSIM values for all sets, \mathbf{c}_i 's, the final CW-SSIM value for two shape images $\text{CW-SSIM}(S_i, S_j)$ is obtained by averaging CW-SSIM value over all wavelet subbands as described in [37].

Starting with the domain expert provided shape set $\mathbf{S} = \{S_i | i = 1, 2, \dots, N\}$, we eliminate redundancy by using the aforementioned CW-SSIM measure to compute pairwise comparisons of shapes. Near identical (redundant) shapes are grouped together and one representative is extracted from the group. The complete procedure is detailed in Algorithm 1.

B. Shape Refinement

The output of Algorithm 1 is what we address as $\mathbf{S}_r = \{S_{r,j} | j = 1, 2, \dots, Q\}$ the Reference Shape Set – which is essentially a cardinality (size) reduced version of the domain expert provided shape set \mathbf{S} , i.e. $Q < N$. Our approach to learn or refine shapes is then to optimize a Learnable Shape Set $\mathbf{S}_l = \{S_{l,i} | i = 1, 2, \dots, Q\}$ under the important physical constraint that the shapes in \mathbf{S}_l bear similarity to those in \mathbf{S}_r .

Ideally, as in the shape elimination stage, CW-SSIM should be used as a similarity measure between \mathbf{S}_l and \mathbf{S}_r . However,

⁶In Algorithm 1, $\{\}$ is an empty set, $|\mathbf{S}|$ is the cardinality of set \mathbf{S} , and $\text{Rep}(\mathbf{S}_n)$ is one representative shape from grouped similar shapes in set \mathbf{S}_n .

Algorithm 1 Shape elimination procedure using CW-SSIM

```

Input:  $\mathbf{S} = \{S_i | i = 1, 2, \dots, N\}$ ,  $k, n = 1$ ,  $\widehat{\mathbf{S}}, \mathbf{S}_n = \{\}$ 6
1: while  $k = 1$  do
2:    $\mathbf{S}_n \leftarrow \mathbf{S}_n \cup \{S_k\}$ 
3:   for  $l = 2$  to  $N$  do
4:      $C = \text{CW-SSIM}(S_k, S_l)$ 
5:     if  $C > 0.8$  then
6:        $\mathbf{S}_n \leftarrow \mathbf{S}_n \cup \{S_l\}$ 
7:     else
8:       Go to line 3
9:     end if
10:   end for
11:    $\mathbf{S} \leftarrow \mathbf{S} \setminus \{\mathbf{S} \cap \mathbf{S}_n\}$ 
12:    $\widehat{\mathbf{S}} \leftarrow \widehat{\mathbf{S}} \cup \{\text{Rep}(\mathbf{S}_n)\}$ 
13:    $n \leftarrow n + 1$ 
14:    $N = |\mathbf{S}|$ 
15:   if  $N = 0$  then
16:      $k = 0$ 
17:   end if
18: end while
Output: Reference shape set  $\mathbf{S}_r = \widehat{\mathbf{S}}$ .

```

CW-SSIM is not differentiable and hence arduous to incorporate in standard CNN learning frameworks which rely on derivative based back-propagation. SSIM, a pre-cursor to CW-SSIM on the other hand is differentiable and implementable in a deep learning framework [44]. In Fig. 6, we compare CW-SSIM and SSIM values in five cases of binary shape

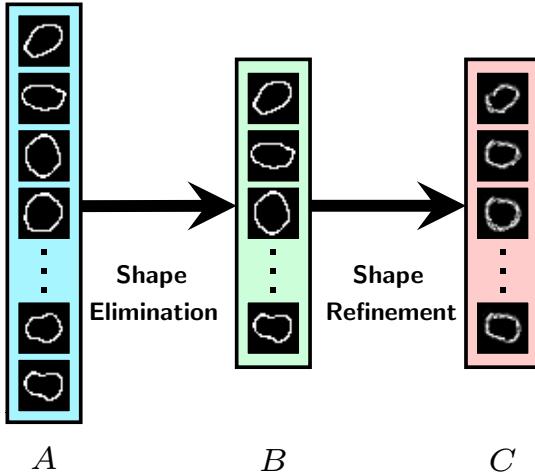


Fig. 5: An illustration of the two-stage process of shape elimination and refinement. Starting with the domain expert provided shape set, a redundancy removing elimination step is performed first using Algorithm 1. Second, refinement of this reduced size shape set is performed which adapts to training imagery and arrives at a compact *Basis Shape Set*.

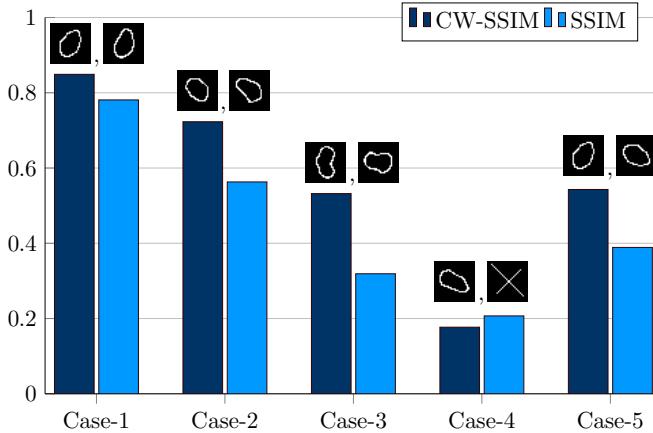


Fig. 6: SSIM & CW-SSIM values for different cases of shapes.

similarity by using representative nuclei shapes corresponding to the image datasets we work with: case-1: very similar, case-2: similar, case-3: different, case-4: very different, and case-5: rotation (90°). From Fig. 6 we can infer two facts: 1.) CW-SSIM can be quite effective in comparing binary nuclei shapes, and 2) SSIM forms an approximation to CW-SSIM. Given the tractability benefits of SSIM in deep learning set-ups, we employ it to build our shape similarity regularizer. As is corroborated later in Section IV, using an SSIM based regularizer suffices to afford TSP-CNN significant practical gains. The design of more sophisticated e.g. rotation invariant, similarity measures that can be tractably incorporated in deep learning frameworks is a viable direction for future research.

