Softmax-Driven Active Shape Model for Segmenting Crowded Objects in Digital Pathology Images

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 19 January 2024, accepted 19 February 2024, date of publication 26 February 2024, date of current version 1 March 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3369916*

RESEARCH ARTICLE

# Softmax-Driven Active Shape Model for Segmenting Crowded Objects in Digital Pathology Images

**MASSIMO SALVI, KRISTEN M. MEIBURGER, (Member, IEEE), AND FILIPPO MOLINARI, (Senior Member, IEEE)**

Biolab, PolitoBIOMed Lab, Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy

Corresponding author: Massimo Salvi (massimo.salvi@polito.it)

This work was supported in part by the Public-Private Partnership (PPP) Calls for Photonics: H2020-ICT-2020-2 under Grant ID 101016964 (REAP).

**ABSTRACT** Automated segmentation of histological structures in microscopy images is a crucial step in computer-aided diagnosis framework. However, this task remains a challenging problem due to issues like overlapping and touching objects, shape variation, and background complexity. In this work, we present a novel and effective approach for instance segmentation through the synergistic combination of two deep learning networks (detection and segmentation models) with active shape models. Our method, called softmax-driven active shape model (SD-ASM), uses information from deep neural networks to initialize and evolve a dynamic deformable model. The detection module enables treatment of individual objects separately, while the segmentation map precisely outlines boundaries. We conducted extensive tests using various state-of-the-art architectures on two standard datasets for segmenting crowded objects like cell nuclei - MoNuSeg and CoNIC. The proposed SD-ASM consistently outperformed reference methods, achieving up to 8.93% higher Aggregated Jaccard Index (AJI) and 9.84% increase in Panoptic Quality (PQ) score compared to segmentation networks alone. To emphasize versatility, we also applied SD-ASMs to segment hepatic steatosis and renal tubules, where individual structure identification is critical. Once again, integration of SD-ASM with deep models enhanced segmentation accuracy beyond prior works by up to 6.2% in AJI and 38% decrease in Hausdorff Distance. The proposed approach demonstrates effectiveness in accurately segmenting touching objects across multiple clinical scenarios.

**INDEX TERMS** Digital pathology, deep learning, hybrid frameworks, nuclei instance segmentation, active shape models.

## I. INTRODUCTION

Cellular microscopy imaging methods are essential for evaluating and quantifying tissue characteristics like cell integrity, viability, and disease state. Some of the main cellular imaging methods include phase-contrast microscopy, fluorescent microscopy, brightfield microscopy and confocal microscopy [1], [2]. These imaging methods often aim to provide an analysis of a single object, which can often be hindered by the fact that cellular structures overlap and are not easily separable [3], [4], [5].

One of the main medical applications of brightfield microscopy is histopathological image analysis, where clinicians analyze images of explanted tissue from biopsies or surgeries, which is fundamental for cancer diagnosis and grading. Digital pathology has rapidly grown over the last decade thanks mainly to slide scanners which are now able to produce digital super-resolution whole slide images (WSI) [6] that can be viewed on a traditional monitor as an alternative to the use of a conventional microscope. Digital pathology presents a wide range of potential applications

The associate editor coordinating the review of this manuscript and approving it for publication was György Eigner.

such as, but not limited to: normal and pathological cell counting, region or object segmentation, tissue structure recognition, cancer grade classification, cancer prognosis, etc.

An accurate object detection is fundamental for cellular-level analysis in computational pathology. In a clinical setting, quantitative analysis of cell morphology is useful for diagnostic decision making, prediction of treatment outcome, and prognostic stratification. Precise segmentation of cellular object boundaries is also important for grading cancer, from which clinicians can obtain more pathological information in clinical medicine and concurrently analyze the details of spatial characteristics of tissues and targets. Segmentation of individual histological structures is a very challenging task due to large variations in morphology of the object to be recognized as well as the existence of touching and poorly defined structures in pathological conditions.

Deep learning-based methods have become the state-of-the-art computational methods for medical image segmentation, due to their dominating performance in numerous computer vision tasks, such as object classification, detection, and segmentation. In particular, Convolutional Neural Networks (CNN) present exceptional feature extraction and generalization capabilities for image segmentation tasks, which is one of the most common and relevant challenge for medical image analysis.

Regarding the specific task of segmentation in histopathological images, there are two main challenges:

- Stain variability: during tissue preparation, a crucial step is the staining process, where dyes are applied to highlight specific cellular components. However, this process introduces significant variability, resulting in histological slides with the same stain appearing different in terms of color and intensity. A solution to this challenge is known as stain normalization, in which stain-normalized images have their intensity distribution mapped to match the color distribution of a chosen template image [7].
- Overlapping and touching structures: histopathological images often contain structures that overlap or touch each other, making it difficult to distinguish them as individual objects. This difficulty is exacerbated when using deep learning techniques, as the encoder and decoder of the CNN tend to lose some image details, making the segmentation of touching objects more challenging. Additionally, boundaries of histological structures may become blurred during the staining and scanning process, further complicating the segmentation task.

Despite the various benefits offered by AI methods, deep learning techniques are still mostly viewed as black-box systems, making it challenging to understand the factors that control their performance. Specifically, the semantic segmentation performed by CNNs usually shows very high pixel-level accuracy but is prone to merge different objects that are close to each other or intersect [8]. Rather than developing a new network architecture, this work presents a hybrid framework that integrates deep learning segmentation models with traditional deformable models. By combining semantic predictions with localized shape refinement, our goal is to overcome limitations in current methods for refining instance segmentation. The key contributions of this paper can be summarized as follows:

- We introduce a novel formulation of softmax-driven active shape models (SD-ASMs). These dynamic deformable models leverage object localization and semantic content provided by two deep networks. Specifically, the bounding boxes generated by a detection network are used to initialize each ASM. Additionally, the probability map from a segmentation network is employed to balance the internal and external forces of each active shape model. SD-ASMs evolve directly on the probability map provided by the segmentation network, enabling accurate instance segmentation.
- We propose a hybrid deep learning framework that combines the accuracy of active shape models (ASMs) with the localization and semantic content created by two deep networks. Our dynamic deformable models excel at accurately separating individual cell instances by leveraging the detection and the segmentation modules. The controllability of these hybrid ASMs allows to overcome the limitations of current deep learning methods.
- To address the challenge of touching and partial overlapping nuclei, we incorporate a repulsive interaction term in the mathematical formulation of each ASM. This interaction term is introduced to guide the evolution of each ASM and effectively mitigate the issue of ''partial overlap''.
- We conduct an extended validation by integrating the proposed approach with state-of-the-art techniques for nuclei segmentation in hematoxylin and eosin (H&E) stained images. Results demonstrate that the introduction of SD-ASMs inside existing methods provide a significant boost in instance recognition. Additionally, we quantitatively assess generalizability on two clinical applications - liver steatosis and kidney tubule segmentation. In both cases, SD-ASM integration with deep models enhances segmentation accuracy beyond prior works.

The rest of this paper is organized as follows: Section II presents an overview of the current approach for segmenting individual instances of histological structures; Section III provides an exhaustive description of the proposed approach; Sections IV and V report and discuss the experimental results.

## II. RELATED WORKS

In this section, we provide a concise overview of recent advancements in instance segmentation for H&E-stained histopathology images. While some existing methods rely on traditional image processing techniques such as thresholding, clustering, and edge detection [9], [10], our focus here is

primarily on deep learning techniques, which have emerged as the state-of-the-art in this field.

One of the earliest applications of deep learning in histopathology involved nuclei segmentation, a crucial step in cancer analysis and grading [11]. Veta et al. [11] proposed a spatially constrained CNN that operates on the likelihood of a pixel being the center of a nucleus. Pan et al. [12] introduced a method based on sparse reconstruction to remove background and enhance nuclei, followed by refining the CNN segmentation using morphological operations and prior knowledge. Lal et al. [13] presented NucleiSegNet, which consists of robust residual, bottleneck, and attention decoder blocks. While these studies showed promising results, they did not specifically address the challenge of overlapping nuclei.