The SSIM index between patches $\mathbf{x} \in \mathcal{S}_{l_i}$ and $\mathbf{y} \in \mathcal{S}_{r_j}$ is

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + C_1}{\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + C_1} \cdot \frac{2\sigma_{\mathbf{x}\mathbf{y}} + C_2}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + C_2}, \quad (7)$$

where μ and σ are mean and standard deviation of the patches,

respectively. The patch selection, comparisons and estimation of local statistics is done as in [43]. The overall SSIM value between two shapes, $\text{SSIM}(\mathcal{S}_{l_i}, \mathcal{S}_{r_j})$, is the average of SSIM value over all the patches in the shapes.

In [44], the use of SSIM is motivated by image quality concerns. In TSP-CNN, differently from [44], we use SSIM to define a shape learning regularizer in the following manner:

$$\mathcal{L}^{\text{SSIM}} = -\gamma \text{SSIM}(\mathbf{S}_l, \mathbf{S}_r) = -\gamma \sum_{i=1}^Q \sum_{j=1}^Q \text{SSIM}(\mathcal{S}_{l_i}, \mathcal{S}_{r_j}) \quad (8)$$

and optimize the network parameters (including shapes) with:

$$\begin{aligned} \widehat{\Theta} &= \arg \min_{\Theta} \mathcal{L}^{\text{Loss}} + \mathcal{L}^{\text{SP}} + \mathcal{L}^{\text{SSIM}} \\ &= \arg \min_{\Theta} \|f(\mathbf{x}; \Theta) - \mathbf{y}\|_2^2 - \lambda \sum_{i=1}^Q \|(g_p(\hat{\mathbf{y}}) \odot \hat{\mathbf{x}}) * \mathcal{S}_{l_i}\|_2^2 \\ &\quad - \gamma \sum_{i=1}^Q \sum_{j=1}^Q \text{SSIM}(\mathcal{S}_{l_i}, \mathcal{S}_{r_j}) \end{aligned} \quad (9)$$

where $\widehat{\Theta} = \{\Theta, \mathbf{S}_l\} = \{\mathbf{W}, \mathbf{b}, \mathbf{S}_l\}$. In Eq. 9, $\mathcal{L}^{\text{Loss}}$ is the standard mean squared loss term between the ground-truth and output of the network.

Note that reference shape set \mathbf{S}_r is obtained using the Algorithm 1, that is \mathbf{S}_r is a pruned shape set obtained using the shape elimination procedure. The shapes that are actually learned are in the set \mathbf{S}_l , and therefore \mathbf{S}_l once optimized represents the refined version of the shapes in \mathbf{S}_r .

The shapes in set \mathbf{S}_l are learned according to the shape prior term (\mathcal{L}^{SP}) and the shape similarity measure ($\mathcal{L}^{\text{SSIM}}$): in particular, as argued before the shape prior term \mathcal{L}^{SP} should be minimized – so in essence, we are looking for the best shape set \mathbf{S}_l that can do so. Note that the second regularizer $\mathcal{L}^{\text{SSIM}}$ emphasizes similarity to the pruned reference shape set \mathbf{S}_r and therefore ensures that physically meaningful shapes are learned – unconstrained optimization of \mathbf{S}_l could lead to learned shapes that do not conform to reality.

C. Back-propagation: Shape Learning Cost Term

The back-propagation for TSP-CNN is similar to that of SP-CNN, except for the new shape learning term. The back-propagation for shape learning term only depends on the shapes \mathcal{S}_{l_i} . Let the cost function be $\widehat{\mathcal{L}} = \mathcal{L}^{\text{Loss}} + \mathcal{L}^{\text{SP}} + \mathcal{L}^{\text{SSIM}}$, hence its gradient w.r.t. $\mathcal{S}_{l_i}^a$, an arbitrary scalar entry within the shape \mathcal{S}_{l_i} , is as follows:

$$\begin{aligned} \frac{\partial \widehat{\mathcal{L}}}{\partial \mathcal{S}_{l_i}^a} &= -\lambda \sum_{i=1}^Q \langle (g_p(\hat{\mathbf{y}}) \odot \hat{\mathbf{x}}) * \mathcal{S}_{l_i}, (g_p(\hat{\mathbf{y}}) \odot \hat{\mathbf{x}}) * \mathbf{I} \rangle_F \\ &\quad - \gamma \sum_{j=1}^Q \frac{\partial}{\partial \mathcal{S}_{l_i}^a} \text{SSIM}(\mathcal{S}_{l_i}, \mathcal{S}_{r_j}) \end{aligned} \quad (10)$$

where \mathbf{I} denotes a matrix with same size as \mathcal{S}_{l_i} with only entrance a active as one. Second term in right hand side of Eq. 10 is derived in detail in [45], [44]. For a complete derivation of the back-propagation procedure, refer to our supplementary document in [29].

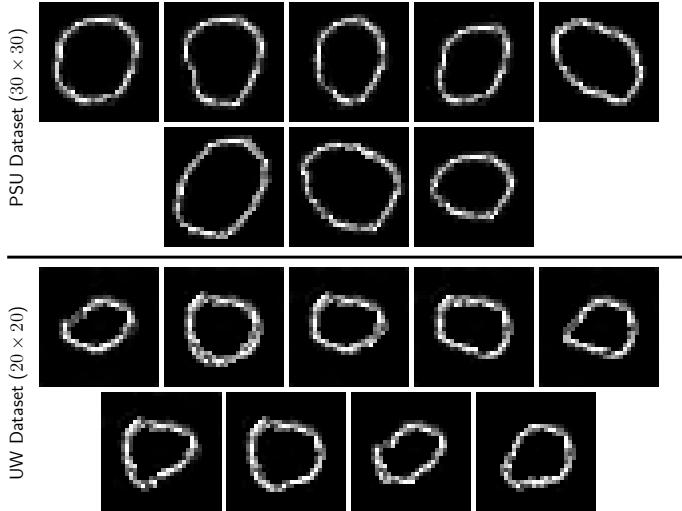


Fig. 7: Learned basis shapes using TSP-CNN for both UW and PSU datasets. Note that our optimization of shapes does not constrain them to be binary and hence these shapes have thicker boundaries with values in between 0 and 1.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup and Datasets