Other studies have specifically tackled the issue of overlapping nuclei in histology images. Xing et al. [14] combined a two-class CNN for generating probability maps with sparse shape models to separate touching nuclei. Naylor et al. [15] tackled the problem of overlapping nuclei by regressing the distance map and employing post-processing methods. Graham and Rajpoot [16] introduced a CNN method for simultaneous nuclear segmentation and classification, utilizing pixel distances to separate clustered nuclei. Cui et al. [17] proposed a contour-aware CNN that outputs estimated nuclei and boundary maps. Xie et al. [3] employed the Mask R-CNN model and introduced structure preserving color normalization along with a watershed algorithm to separate fused nuclei, which however tended to produce irregular nuclei shapes, which is detrimental to the subsequent morphological analysis of nuclei. Wan et al. [18] proposed a technique based on the integration of an atrous spatial pyramid pooling U-Net (ASPPU-Net) and an improved concave point detection algorithm to segment nuclei with high overlap. Chanchal et al. [19] introduced an encoder–decoder architecture with a separable convolution pyramid pooling network (SCPP-Net), which focuses on increasing the receptive field. The authors do not present any specific approach for the problem of overlapping nuclei but claim that the increase of the receptive field at a higher level helps overcome the issue. Liu et al. [20] proposed the MDC-Net, incorporating distance maps and contour information for overlapping nuclei segmentation. Wazir and Fraz [21] utilized a quick attention module and multi-loss function to capture relevant local and global features at multiple scales, enabling more precise boundary detection. Vo et al. [22] integrated multiple filter blocks into their Hover-Net architecture. Testing showed these blocks enhanced the original model's performance [23] for instance segmentation tasks. Additionally, Saednia et al. [24] employed a cascaded deep learning framework with a customized pixel-wise weighted loss function. This allowed their approach to segment nuclei with higher accuracy by optimizing the loss calculation at each stage.

Another application in which it is critical to recognize individual instances of histological structures involves the analysis of tubules and glands. Their morphology is routinely used by experienced pathologists to assess the degree of cancer malignancy in various epithelial tissues such as the prostate, breast, and colon [25], [26]. To obtain reliable morphologic statistics for quantitative diagnosis, individual tubules/glands must be precisely segmented. Qu et al. [27] proposed a three-class CNN to segment touching glands, while Xu et al. [28] combined three different CNNs (foreground segmentation, edge detection, and single gland detection) for gland instance generation. Chen et al. [25] introduced a deep contour-aware network that depicted gland contours and produced probability maps, addressing touching structures using upsampling and morphological operators. Zhang et al. [29] proposed the SGDANet model for gland instance segmentation, which incorporates a geometric module to capture cell shapes along with a parallel spatial-geometric dual-path attention mechanism for enhanced feature extraction.

It is evident that the segmentation of overlapping and touching histological structures remains an open problem. In this work, we propose a novel approach that synergistically combines two deep learning networks with active shape models. We demonstrate the effectiveness of our proposed method through comparative evaluations with other competitive state-of-the-art instance segmentation models across different tissue types.

## III. MATERIALS AND METHODS

In this paper, we present a fully supervised method for instance segmentation in histopathological images. Fig. 1 illustrates the flowchart of our approach. Our method consists of three steps: i) image normalization, ii) object localization & semantic content extraction, iii) softmax-driven active shape models. We will publicly release the dataset and codes used in this study at the following doi: 10.17632/nwnfh5x5j3.1. A detailed description of our approach is provided in the next section.

### A. IMAGE NORMALIZATION

The first step of the proposed pipeline is image normalization. Stain normalization is a common pre-processing step in almost all deep learning frameworks for digital pathology [30], [31], [32]. Briefly, a stain normalizing procedure allows to standardize the color appearance of a source image with respect to the color profile of a template image. This operation reduces the stain variability and improves the robustness of computer-aided diagnostic and image quantification algorithms [33].

In this work, we employed a GAN (Generative Adversarial Network) [34] to address the task of stain normalization. The goal is to translate the color pattern of images from one domain (domain A) to the color pattern in another domain (domain B), where domain A exhibits a wide range of color patterns and domain B has a relatively uniform color pattern [35]. Our generative model is a Pix2Pix GAN consisting of a U-net architecture as the generator (G) and a three-layer fully convolutional PatchGAN as the discriminator (D). The

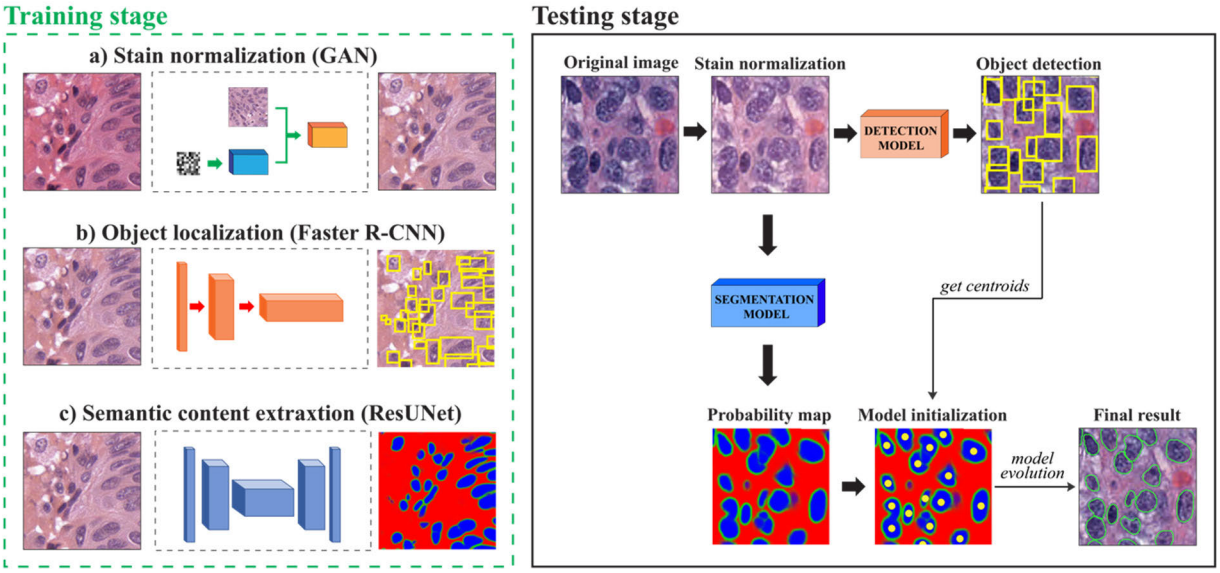



**FIGURE 1.** Overview of the proposed instance segmentation framework. During the training phase, three different deep networks are developed: a generative adversarial network (GAN) for stain normalization, a faster R-CNN for object localization, and a DeepLabv3+ for semantic content extraction. During inference, the outputs of these three networks are combined to initialize and evolve the softmax-driven ASMs.

generator aims to generate normalized images that can fool the discriminator into classifying them as real target images, while the discriminator aims to correctly distinguish between the real and generated images. A pixel-wise loss is used to preserve the pixel-level correspondence between the original and normalized images. The GAN is trained using a least square objective function [36], which is defined as follows:

$$\min_{\mathbf{D}} \mathbf{V}_{\mathbf{LSGAN}}(\mathbf{D}) = \frac{1}{2}\mathbf{E}_{\mathbf{x},\mathbf{y}}\left[\mathbf{D}(\mathbf{x}) - \mathbf{b}^2\right] + \frac{1}{2}\mathbf{E}_{\mathbf{x},\mathbf{z}}\left[(\mathbf{D}(\mathbf{x}, \mathbf{G}(\mathbf{z})) - \mathbf{a})^2\right] \quad (1)$$

$$\min_{\mathbf{G}} \mathbf{V}_{\mathbf{LSGAN}}(\mathbf{G}) = \frac{1}{2}\mathbf{E}_{\mathbf{x},\mathbf{z}}\left[(\mathbf{D}(\mathbf{x}, \mathbf{G}(\mathbf{x}, \mathbf{z})) - \mathbf{c})^2\right] \quad (2)$$

Here, a and b represent the labels for fake and real data, respectively, and c denotes the value that G aims for D to perceive as fake data.