We train and test SP-CNN and TSP-CNN on two colon cancer datasets: 1) publicly available UW Dataset [5] which includes 100 H&E stained histology images of colorectal adenocarcinomas. There are a total number of 29756 nuclei marked at the nucleus center (please refer to Sec. VII.A. of [5] for more information). Our choice of the UW Dataset is because it represents real-world challenges such as overlapping nuclei and contains other shapes that are often confused with nuclei. Further, it is one of the few publicly available datasets that is widely used in many recent deep learning-based nuclei detection methods [5], [19], [46], 2) a new dataset carefully prepared and labeled manually by medical experts at the Center for Molecular Immunology and Infectious Disease, Penn State University. We call it the ‘PSU Dataset’ and it includes 120 images of colon tissue from 12 pigs at a resolution of $0.55\mu\text{m}/\text{pixel}$. Formalin fixed paraffin embedded pig colon sections were deparaffinized and stained with fluorescent DNA stain DAPI ($4',6\text{-diamidino-2-phenylindole}$) to visualize the cells as described in [47]. The selected images represent cross-sectional view of the colon epithelial cells. It also comprised of areas with artifacts, over-staining, and failed auto focusing, to represent outliers normally found in real scenarios. A total number of 25462 nuclei are annotated manually by an expert. For reproducing research results, we have made this dataset publicly available at the SP-CNN web-page [29]. Sample images from both datasets are shown in Fig. 8.

We construct $\mathbf{y} \in [0, 1]$ by processing the binary image of ground-truth nuclei center locations, which has 1 at the nucleus center and 0 elsewhere. This is accomplished by convolving the said ground-truth binary image with a zero mean Gaussian ($\sigma = 2$) filter of size 7×7 . Then the (luminance) input image, the raw edge image, and the labeled image form a training

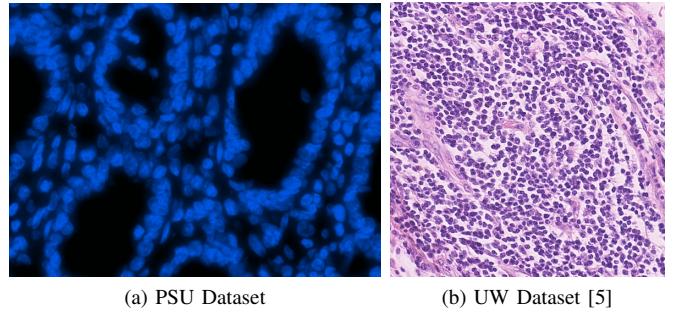


Fig. 8: Sample images from the datasets used for evaluation

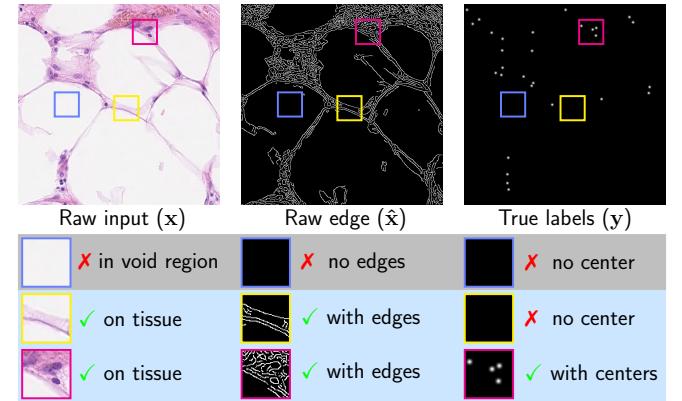


Fig. 9: Void training tuple ($\mathbf{x}, \hat{\mathbf{x}}, \mathbf{y}$); patches of size 40×40 are extracted and used for training. There are void regions in the raw image $\hat{\mathbf{x}}$, which do not include nuclei, as a result the label image \mathbf{y} is also void. These redundant void regions can mislead the network while wasting computation. To avoid this, we develop a procedure to eliminate the training patches which are empty in $(\mathbf{x}, \hat{\mathbf{x}}, \mathbf{y})$ – this is visually illustrated in Fig. 9.

tuple $(\mathbf{x}, \hat{\mathbf{x}}, \mathbf{y})$; patches of size 40×40 are extracted and used for training. There are void regions in the raw image $\hat{\mathbf{x}}$, which do not include nuclei, as a result the label image \mathbf{y} is also void. These redundant void regions can mislead the network while wasting computation. To avoid this, we develop a procedure to eliminate the training patches which are empty in $(\mathbf{x}, \hat{\mathbf{x}}, \mathbf{y})$ – this is visually illustrated in Fig. 9.

B. Assessment Metrics and Parameters

The output of SP/TSP-CNN is normalized as $\hat{\mathbf{y}} \in [0, 1]$, which is then processed via a thresholding operation with a pre-determined threshold T . Local maxima of the resulting thresholded image are identified as detected nuclei locations. To evaluate the detected locations against the true ones, we need some tolerance since it is unlikely that they will exactly match. This is handled in the literature by defining a golden standard region around each ground-truth nuclei center as described in [5]: we define this to be a region of 6 pixels around each nuclei center for UW Dataset and a region of 10 pixels for PSU Dataset (since the nuclei are larger in this dataset). Note that for fairness, the *same* ‘Golden Region’ is used across all methods that are compared in Section IV-D.

A detected nuclei location is considered to be true positive (TP), if it lies inside this region, otherwise it is considered to be false positive (FP), and the ones that are not matched by any of golden standard regions are considered to be false negative (FN). For quantitative assessment of SP-CNN

TABLE I: Effect of different parameters (PSU Dataset)

| SP-CNN | | | TSP-CNN ($p=11 \times 11$) | | | | |
|-------------|--------------|--------------|------------------------------|----------|--------------|--------------|--------------|
| λ | F1 | p | F1 | γ | F1 | λ | F1 |
| 1e-6 | 0.777 | 5x5 | 0.806 | 0.01 | 0.803 | 1e-6 | 0.824 |
| 5e-7 | 0.863 | 7x7 | 0.762 | 0.1 | 0.807 | 1e-8 | 0.818 |
| 1e-8 | 0.818 | 9x9 | 0.809 | 2 | 0.892 | 1e-10 | 0.892 |
| 1e-12 | 0.731 | 11x11 | 0.863 | 10 | 0.804 | 1e-12 | 0.812 |

TABLE II: The Configuration of The CNN Used in SP-CNN

| Layer No. | Layer Type | Filter Dimensions | Filter Numbers |
|-----------|---------------|------------------------|----------------|
| 1 | Conv. + ReLU | $5 \times 5 \times 1$ | 64 |
| 2 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 3 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 4 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 5 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 6 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 7 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 8 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 9 | Conv. + ReLU | $3 \times 3 \times 64$ | 64 |
| 10 | Convolutional | $3 \times 3 \times 64$ | 1 |

and TSP-CNN and comparison with other methods we use Precision (P), Recall (R), and F1 score ($F1$) which are

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, \text{ and } F1 = \frac{2PR}{P + R}. \quad (11)$$

To determine parameters for both networks, we use cross validation [48], which leads to the following choice⁷:

SP-CNN: $D = 7$ layers are used as validated in Section IV-C. These layers are with ‘SAME’ padding scheme; its configuration details are provided in Table. II. We use $N = 24$ different nuclei shapes in the domain expert provided shape set. Each shape is described by a 20×20 patch for UW Dataset [5], while for the PSU dataset, we use a 30×30 patch. The active part of the shapes are labeled as 1 and 0 otherwise. All the parameters used in SP-CNN are chosen by cross validation [48]. Most important of them are: trade-off values $\lambda = 5e-7$ and $\gamma = 0$, pooling window size $p = 11 \times 11$, weight decay parameter $= 1e - 5$, learning rate decay $= 0.75$, where the trade-off parameters are selected by cross-validation and the effect of these parameters is shown in Table I.

TSP-CNN: The parameters in TSP-CNN are same as SP-CNN except for the trade-off parameters which are $\lambda = 1e-10$ and $\gamma = 2$, and number of layers which is chosen (as per Section IV-C) to be $D = 10$. With this parameter setting, the Basis Shape Set using TSP-CNN for the UW-dataset converges to $Q = 9$ basis shapes, while for the PSU dataset $Q = 8$ shapes are determined. For both datasets, the optimized *Basis Shape Set* is visualized in Fig. 7.

C. Determining Number of Layers for the Network

In order to come up with the number of layers which suffices for accurate nuclei detection, we train the network with layers $D = 3$ to $D = 11$. The impact of number of layers is investigated in Fig. 10 for both SP-CNN and TSP-CNN. Fig. 10 reveals that $D = 10$ for TSP-CNN and $D = 7$

⁷During the training, both SP- and TSP-CNN network parameters are initialized using the same random seed to make for a fair comparison.

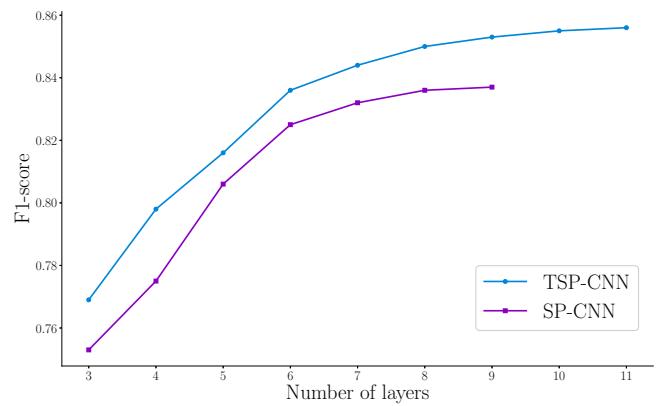


Fig. 10: F1-score for different number of layers: UW Dataset.

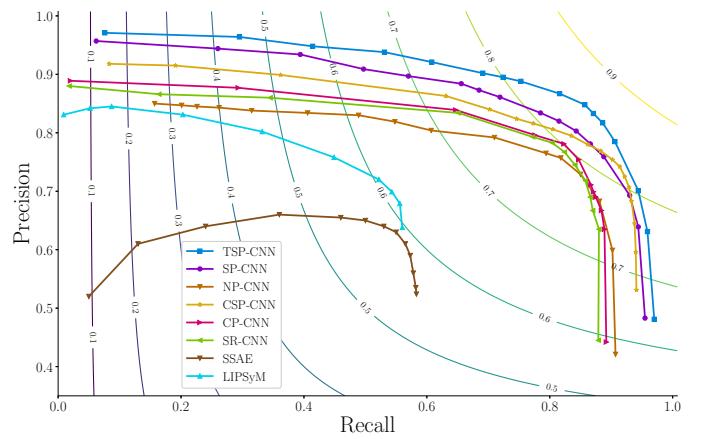


Fig. 11: Precision-recall curve for UW Dataset [5].

layers for SP-CNN facilitate the best computation-performance balance, i.e. more layers can mildly help but also increase computational burden. Other competing deep networks also employ 6-10 layers [5], [18].

TSP-CNN produces better results than SP-CNN even using shallower CNNs because the learned shapes adapt to the dataset. Note from Fig. 10 that a 6 layer TSP-CNN has the same level of performance as the 9 layer SP-CNN. Finally, while Fig. 10 is plotted for the UW Dataset, we found similar trends for the PSU dataset.

TABLE III: Nucleus detection results for UW Dataset [5]

| UW Dataset [5] | Precision | Recall | F1 score |
|----------------|-----------|--------|--------------|
| TSP-CNN | 0.848 | 0.857 | 0.852 |
| SP-CNN | 0.803 | 0.843 | 0.823 |
| WNP-CNN | 0.761 | 0.829 | 0.793 |
| NP-CNN | 0.757 | 0.818 | 0.786 |
| CSP-CNN [19] | 0.788 | 0.886 | 0.834 |
| SC-CNN [5] | 0.781 | 0.823 | 0.802 |
| SR-CNN [16] | 0.783 | 0.804 | 0.793 |
| SSAE [49] | 0.617 | 0.644 | 0.630 |
| LIPSYM [7] | 0.725 | 0.517 | 0.604 |
| CRImage [30] | 0.657 | 0.461 | 0.542 |

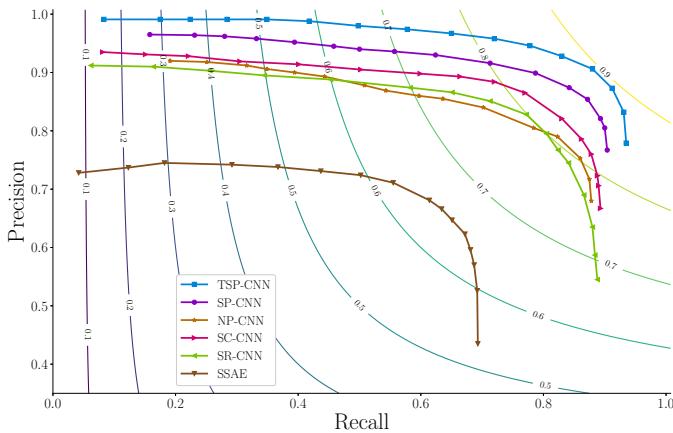


Fig. 12: Precision-recall curve for PSU Dataset.