The training process continues until convergence, during which the generator network learns to produce high-quality normalized images that closely resemble the target images. Our model is trained for 150 epochs using instance normalization. The initial learning rate is set to 0.0001 and reduced by a factor of 0.5 every 50 epochs. Fig. 2 illustrates the effect of the stain normalization procedure on three sample images.

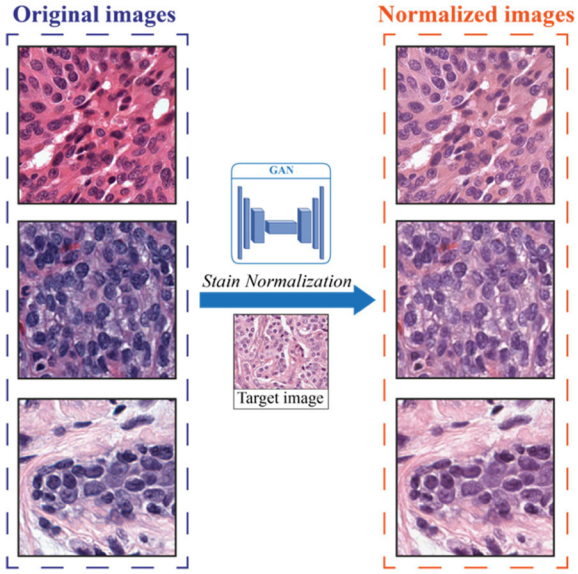



**FIGURE 2.** Stain normalization performed by the generative adversarial network (GAN) for three tiles. The generative model is capable of normalizing and reducing the color variability of histological images.

### B. OBJECT LOCALIZATION AND SEMANTIC CONTENT EXTRACTION

In this study, a dual-branch deep learning approach is used to both localize and perform an initial segmentation of individual cells. The approach works by first decomposing the training images into overlapping patches. These patches are then input to separate detection and segmentation networks, as depicted in Fig. 3. The figure provides an overview of the localization strategy employed, which aims to extract relevant semantic content from each input patch for downstream analysis.

The detection branch is performed by a Faster R-CNN using the detectron2 library [37]. The Faster R-CNN is a region proposal network that simultaneously predicts object boundaries and scores at each position. This detection network consists of two modules: the first module is a deep fully

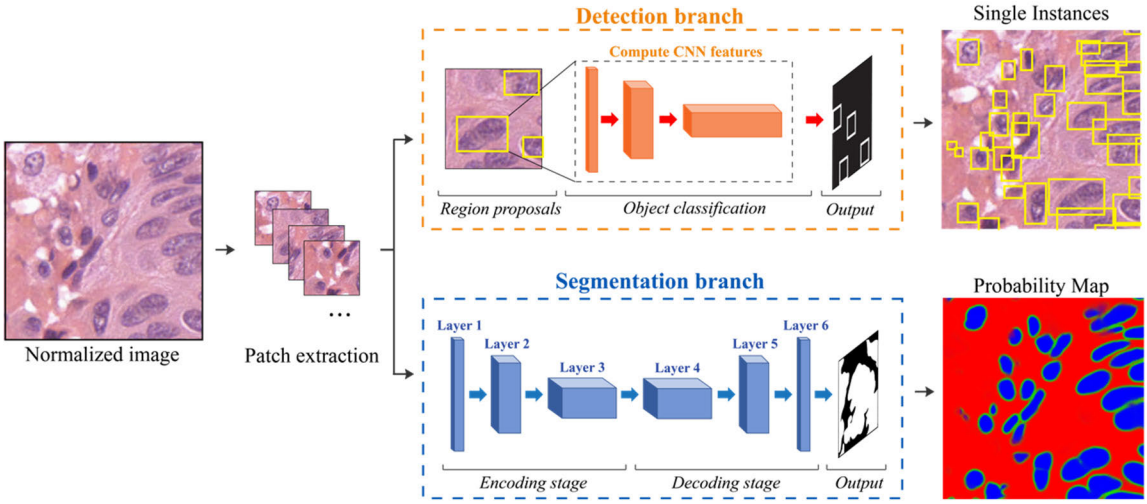



**FIGURE 3.** Dual-branch approach for nuclei localization and extraction of the semantic content. The detection network (first branch) identifies each individual object while the segmentation network performs pixel-wise segmentation (blue: inner area, green: border, red: background).

convolutional network (ResNet50) that proposes regions, and the second module is a detector [38] that analyze and classify the proposed regions. Instead of using pyramids of images (i.e., multiple instances of the image but at different scales), our detection network employs anchor boxes. An anchor box is a reference box with a specific scale and aspect ratio. The network was trained for 150 epochs with an early stopping if validation performance did not improve for 25 epochs. The learning rate is set to $10^{-3}$ with a batch size equal to 8. Performance was calculated as a weighted average of precision, recall, mAP@0.5 and mAP@0.5:0.95, focusing on mAP@0.5 to roughly localize nuclei. During the training process, data augmentation is applied to contain overfitting problems. The applied transformations are summarized with their respective probabilities (p): flipping (upside-down, left-right, $p = 0.30$), random rotation (range 0-180°, $p = 0.5$) and scale change (range: 0-0.2, $p = 0.25$). Prediction on test tiles were made using a sliding windows approach with merged predictions and the non-maximum suppression (NMS) algorithm [39] to eliminate overlapping bounding boxes (Fig. 3).

The second branch of our pipeline is composed of the segmentation network. Specifically, we employed the ResUNet [40] to extract the semantic content of the image. The ResUNet combines the UNet architecture with residual connections from the ResNet backbone. This enables the network to encode multi-scale contextual information and gradually recover the spatial information by capturing object boundaries. We designed a contour-aware model to simultaneously predict the inner areas and boundaries of histological structures using the ResUNet. The entire network is trained on a three-class problem: the contour of each object is extracted from the manual mask and a dilation with a 2-pixel radius is performed to obtain a binary mask of the contours. The pixels are then labeled into three classes: (i) background, (ii) boundary and (iii) inner area. This joint approach of object and edge detection helps define the spatial limits of each histological structure based on location and contour information, which has been shown to outperform single approaches [27], [28]. The deep network is trained with a mini-batch size of 8 and an initial learning rate of $10^{-3}$. Categorical cross-entropy and the Adam optimizer are employed as a loss and optimization function, respectively. The training process is stopped after a maximum of 50 epochs, with a validation patience of 10 epochs for early stopping.

During inference, we adopted the same patch aggregation strategy as in our previous work [41]. Briefly, we synthesize a mirror border in each direction and use a sliding window approach to construct the probability map. The output of our segmentation network is encoded as an RGB image: the red channel represents the softmax probabilities for the ''background'' class; the green channel denotes the ''boundary'' class; and the blue channel indicates the ''inner area'' class. Being a graphical representation of probabilities, the sum of the three channels for any given pixel is always equal to one. As can be seen from Fig. 3, the semantic segmentation alone may not guarantee satisfactory detection of all cell nuclei. The network output could contain incomplete cell contours, especially when two or more nuclei are fused together. To address this issue, we propose a novel deformable model in the next section, which can properly handle the segmentation of fused nuclei that would otherwise be merged by the segmentation network.

### C. SOFTMAX-DRIVEN ACTIVE SHAPE MODELS

Active shape models (ASMs) [42], also known as dynamic deformable models, are a set of computer vision techniques used for object segmentation. These models utilize one or more profiles to define the borders of the objects of interest,

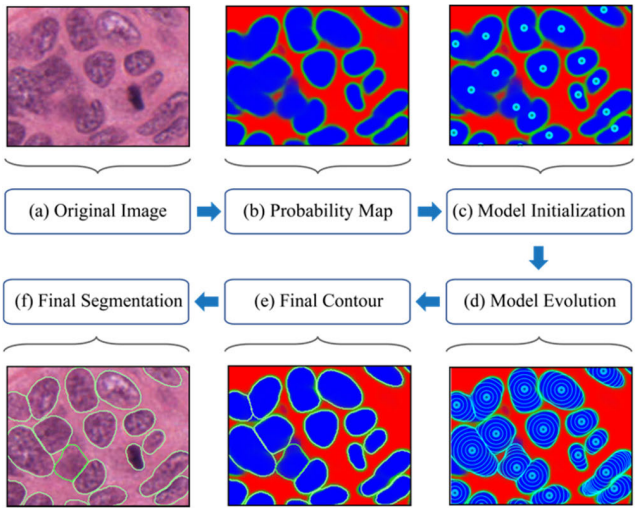



**FIGURE 4. Steps followed by our softmax-driven active shape model (SD-ASM) to segment individual instances of histological structures. The deformable models are initialized at the center of the bounding boxes generated by the detection network, and their evolution occurs directly on the probability map created by the segmentation network. This allows for precise segmentation of individual objects.**

generating a parametric curve or contour. The curvature of the contour is determined by the balance of external and internal forces specified within the model. ASMs offer advantages such as computational simplicity, ease of control, and the ability to consider the simultaneous evolution of multiple instances and their interaction. However, they are sensitive to noise, initialization, and the complexity of the object being segmented. Consequently, they can get stuck in local minima and achieve suboptimal convergence.