TABLE IV: Nucleus detection results for PSU dataset

| PSU Dataset | Precision | Recall | F1 score |
|-------------|-----------|--------|--------------|
| TSP-CNN | 0.874 | 0.911 | 0.892 |
| SP-CNN | 0.854 | 0.871 | 0.863 |
| NP-CNN | 0.746 | 0.859 | 0.799 |
| SC-CNN [5] | 0.821 | 0.830 | 0.825 |
| SR-CNN[16] | 0.797 | 0.805 | 0.801 |
| SSAE[49] | 0.665 | 0.634 | 0.649 |

D. Comparison With State-of-the-Art

As is common, the threshold value T is varied to generate a Precision-Recall curve. The Precision-Recall curve for averaged values over all images in the test set from UW Dataset [5] is plotted in Fig. 11, and for PSU Dataset the PR-curve is plotted in Fig. 12. For consistency, results for UW Dataset for the proposed SP-CNN and TSP-CNN are based on using the same assessment procedure as in [5], which used a 50-50 split of training vs. test images (using the official assessment source codes of [5] provided by the paper's author). For PSU Dataset, a 100-20 split of training vs. test images was employed by all competing methods. Fig. 11 essentially compares SP-CNN and TSP-CNN against state of the art deep learning methods: 1.)

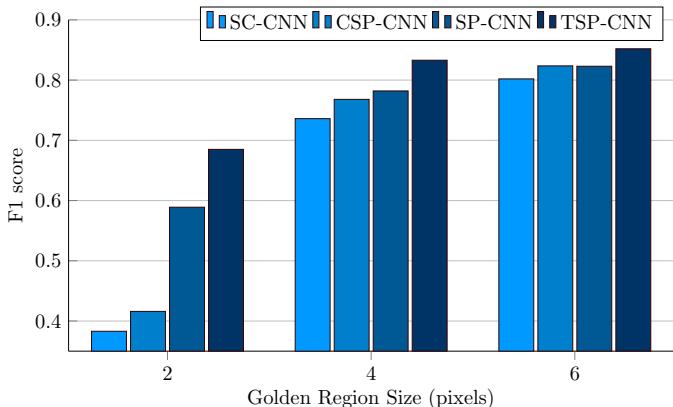


Fig. 13: F1 scores for the top competing methods over varying choice of the Golden Region for the UW Dataset [5].

SC-CNN [5], 2.) SR-CNN [16], 3.) SSAE [49], 4.) CSP-CNN [19] and two other popular feature and morphology based methods 5.) LIPSSyM [7], and 6.) CRImage [30].

In the case of UW Dataset, P , R , and $F1$ results for the aforementioned competing methods are obtained directly from comparisons already reported in [5], [19]. Figures 11 and 12 reveal that SP-CNN and TSP-CNN achieve Precision-Recall curves that outperform state of the art alternatives. To obtain a single representative figure of merit for each method, we chose the threshold value, T , which maximizes the F1-score correspondingly for each method. These ‘best F1-scores’ for each method are then reported in Table III for the UW Dataset [5]. In Table IV, ‘best F1-scores’ are reported for the PSU Dataset. Note that in Table IV and Fig. 12, we focus on deep learning methods only because: 1.) they are shown to comfortably outperform traditional feature and morphology based methods and 2.) the particular methods in [7], [30] employ features that are not really appropriate for the PSU dataset leading to vastly degraded results. Further, comparisons with CSP-CNN [19] are not reported for the PSU dataset because CSP-CNN is highly customized for the UW dataset.

To show the value of prior guided regularization, we also report in Tables III and IV, the results of our network for the case with $\lambda = \gamma = 0$ in Eq. (9). Because no regularizers are involved, we call this – No Prior CNN (NP-CNN). The case with only $\gamma = 0$ will reduce Eq. (9) to Eq. (3) and hence corresponds to SP-CNN. Consistent with SP-CNN, the aforementioned NP-CNN also uses $D = 7$ layers. The quantitative gains of using shape priors and learning shapes, i.e. (T)SP-CNN vs. NP-CNN are readily apparent in Tables III and IV. Similar trends can also be seen in Figs. 11 and 12 in terms of the benefits of shape priors.

As an alternate strategy to address false positives, we perform an experiment which involves training a (prior-less) network with a weighted cost function. That is, we divide the Mean Square Error (MSE) term in Eq. (1) into two terms: 1) MSE between ground-truth and false positive detections, and 2) MSE that results from missed detections, i.e the loss from false negative. During training, in every iteration (containing two training pairs: input image and corresponding ground-truth labels), false positives and negatives are determined using the ground-truth and subsequently a weighted MSE is computed with a larger weight assigned to false positives⁸ – we call this design Weighted NP-CNN (WNPCNN). We train WNP-CNN with exactly the same network setting and parameters as NP-CNN. The detection results are presented in Table III which shows that WNP-CNN with a more complicated training process can achieve modest gains over NP-CNN but is still outperformed by SP/TSP-CNN.

Note that the results reported in Table III and Fig. 11 use the same Golden Region specified in Section IV-B: 6 pixels for the UW Dataset consistent with existing literature [5]. To investigate the effect of Golden Region selection, we report F1 scores generated by selecting different Golden Region sizes for the UW dataset in Fig. 13. We select the SC-, CSP-,

⁸The best weights for these two terms were found using cross-validation [48] to be 0.7 and 0.3, respectively.

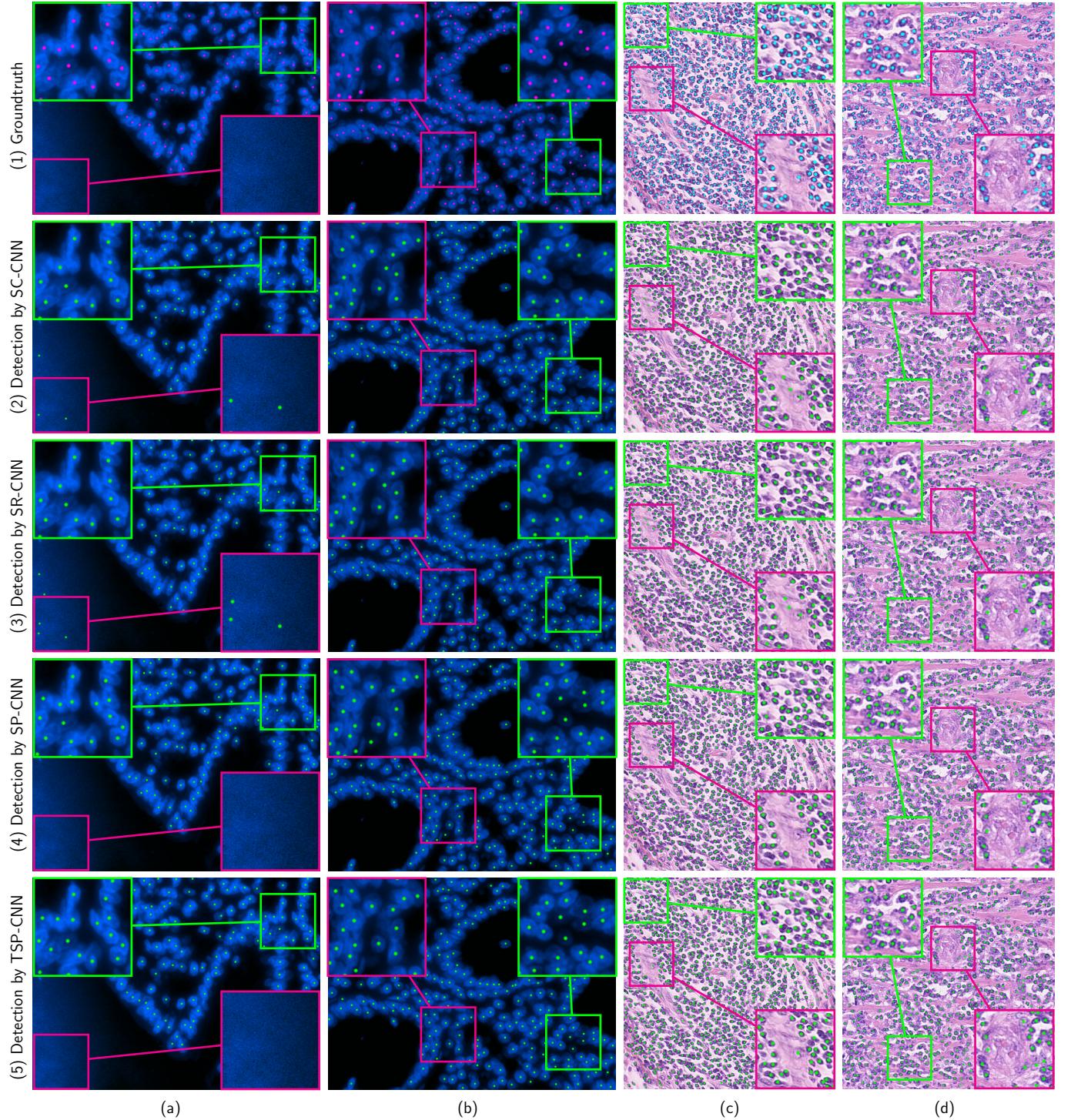


Fig. 14: Ground-truth nuclei centers and detection results for two images from each dataset presented here. While ‘green’ zoomed areas show the missed detection by SC-CNN and SR-CNN, ‘magenta’ ones show the wrong detection (FP) of nuclei by those methods. F1-scores for them are as follows: a.2) 0.815, a.3) 0.809, a.4) 0.887, a.5) 0.920; b.2) 0.906, b.3) 0.895, b.4) 0.912, b.5) 0.926; c.2) 0.830, c.3) 0.801, c.4) 0.908, c.5) 0.948; d.2) 0.901, d.3) 0.898, d.4) 0.909, d.5) 0.949.

SP- and TSP-CNN to conduct the experiments as they are the top-4 methods from Table III. From Fig. 13, as expected, F1 score increases with an increase in the golden region size. However, for a fixed Golden Region size, it is the relative performance of different methods that matters and as is evident from Fig. 13, this trend remains unchanged. In fact, with a

smaller Golden Region such as 2, the SC-CNN and CSP-CNN performance degrades heavily even as the SP-CNN and TSP-CNN exhibit a much more graceful decay emphasizing higher spatial accuracy of (T)SP-CNN.

Figure 14 provides further insight into the merits of SP-CNN and TSP-CNN for two test images from each dataset.