In this work, we propose a hybrid model (deep learning + ASMs) to overcome the limitations of the individual models. By leveraging the advantageous features of each model, we aim to achieve robust segmentation. Specifically, we introduce a softmax-driven active shape model (SD-ASM) capable of performing reliable segmentation even in noisy images with overlapping nuclei (Fig. 4). Our models evolve directly on the probability map generated by the segmentation network, as this map provides stronger contrast between the background and foreground compared to the original image (Fig. 4a-b). The SD-ASM is parametrically defined in the (x,y) plane of an image as:

$$\mathbf{v}(\mathbf{s}) = [\mathbf{x}(\mathbf{s}), \mathbf{y}(\mathbf{s})] \tag{3}$$

where x(s), y(s) represents the coordinates along the active shape curve, and s is the normalised arc length. The curve v(s) is represented by a set of control points $v_0, v_1, \ldots, v_{N-1}$ and is linearly obtained by connecting these control points. Each model is initialized at the center of each bounding box generated by the detection network (Fig. 4c) using a circumference of radius 5 pixels with a number of points (N) equal to 40. Next, any ASM that lies in regions with a predominant background probability (red channel of the segmentation map) is removed, as they represent false positives of the detection network.

After initialization, the SD-ASMs evolve on the probability map generated by the segmentation network (Fig. 4d). The evolution of all contours is processed simultaneously to take advantage of code vectorization and increase overall computational efficiency. The evolution of the contour v(s) is regulated by four energy terms, two of which are modulated by the local values of the segmentation map:

A. *Smoothness contribution*: this term seeks to maintain a regular contour shape during evolution:

$$\mathbf{E_{INT}}(\mathbf{s}) = \frac{\mathbf{1}}{\mathbf{2}}(\alpha \cdot \left|\frac{\mathbf{dv}}{\mathbf{ds}}\right|^{\mathbf{2}}) = \frac{\mathbf{1}}{\mathbf{2}}\boldsymbol{\alpha}[(\mathbf{x_{left}} + \mathbf{x_{right}}) - \mathbf{x}] \tag{4}$$

This first-order term is designed to hold the curve together, while the weight parameter $\alpha$ controls the overall elasticity of the model. $x_{left}$ and $x_{right}$ represent the coordinates of the previous and next contour points from the currently considered one.

B. *Expansion contribution*: this term controls the radial inflation of the contour (outward force in radial direction). It is modulated by the local nuclear probability (blue channel), so that the expansion can be reduced and stopped as the contour approaches the edge of the cell (green channel) or the background (red channel). It is formulated as follows:

$$\mathbf{E_{EXT1}}(\mathbf{s}) = \beta \cdot \vec{n}(\mathbf{s}) = \beta \cdot \mathbf{p_{inner}}(\mathbf{x}) \cdot \frac{\mathbf{x} - \mathbf{x_b}}{\mathbf{max}(\mathbf{x} - \mathbf{x_b})} \tag{5}$$

where $\beta$ is the magnitude of the force and $\vec{n}(s)$ stands for the normal unitary vector of the curve at x(s). In terms of coordinates in the plane (x,y), the new coordinate of the point depends on the local inner cell probability ($p_{inner}$) of the pixel at the actual coordinate x and the barycenter of the entire SD-ASM ($x_b$).

C. *Border contribution*: this term is a gradient-based external energy. It is used to stop the evolution of the model near the actual border of the object by taking advantage of the probability map of the boundary class (green channel). It is described as:

$$\begin{aligned}\mathbf{E_{EXT2}}(\mathbf{s}) &= -\boldsymbol{\gamma_1} \cdot \nabla \boldsymbol{P}(\mathbf{v}(\mathbf{s})) \\ &= -\boldsymbol{\gamma_1}[\mathbf{p_{border}}(\mathbf{x}+\mathbf{1}, \mathbf{y}) - \mathbf{p_{border}}(\mathbf{x}-\mathbf{1}, \mathbf{y})]\end{aligned} \tag{6}$$

where P is the potential associated to the external forces, $\gamma_1$ is the weight parameter associated to this term and $p_{border}$ represents the local object boundary probability. In this way the curve is attracted by the local minima of the potential (i.e., the local maxima of the gradient). This term acts as a typical "line" term, attracting contour points to high values of green intensity (nuclear boundary probability).

D. *Mutual interaction contribution*: this last term regulates the interaction between two or more contours in the case of any overlap during their evolution. Two given contours are considered interacting when at least one of

their points intersects the other. Each point that interacts with a second contour is pushed back toward the center of its own contour, following the relationship:

$$\mathbf{E_{EXT3}}(\mathbf{s}) = \frac{\gamma_2}{\mathbf{d}(\mathbf{s})} = \gamma_2 \cdot \frac{\mathbf{1}}{\|\mathbf{x} - \mathbf{x}''\|} \tag{7}$$

where $d(s) = \|x - x''\|$ represents the Euclidean distance between the considered point and the centroid of the interacting contour.

Each of the previously described energy terms has its corresponding weighting parameters ($\alpha$, $\beta$, $\gamma_1, \gamma_2$). These parameters establish the impact of each energy term during the evolution of the contours. The final values of these parameters were derived following a tuning procedure by monitoring performance on the training and validation sets (Section IV-B).

The evolution of each model is controlled by these stopping criteria:

1. Convergence of the SD-ASM: if after a minimum number of iterations (set to 50), the average motion of all points within a single iteration falls below 1 pixel, the contour is considered converged, and its evolution is stopped.
2. Maximum iterations: evolution of each SD-ASM terminates if it reaches the maximum number of iterations (set to 1000).

Once all models have evolved, any segmented objects from the segmentation network that are not contained within the final mask (i.e., nuclei without bounding boxes of the detection network) are added to the final mask. Each contour defines a single segmentation instance, resulting in the output shown in Fig. 4e-f.

### D. PERFORMANCE METRICS

The automatic segmentation obtained from the proposed SD-ASM model are compared with the corresponding manual segmentation to evaluate the performance of our instance segmentation approach on histological structures. We quantify segmentation accuracy using several pixel-based and object-based metrics. At the pixel-level, we compute the binary Dice score, which measures the spatial overlap between predictions and ground truth. At the instance-level, we use the binary Panoptic Quality (PQ) metric [43], which jointly evaluates segmentation quality and detection correctness. We also calculate the Hausdorff Distance (HD) to measure contour adherence. The Aggregated Jaccard Index (AJI) [44] is used as well, as it specifically evaluates nuclear instance segmentation methods by penalizing both object-level and pixel-level errors.

## IV. EXPERIMENTAL RESULTS

### A. DATASETS

In this section, we used the Multi-Organ Nuclei Segmentation (MoNuSeg) dataset [45] and the Colon Nuclei Identification and Counting (CoNIC) dataset [46] to evaluate our method on multiple tissue types and staining conditions. The MoNuSeg dataset consists of 30 H&E-stained images of size 1000 × 1000 pixels acquired at 40x magnification from 18 hospitals. It contains images from seven organs (liver, kidney, prostate, breast, colon, stomach, bladder) with 21,623 delineated nuclei in total. For our study, we used the training, validation and test image splits provided for four organ types: liver, kidney, prostate, and breast. The CoNiC dataset contains 4981 images of colon histology sections stained with H&E. The images, acquired at 20x magnification, were manually annotated by expert pathologists, containing around half a million labelled nuclei. We downloaded the pre-extracted $256 \times 256$ patches and corresponding pixel-level nuclei masks for training and evaluation. The 4,981 images were randomly divided into three subsets for training (4000 images), validation (500 images), and testing (481 images).

By leveraging both datasets, which vary in tissue type, staining protocol and data collection processes, we aimed to develop a segmentation approach capable of handling diversity present in real histopathology datasets. This allows for more thorough validation of the model's generalization abilities.

**TABLE 1. Tuned hyperparameters and search ranges for the faster R-CNN, and ResUnet evaluated on the MoNuSeg and conic datasets.**

| Dataset | Detection branch (Faster R-CNN) | | | Segmentation branch (ResUNet) | |
|---|---|---|---|---|---|
| | Input size | Anchor boxes | Aspect ratios | Input size | Backbone |
| MoNuSeg | 224x224 | [5,15,25, 38,60] | [0.5,1,1.6] | 224x224 | ResNet34 |
| *CoNIC* | 256x256 | [2,5,8,11, 19] | [0.5,1,2] | 128x128 | ResNet50 |

### B. HYPERPARAMETERS TUNING

In this study, we implemented three different methods: Faster R-CNN, ResUNet and SD-ASMs. For each of them, the most relevant hyperparameters were optimized:

1. Faster R-CNN: this network has many hyperparameters most of which we set to recommended values [38]. The network was pretrained on the COCO dataset [47] to reduce training time. We tuned three hyperparameters: input size, anchor boxes and aspect ratios. Anchor boxes and aspect ratios are important for object detection as they define the default bounding boxes used for training. Specifically, anchor boxes of different sizes and aspect ratios are matched to ground truth boxes to train the network. Tuning these anchors helps the network learn more robust size features to detect objects of various shapes and scales. We took the set of hyperparameters that minimized the validation error and measured the performance on the independent test

**TABLE 2. Hyperparameters tuning of the SD-ASMs for the MoNuSeg and conic datasets.**

| Param. | Definition | Range | Best values (MoNuSeg) | Best values (CoNIC) |
|---|---|---|---|---|
| $\alpha$ | smoothness parameter | $[i \cdot 10^{-2}\ for\ i\ \in [1:8]]$ | 0.05 | 0.05 |
| $\beta$ | expansion contribution | $[10^{-i}\ for\ i\ \in [1:4]]$ | 0.1 | 0.1 |
| $\gamma_1$ | cell border contribution | $[i \cdot 10^{-1}\ for\ i \in [1:10]]$ | 0.2 | 0.2 |
| $\gamma_2$ | mutual interaction contribution | $[0.5, 1.0, 1.5, 2.0]$ | 1.5 | 1.5 |

set. Table 1 shows the hyperparameters used in this study.

2. ResUNet: this network was pretrained on ImageNet [48] to take advantage of transfer learning and reduce the training time. For the segmentation network, we tuned two hyperparameters: input size, and backbone. The final architecture was chosen as the one that maximized the performance on the validation set in terms of Dice score (Table 1).
3. SD-ASMs: during the mathematical formulation of the softmax-driven active shape models, we tuned the four weight parameters that govern the balance of internal and external forces. In particular, the following were optimized: $\alpha$ (smoothness contribution), $\beta$ (expansion contribution), $\gamma_1$ (cell border contribution), and $\gamma_2$ (mutual interaction contribution). The final values were chosen as those that maximized performance (i.e., AJI) on the validation set. As shown in Table 2, the optimal parameters for the deformable models are identical across both nuclei segmentation datasets. This lack of variation is expected since the SD-ASMs evolve based on the probability maps, which in turn depend solely on the shape of the segmented objects. Given that both datasets involve nuclear instance segmentation, the morphological properties of the nuclei remain consistent between the two sets.

### C. ABLATION STUDY AND EXECUTION TIME

We conducted an ablation study to understand the effects of different module components of our pipeline. Specifically, we evaluated:

- The proposed full model (ResUNet + SD-ASMs);
- ResUNet without stain normalization (ResUNet w/o SN);
- ResUNet alone without deformable models;
- ResUNet with marker-controlled watershed [49].

Results on the MoNuSeg and CoNIC test sets are reported in Table 3. The SD-ASM module consistently improved performance, achieving the best AJI scores. This indicates deformable models help overcome limitations from missing contour in histopathology images. Compared to ResUNet alone, the full model attained up to 8.93% higher AJI, demonstrating improved boundary adherence. Hausdorff distances were also lowest for our approach, confirming more accurate segmentation. The integration of SD-ASMs with deep networks outputs enables additional gains in PQ score above the original baseline, increasing performance by up to 9.84%. Notably, SD-ASMs consistently provide more accurate segmentations than commonly used methods like watershed transforms (Table 3).

Runtime was measured on a 10-core 3.8GHz CPU with 128G RAM. Deformable model evolution required average times of 11.36 seconds/image for MoNuSeg and 5.72 seconds/image for CoNIC. Images with many touching cells take longer due to the large number of interactions that must be handled by the mutual interaction term. Visual performance of the proposed method is shown in Fig. 5.

### D. INTEGRATION AND COMPARISON WITH OTHER SEGMENTATION METHODS

Three representative segmentation results of our method from different organs are shown in Fig. 6. From the same figure, we can see the ability of our method to correctly separate individual cell instances (thanks to the detection module) and perform accurate contour segmentation (thanks to the semantic map and SD-ASM controllability). In simple cases where nuclei are well separated (Fig. 6a), there is no difference between network segmentation and SD-ASM. This is not surprising: there are, for example, images with relatively few nuclei and well-separated cells, where separation of touching nuclei brings no benefit. In this case, the use of SD-ASM processing is not advantageous. On the other hand, in the case of fused and partially overlapping nuclei (Fig. 6b-c), the combined use of the segmentation network and ASMs ensures contours with high agreement with respect to ground truth. The above experimental data can demonstrate the accuracy and reliability of our approach in the segmentation of overlapping cells. The proposed model can generate segmentations closer to ground truth and successfully separate touching nuclei into individual instances.

To further evaluate the effectiveness of SD-ASMs, we conducted experiments using various state-of-the-art segmentation networks as baseline approaches. Specifically, we examined the impact of SD-ASM as a post-processing strategy starting from the probability maps generated by the following networks: DeepLabv3+ [50], K-Net [51], Twins [52], large ConvNext [53], small ConvNext, SegNeXt [54], and Mask2Former [55]. We then performed a paired t-test ($p < 0.05$) to assess whether the performance differences with and without SD-ASMs were significant. The results for the MoNuSeg test set are summarized in Tables 4 and 5, while the CoNIC results are presented in Table 6.

In the MoNuSeg challenge paper [45], the authors re-annotated all 14 test images and calculated the average AJI between the new and previous annotations. The resulting AJI

**TABLE 3.** Average computational time and performance of the different modules that compose our segmentation framework. SN: stain normalization, SD-ASMs: softmax-driven active shape models. Best values are highlighted in bold.

| Dataset | Method | Comp. Time (s) | Binary Dice | Binary PQ | HD (pixels) | AJI |
|---|---|---|---|---|---|---|
| MoNuSeg - test1 | ResUNet w/o SN | 2.81 | 0.7709 | 0.5523 | 6.80 | 0.5572 |
| | ResUNet | 2.51 | 0.7985 | 0.5867 | 6.64 | 0.5604 |
| | ResUNet + watershed | 4.15 | 0.7930 | 0.5468 | 5.36 | 0.5968 |
| | ResUNet + SD-ASMs | 11.22 | **0.8077** | **0.6300** | **5.07** | **0.6243** |
| MoNuSeg – test2 | ResUNet w/o SN | 2.80 | 0.7836 | 0.5377 | 7.54 | 0.5507 |
| | ResUNet | 3.52 | 0.8066 | 0.5623 | 7.21 | 0.5551 |
| | ResUNet + watershed | 4.87 | 0.8016 | 0.5389 | 5.93 | 0.5984 |
| | ResUNet + SD-ASMs | 11.34 | **0.8221** | **0.6161** | **5.52** | **0.6400** |
| CoNIC – test | ResUNet w/o SN | 1.15 | 0.6481 | 0.3455 | 3.67 | 0.4062 |
| | ResUNet | 1.12 | 0.6659 | 0.3705 | 3.36 | 0.4221 |
| | ResUNet + watershed | 2.68 | 0.6586 | 0.3663 | 2.78 | 0.4452 |
| | ResUNet + SD-ASMs | 5.72 | **0.6920** | **0.4439** | **2.89** | **0.4696** |