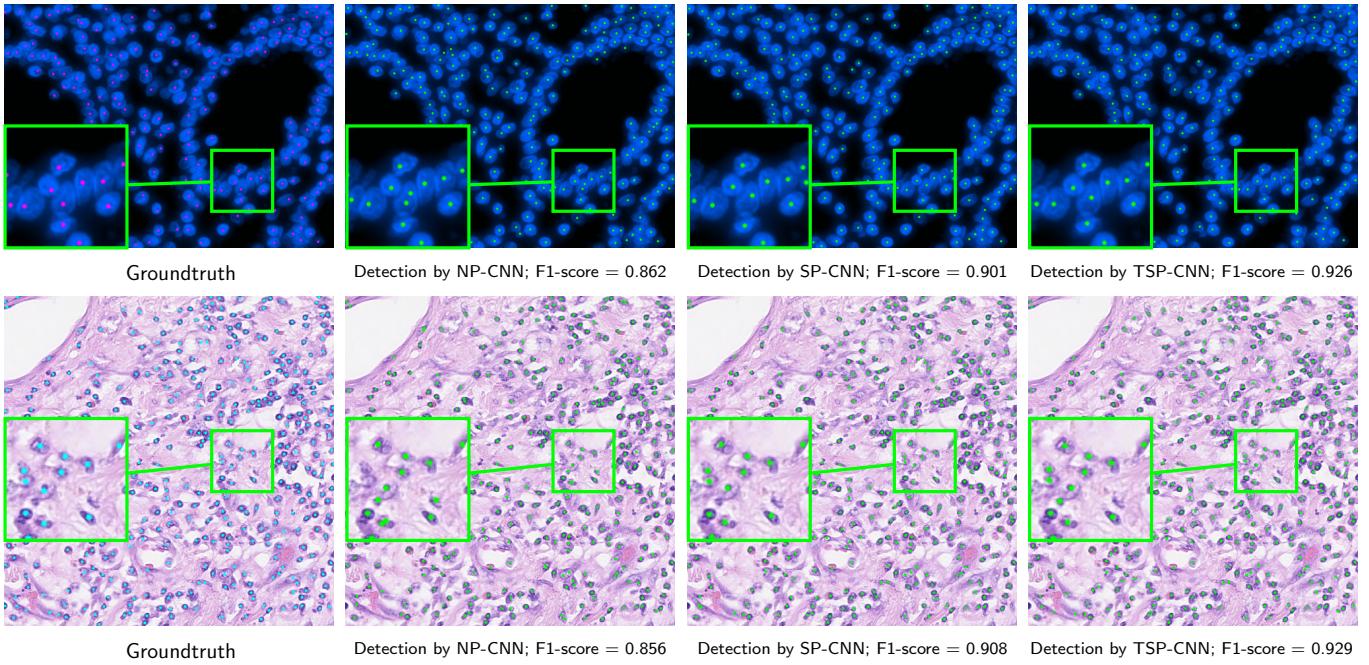


Fig. 15: Nucleus detection results for TSP-CNN, SP-CNN ($\gamma = 0$), and NP-CNN ($\gamma = \lambda = 0$) for an example image from the PSU Dataset as well as the UW dataset. The value of prior guided regularization is evident in that false positives are significantly reduced in both SP and TSP-CNN. NP-CNN detects two cells in a region with only one cell (PSU dataset image), and in the example image from UW dataset, there are two missed detections plus one non-cell texture is detected as nuclei.

We compare TSP-CNN and SP-CNN with the top two methods from Table III and Table IV, i.e. SC-CNN [5] and SR-CNN[16]. Two parts of each image in Fig. 14 are magnified for convenience. While ‘green’ zoomed areas show the missed detection by SC-CNN and SR-CNN, ‘magenta’ ones show the wrong detection (FP) of nuclei by those methods. Thanks to learning guided by pertinent prior information, SP-CNN and TSP-CNN avoid false positives, which are still detected as nuclei by competing state of the art methods.

In the same spirit, we present results for TSP-CNN, SP-CNN, and NP-CNN in Fig. 15. In this figure, we can observe that shape priors indeed help to improve the detection results and tunable shapes improves the performance of SP-CNN even further. In Fig. 15, note that the detections made by NP-CNN, SP-CNN and TSP-CNN nearly capture all the nuclei centers for the example image from the PSU dataset. Yet, SP-CNN and TSP-CNN perform better because guided by shape priors, false positives are significantly reduced. For the UW dataset, benefits of SP/TSP-CNN are evident over NP-CNN for which two missed detections and one false positive can be seen in the zoomed part of the image.

V. DISCUSSION AND CONCLUSION

Our contribution is a prior-guided deep network that can enhance cell nuclei detection significantly. Analytically, we develop methods for tractable incorporation of nuclei shape priors (provided by a domain expert) as regularizers in the learning. We also extend the said framework to allow the shapes to be learned from the same training samples instead of assuming them to be fixed. Experimentally, higher F1 scores

as well as superior precision-recall curves are achieved by the proposed prior guided networks.

In this work, we used SSIM based shape similarity to facilitate learning of meaningful shapes largely for tractability reasons. While there are other known similarity measures that can be better than SSIM for comparing nuclei shapes, they are not differentiable w.r.t the underlying shapes and hence the network parameters. A significant direction for future work is the design of custom, new shape similarity measures that can be effective in comparing nuclei shapes while being analytically tractable for incorporation in deep learning frameworks. Multi-scale extension of our prior guided framework forms another viable future research direction.