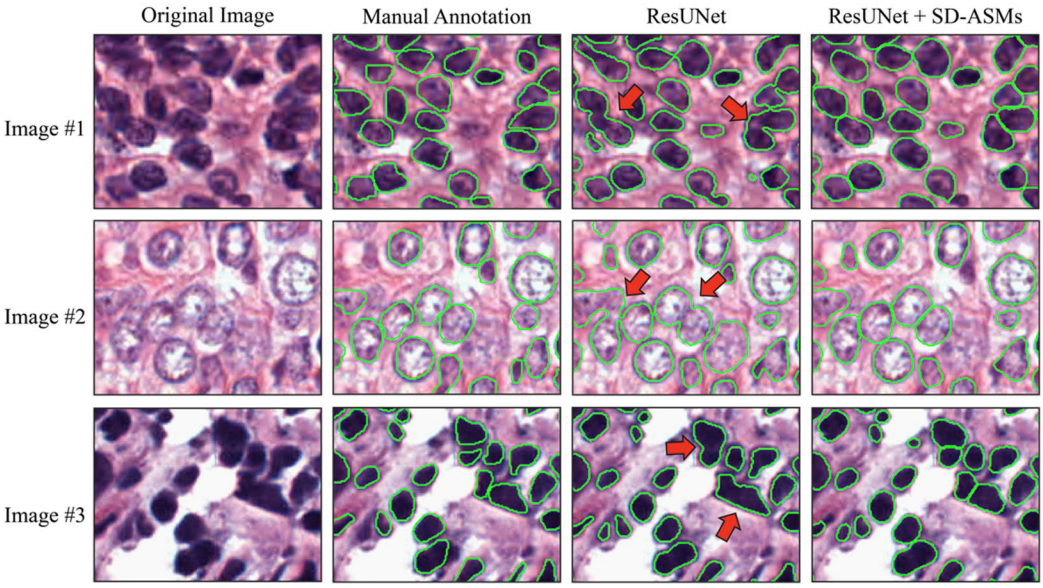



**FIGURE 5.** Examples of tiles with different nuclear densities from the validation and test sets. In the case of well-spaced cells (first row), the segmentation provided by solely the segmentation network is approximately equal to that of our method. If the nuclei start to touch or be partially overlapping (second and third row), the advantage of adopting our segmentation framework over the deep network alone is evident.

for the 14 test images was 0.653. For reference, we included this value in Table 4 as an indication of inter-observer variability. For MoNuSeg dataset, the introduction of SD-ASM systematically improves the performance of the entire deep learning framework. For each of the tested architectures, integrating SD-ASM as a post-processing strategy leads to an increase in both pixel-level and object-level performance. This improvement is particularly notable in the PQ score and AJI, which are fundamental metrics for measuring instance segmentation tasks. In Table 5, we summarized the performance of prior published methods in the MoNuSeg challenge.

The CoNIC dataset also shows similar results, where the integration of SD-ASM statistically enhances performance across all tested architectures. In fact, the AJI metric can increase by up to 5.59% with the integration of SD-ASM.

### *E. GENERALIZATION EVALUATION AND EXTENSION TO OTHER APPLICATIONS*

The proposed modular SD-ASM framework directly segments objects on probability maps, enabling new digital pathology applications requiring single-instance recognition.

Specifically, we tested hepatic steatosis and renal tubule segmentation where individual structure identification is critical. According to Banff consensus guidelines for transplantation, histological assessment now analyzes individual objects rather than occupied areas [64]. Single-instance recognition is therefore essential. The aim of new standards is to improve slide evaluation reproducibility and reduce inter-observer variability.

Deep networks for steatosis/tubule detection and segmentation were trained as described in Section III-B. SD-ASMs

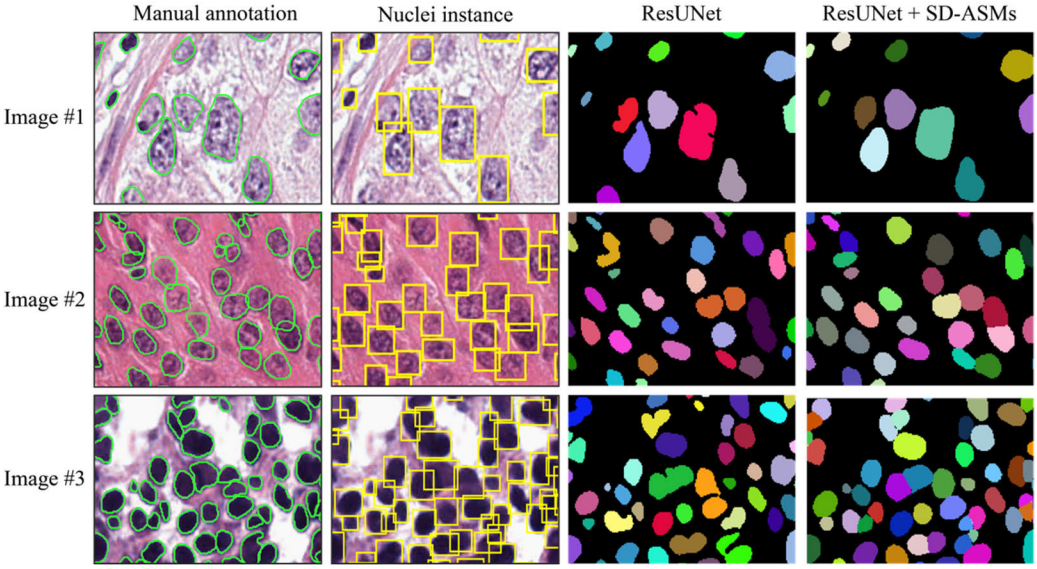



**FIGURE 6.** Performance of the segmentation network with and without the SD-ASMs module. Each row represents sub-images taken from different samples of the test set, while segmentation results are shown in columns. The original image and the corresponding manual annotations is displayed in the first and second columns, respectively. The results provided by the segmentation network alone (ResUNet) are shown in the third column, while the results provided by our proposed segmentation framework (ResUNet + SD-ASMs) are presented in the last column. Red arrows indicate the nuclei fused by the segmentation network but separated correctly by our method.

**TABLE 4.** Performance of different state-of-the-art segmentation networks with and without the SD-ASM module as post-processing in the test set of MoNuSeg. Best values for each network are highlighted in bold. (*) denotes statistical difference respect to baseline.

| Network | Configuration | Binary Dice | Binary PQ | HD (pixels) | AJI |
|---|---|---|---|---|---|
| ResUNet | baseline | 0.8019 | 0.5762 | 6.88 | 0.5581 |
| | with SD-ASMs | **0.8138**$^{(*)}$ | **0.6240**$^{(*)}$ | **5.27**$^{(*)}$ | **0.6310**$^{(*)}$ |
| DeepLabv3 + | baseline | 0.8077 | 0.5723 | 6.49 | 0.5610 |
| | with SD-ASMs | **0.8151**$^{(*)}$ | **0.6018**$^{(*)}$ | **5.23**$^{(*)}$ | **0.6329**$^{(*)}$ |
| K-Net | baseline | 0.7992 | 0.5671 | 6.33 | 0.5695 |
| | with SD-ASMs | **0.8142**$^{(*)}$ | **0.6103**$^{(*)}$ | **5.51**$^{(*)}$ | **0.6312**$^{(*)}$ |
| Twins | baseline | 0.8065 | 0.5801 | 5.93 | 0.5822 |
| | with SD-ASMs | **0.8121**$^{(*)}$ | **0.6196**$^{(*)}$ | **5.04**$^{(*)}$ | **0.6615**$^{(*)}$ |
| ConvNeXt large | baseline | 0.7969 | 0.5721 | 6.01 | 0.5923 |
| | with SD-ASMs | **0.8141**$^{(*)}$ | **0.6169**$^{(*)}$ | **5.13**$^{(*)}$ | **0.6350**$^{(*)}$ |
| ConvNeXt small | baseline | 0.7891 | 0.5713 | 5.84 | 0.5910 |
| | with SD-ASMs | **0.8210**$^{(*)}$ | **0.6326**$^{(*)}$ | **4.99**$^{(*)}$ | **0.6446**$^{(*)}$ |
| SegNeXt | baseline | 0.7924 | 0.5781 | 5.88 | 0.5849 |
| | with SD-ASMs | **0.8230**$^{(*)}$ | **0.6237**$^{(*)}$ | **5.27**$^{(*)}$ | **0.6342**$^{(*)}$ |
| Mask2 Former | baseline | 0.8003 | 0.5730 | 5.94 | 0.5781 |
| | with SD-ASMs | **0.8148**$^{(*)}$ | **0.6320**$^{(*)}$ | **5.33**$^{(*)}$ | **0.6315**$^{(*)}$ |
| Inter-observer variability | | - | - | - | 0.6530 |