REFERENCES

- [1] M. Al-Hajj, M. S. Wicha, A. Benito-Hernandez, S. J. Morrison, and M. F. Clarke, “Prospective identification of tumorigenic breast cancer cells,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 7, pp. 3983–3988, 2003.
- [2] M. E. Prince, R. Sivanandan, A. Kaczorowski, G. T. Wolf, M. J. Kaplan, P. Dalerba, I. L. Weissman, M. F. Clarke, and L. E. Ailles, “Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 3, pp. 973–978, 2007.
- [3] S. V. Costes, D. Daelemans, E. H. Cho, Z. Dobbin, G. Pavlakis, and S. Lockett, “Automatic and quantitative measurement of protein-protein colocalization in live cells,” *Biophysical journal*, vol. 86, no. 6, pp. 3993–4003, 2004.
- [4] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, et al., “Fiji: an open-source platform for biological-image analysis,” *Nature methods*, vol. 9, no. 7, pp. 676, 2012.
- [5] K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, D. R. J. 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, May 2016.
- [6] P. S. Mitchell, R. K. Parkin, E. M. Kroh, B. R. Fritz, S. K. Wyman, E. L. Pogosova-Agadjanyan, A. Peterson, J. Noteboom, K. C. O'Briant, A. Allen, et al., “Circulating microRNAs as stable blood-based markers for cancer detection,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 30, pp. 10513–10518, 2008.
- [7] M. Kuse, Y. Wang, V. Kalasannavar, M. Khan, and N. Rajpoot, “Local isotropic phase symmetry measure for detection of beta cells and lymphocytes,” *Journal of pathology informatics*, vol. 2, 2011.
- [8] M. Veta, J. P. W. Pluim, P. J. Van Diest, and M. A. Viergever, “Breast cancer histopathology image analysis: A review,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 5, pp. 1400–1411, 2014.
- [9] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, “Improved automatic detection and segmentation of cell nuclei in histopathology images,” *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 841–852, 2010.
- [10] S. Ali and A. Madabhushi, “An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery,” *IEEE transactions on medical imaging*, vol. 31, no. 7, pp. 1448–1460, 2012.
- [11] D. Shen, G. Wu, and H.-Il Suk, “Deep learning in medical image analysis,” *Annual review of biomedical engineering*, vol. 19, pp. 221–248, 2017.
- [12] G. Litjens, Th. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, and C. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [13] Y. Xie, F. Xing, X. Shi, X. Kong, H. Su, and L. Yang, “Efficient and robust cell detection: A structured regression approach,” *Medical image analysis*, vol. 44, pp. 245–254, 2018.
- [14] S. Ram, V. T. Nguyen, K. H. Limesand, and M. R. Sabuncu, “Joint cell nuclei detection and segmentation in microscopy images using 3d convolutional networks,” *arXiv preprint arXiv:1805.02850*, 2018.
- [15] A. A. Cruz-Roa, J. E. Arevalo Ovalle, A. Madabhushi, and F. A. González Osorio, “A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection,” *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pp. 403–410, 2013.
- [16] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang, “Beyond classification: Structured regression for robust cell detection using convolutional neural network,” *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 358–365, 2015.
- [17] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, “Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119–130, Jan 2016.
- [18] F. Xing, Y. Xie, and L. Yang, “An automatic learning-based framework for robust nucleus segmentation,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 550–566, Feb 2016.
- [19] L. Hou, V. Nguyen, A. B. Kanevsky, D. Samaras, T. M. Kurc, T. Zhao, R. R. Gupta, Y. Gao, W. Chen, D. Foran, et al., “Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images,” *Pattern Recognition*, vol. 86, pp. 188–200, 2019.
- [20] Y. Song, E.-L. Tan, X. Jiang, J.-Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang, “Accurate cervical cell segmentation from overlapping clumps in pap smear images,” *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 288–300, 2017.
- [21] X. Pan, L. Li, H. Yang, Z. Liu, J. Yang, L. Zhao, and Y. Fan, “Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks,” *Neurocomputing*, vol. 229, pp. 88–99, 2017.
- [22] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, “Statistical shape influence in geodesic active contours,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2000, vol. 1, pp. 316–323.
- [23] S. Ali and A. Madabhushi, “An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 7, pp. 1448–1460, July 2012.
- [24] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, S. A. Cook, A. de Marvao, T. Dawes, D. P. O'Regan, et al., “Anatomically constrained neural networks (acnn): application to cardiac image enhancement and segmentation,” *IEEE transactions on medical imaging*, vol. 37, no. 2, pp. 384–395, 2018.
- [25] H. S. Mousavi, T. Guo, and V. Monga, “Deep image super resolution via natural image priors,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018.
- [26] V. Cherukuri, T. Guo, S. J. Schiff, and V. Monga, “Deep mr image super-resolution using structural priors,” in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2018, pp. 410–414.
- [27] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [28] M. Tofighi, T. Guo, J. K. P. Vanamala, and V. Monga, “Deep networks with shape priors for nucleus detection,” in *IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 719–723.
- [29] “Shape Priors With Convolutional Neural Networks for Cell Nucleus Detection,” <http://signal.ee.psu.edu/research/SPCNN.html>.
- [30] Y. Yuan *et al.*, “Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling,” *Science translational medicine*, vol. 4, no. 157, pp. 157ra143–157ra143, 2012.
- [31] L. Kang, P. Ye, Y. Li, and D. Doermann, “Convolutional neural networks for no-reference image quality assessment,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 1733–1740.
- [32] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [33] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on pattern analysis and machine intelligence*, , no. 6, pp. 679–698, 1986.
- [34] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [35] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [36] R. C. Veltkamp, “Shape matching: Similarity measures and algorithms,” in *Shape Modeling and Applications, SMI 2001 International Conference on*. IEEE, 2001, pp. 188–197.
- [37] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, “Complex wavelet structural similarity: A new image similarity index,” *IEEE transactions on image processing*, vol. 18, no. 11, pp. 2385–2401, 2009.
- [38] E. Attalla and P. Siy, “Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching,” *Pattern Recognition*, vol. 38, no. 12, pp. 2229–2241, 2005.
- [39] L. J. Latecki and R. Lakamper, “Shape similarity measure based on correspondence of visual parts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1185–1190, 2000.
- [40] G. Lu and A. Sajjanhar, “Region-based shape representation and similarity measure suitable for content-based image retrieval,” *Multimedia Systems*, vol. 7, no. 2, pp. 165–174, 1999.
- [41] B. J. Super, “Learning chance probability functions for shape retrieval or classification,” in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. IEEE, 2004, pp. 93–93.
- [42] N. Ueda and S. Suzuki, “Learning visual models from shape contours using multiscale convex/concave structure matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 337–352, 1993.
- [43] Z. Wang and A. C. Bovik, “Mean squared error: love it or leave it? - a new look at signal fidelity measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan. 2009.
- [44] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, “Loss functions for image restoration with neural networks,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.
- [45] Z. Wang and E. P. Simoncelli, “Maximum differentiation (mad) competition: A methodology for comparing computational models of perceptual quantities,” *Journal of Vision*, vol. 8, no. 12, pp. 8–8, 2008.
- [46] A. Ahmad, A. Asif, N. Rajpoot, M. Arif, and F. Minhas, “Correlation filters for detection of cellular nuclei in histopathology images,” *Journal of Medical Systems*, vol. 42, no. 1, Nov 2017.
- [47] V. Charepalli, L. Reddivari, S. Radhakrishnan, E. Eriksson, X. Xiao, S. W. Kim, F. Shen, V. K. Matam, Q. Li, V. B. Bhat, R. Knight, and J. K. P. Vanamala, “Pigs, unlike mice, have two distinct colonic stem cell populations similar to humans that respond to high-calorie diet prior to insulin resistance,” *Cancer Prevention Research*, 2017.
- [48] V. Monga, *Handbook of Convex Optimization Methods in Imaging Science*, Springer, 2017.
- [49] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, “Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119–130, Jan 2016.