**TABLE 5.** Comparison with state of the art on MoNuSeg dataset.

| Method | Dice | AJI |
|---|---|---|
| U-Net [56] | 0.8362 | 0.5971 |
| Mask R-CNN [57] | 0.7600 | 0.5460 |
| DCAN [58] | 0.7920 | 0.5250 |
| Micro-Net [59] | 0.7970 | 0.5600 |
| DIST [44] | 0.7890 | 0.5590 |
| CIA-Net [60] | 0.8180 | 0.6200 |
| FullNet [27] | 0.8027 | 0.6039 |
| Hover-Net [22] | 0.8260 | 0.6180 |
| BRP-Net [61] | - | 0.6422 |
| PFF-Net [62] | 0.8091 | 0.6107 |
| CD-Net [63] | **0.8316** | 0.6331 |
| ConvNeXt small + SD-ASM | 0.8210 | 0.6446 |
| Twins + SD-ASM | 0.8121 | **0.6615** |

(Section III-C) were then applied to probability maps to refine instance-level results. Quantitative metrics on these tasks (Table 7) show SD-ASMs exceeding reference performances, with significant gains in binary PQ score (steatosis: +6.6%, tubules: +5.3%) and AJI (steatosis: +6.2%, tubules: +4.2%). Furthermore, the Hausdorff Distance decreased by up to 38% for both applications.

Fig. 7 demonstrates excellence across other applications in digital pathology, underlining the framework's generalization. Segmenting individual steatosis lesions and tubules is critical given new guidelines; our method addresses this need through robust single-instance recognition. This modularity and flexibility position SD-ASMs for widespread use in digital pathology and beyond.

## V. DISCUSSION

In the field of digital pathology, accurate segmentation of histological structures is crucial for developing interpretable models that can predict clinical variables, such as treatment efficacy or the risk of relapse. Nuclei play a fundamental role in tissue analysis, making their accurate segmentation a cornerstone in building interpretable models. Once nuclei are correctly segmented, it becomes possible to analyze spatial

**TABLE 6.** Performance of different state-of-the-art segmentation networks with and without the SD-ASM module as post-processing in the test set of conic. Best values for each network are highlighted in bold. (*) denotes statistical difference respect to baseline.

| Network | Configuration | Binary Dice | Binary PQ | HD (pixels) | AJI |
|---|---|---|---|---|---|
| ResUNet | baseline | 0.6659 | 0.3705 | 3.36 | 0.4221 |
| | with SD-ASMs | **0.6920**$^{(*)}$ | **0.4439**$^{(*)}$ | **2.89**$^{(*)}$ | **0.4696**$^{(*)}$ |
| DeepLabv3 + | baseline | 0.6722 | 0.3813 | 3.18 | 0.4245 |
| | with SD-ASMs | **0.7035**$^{(*)}$ | **0.4561**$^{(*)}$ | **2.75**$^{(*)}$ | **0.4739**$^{(*)}$ |
| K-Net | baseline | 0.6818 | 0.3864 | 3.22 | 0.4304 |
| | with SD-ASMs | **0.7067**$^{(*)}$ | **0.4594**$^{(*)}$ | **2.65**$^{(*)}$ | **0.4811**$^{(*)}$ |
| Twins | baseline | 0.6883 | 0.3918 | 2.87 | 0.4362 |
| | with SD-ASMs | **0.7102**$^{(*)}$ | **0.4772**$^{(*)}$ | **2.31**$^{(*)}$ | **0.4921**$^{(*)}$ |
| ConvNeXt large | baseline | 0.7037 | 0.3974 | 2.74 | 0.4531 |
| | with SD-ASMs | **0.7345**$^{(*)}$ | **0.4831**$^{(*)}$ | **2.12**$^{(*)}$ | **0.4977**$^{(*)}$ |
| ConvNeXt small | baseline | 0.6951 | 0.3877 | 2.86 | 0.4432 |
| | with SD-ASMs | **0.7191**$^{(*)}$ | **0.4789**$^{(*)}$ | **2.45**$^{(*)}$ | **0.4862**$^{(*)}$ |
| SegNeXt | baseline | 0.6810 | 0.3789 | 3.11 | 0.4362 |
| | with SD-ASMs | **0.7198**$^{(*)}$ | **0.4601**$^{(*)}$ | **2.52**$^{(*)}$ | **0.4793**$^{(*)}$ |
| Mask2 Former | baseline | 0.6778 | 0.3846 | 2.89 | 0.4511 |
| | with SD-ASMs | **0.7421**$^{(*)}$ | **0.4802**$^{(*)}$ | **2.32**$^{(*)}$ | **0.4808**$^{(*)}$ |

**TABLE 7.** Pixel-based and object-based performance of our method for liver steatosis segmentation and kidney tubules segmentation. Comparison with the baseline method (reference paper) in the test set is also provided. Best values are highlighted in bold.

| Application | Method | Binary Dice | Binary PQ | HD (pixels) | AJI |
|---|---|---|---|---|---|
| Liver steatosis | Ref. paper [64] | 0.7271 | 0.5468 | 5.12 | 0.5375 |
| | SD-ASMs | **0.7422** | **0.6132** | **3.16** | **0.5993** |
| Kidney tubules | Ref. paper [65] | 0.8920 | 0.6431 | 3.82 | 0.7229 |
| | SD-ASMs | **0.9246** | **0.6964** | **2.55** | **0.7652** |

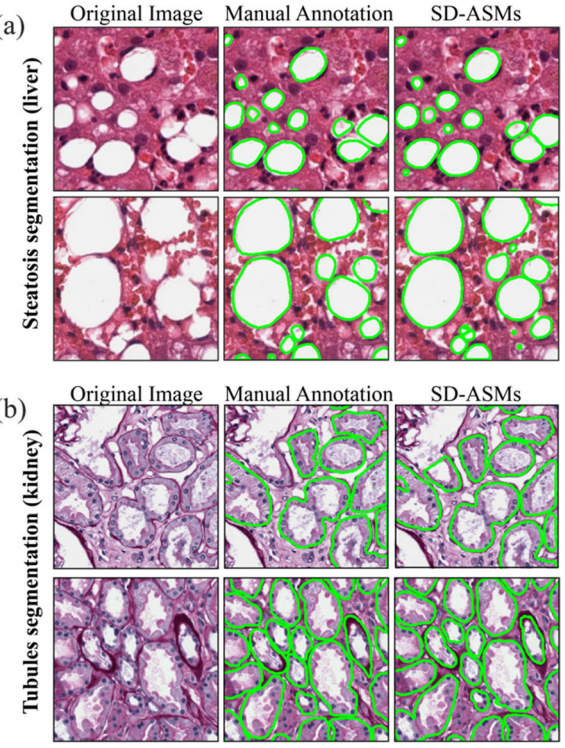



**FIGURE 7.** Segmentation performance of our method in two other digital pathology applications in which there are touching and overlapping objects. (a) liver steatosis segmentation, (b) renal tubules segmentation.

and morphological features of the tissue in-depth, bridging the gap between low-level molecular information and the macroscopic characteristics of diseased tissue.

Recent studies have demonstrated the significance of nucleus segmentation in various cancer types, such as lung, bladder, and colorectal cancer, for grading and assisting in diagnosis and treatment [66], [67], [68]. Accurate nucleus detection and boundary segmentation are vital for these applications. Precise nucleus segmentation enables clinicians to obtain more pathological information, analyze spatial characteristics of tissues and targets, and perform morphometric analysis. However, existing state-of-the-art deep learning models still face challenges in handling the complex structure of histology images, especially when nuclei overlap or touch each other [3], [20], [44].

This study aims to propose an end-to-end pipeline for instance segmentation, with a specific focus on the post-processing stage. Our framework is designed to accurately recognize and segment individual objects within crowded images, effectively addressing the challenge of overlapping objects while maintaining precise boundaries for subsequent morphometry analysis. The proposed pipeline is applicable not only to nucleus segmentation but also to the segmentation of steatosis and tubules in digital pathology images.

The pipeline consists of a dual-branch deep learning approach for object localization and coarse segmentation, followed by a novel formulation of softmax-driven active shape models (SD-ASMs). These dynamic deformable models effectively separate individual object instances and achieve accurate contour segmentation. To address the issue of touching and partially overlapping structures, a mutual interaction term is incorporated into the SD-ASM, effectively alleviating the problem and improving segmentation accuracy.

Experimental results demonstrate that the proposed segmentation framework is well-suited for histological structure segmentation, particularly in handling touching objects. The method is benchmarked on a standard dataset specifically designed for evaluating nuclear segmentation tasks, using various pixel-level and object-level metrics. The evaluation focuses on the Panoptic Quality (PQ) and Aggregated Jaccard Index (AJI), two of the most comprehensive indicators of segmentation quality. Our statistical analysis provides strong evidence that SD-ASM post-processing significantly enhances the performance of segmentation networks. As detailed in Tables 4 and 6, the integration of SD-ASMs downstream consistently delivered higher accuracy scores across multiple state-of-the-art models and benchmark datasets, as validated through paired t-testing. Additionally, experiments on steatosis and tubules datasets show that the proposed method outperforms reference papers, showcasing its effectiveness.

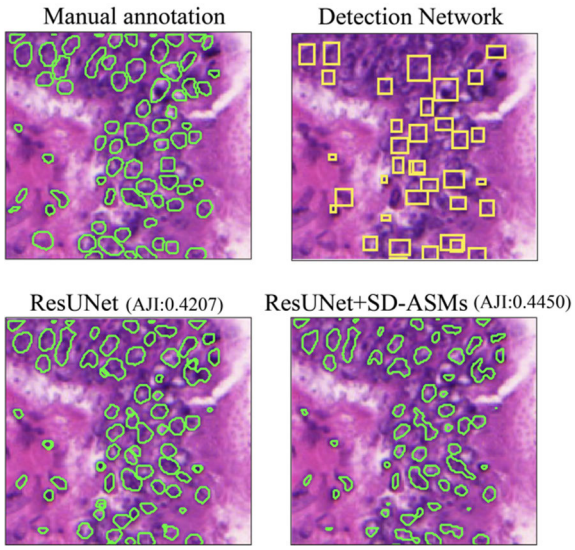



**FIGURE 8.** Examples of tiles where the performance of the detection network is unsatisfactory. This is reflected in the performance of SD-ASM which achieves AJI values comparable to the segmentation network alone.

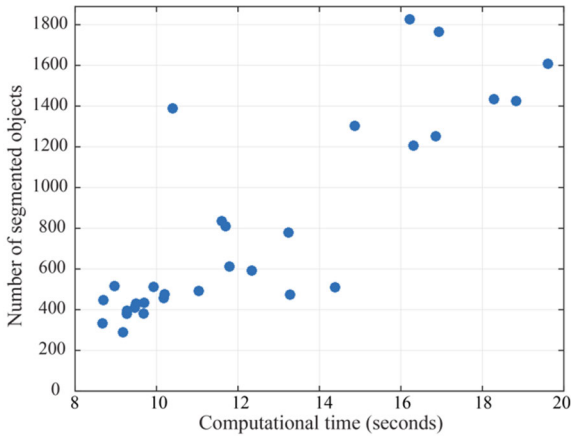



**FIGURE 9.** Joint plot of number of segmented objects and running time for MoNuSeg dataset.

The high performance of the proposed approach can be attributed to the synergistic combination of two deep neural networks with the SD-ASM model. Stain normalization is also essential for standardizing the intensity of histological structures and enhancing the stability and performance of the SD-ASM model. The modular design combining deep networks and SD-ASMs endows flexibility to generalize across diverse histology image domains. Despite datasets exhibiting variability in staining type/intensity and structural appearance, the SD-ASMs could still evolve reliably on the segmentation network's probability maps. This is because regardless of the image source or target structures, the probability maps represent a more stable semantic representation for the deformable models compared to raw pixels. The proposed approach offers several advantages, including automatic initialization, overcoming a major challenge of active shape models.

While larger datasets may refine network performance, the flexible nature of SD-ASMs allows our method to achieve robust and accurate instance segmentation even when training data is limited. This resilience was evident in our experiments utilizing datasets of varying sizes. For example, the MoNuSeg benchmark contains a relatively small training dataset consisting of only 16 labeled images from four organ types. Yet integrating SD-ASMs consistently improved segmentation quality beyond using the deep network alone, as validated through metrics like AJI that systematically increased across all tested architectures (Tables 4-6). This finding underscores the SD-ASM approach's effectiveness in segmenting histological structures even with restricted labeled data availability, thanks to its formulation operating on semantics-rich probability maps rather than raw pixels. Our method also shows competitive results respect to prior works (Table 5)

While our experiments demonstrated the ability of SD-ASM to effectively handle touching and partially overlapping objects, there may be some limitations. Images with highly irregular, elongated or multi-lobed structures that cross with each other could potentially challenge the separation of contours during evolution. Very dense clusters with significant overlap between 3 or more structures may also be difficult to segment correctly if the detection network does not adequately recognize individual instances. This problem is highlighted in Fig. 8 of the paper. In this case, the AJI between just the segmentation network and the combined detection-segmentation model with SD-ASM is comparable. To address such failure modes, future work should systematically analyze failure cases to better characterize challenging configurations. Potential strategies to improve accuracy could include: 1) Incorporating shape priors derived from training data to regularize irregular contour evolutions. 2) Improving the detection and segmentation networks as their performance directly reflects on the ability of SD-ASMs to correctly segment, since SD-ASMs evolve based on the probability maps generated by these networks. An interesting future development is represented by the HoverNet network [22], able to perform simultaneous segmentation and recognition of nuclei centroids.

Moreover, the computational cost is higher compared to typical inference time of a CNN and depends on the number of objects within the image to be segmented, as shown in our runtime analysis (Fig. 9). For example, when segmenting clustered cell nuclei as in the MoNuSeg dataset, SD-ASM processing required an average of 11.36 seconds per image versus 2.51 seconds for the CNN alone. In contrast, for tasks involving fewer structures like tubule segmentation, runtime increases were negligible. The computation of the mutual interaction term, which handles multiple interacting contours simultaneously, currently contributes nearly half of the overall runtime. This highlights the need for optimization to efficiently scale the approach to more densely packed objects. Future work will focus on improving computational

efficiency through parallel computing methods. Additionally, efforts will be made to optimize the combined performance of the two deep learning models and address scalability issues, such as applying the proposed approach to large-scale data, including whole-slide images (WSI). In future works, we also aim to analyze the performance of our method on datasets containing multiple histological structure classes where class imbalance is systematically varied. Extending the deep network training to include classification tasks will allow a more direct assessment of how SD-ASM instance segmentation is impacted by rare class representation.

From a broader perspective, the problem of segmenting touching entities in biological data extends beyond nucleus detection in histopathology. While our study applied the SD-ASM method to segment nuclei, steatosis, and tubules, further evaluation on additional tissue morphologies such as glands or tubular structures could provide insights into its versatility across diverse tissue types. Adapting the internal and external energy terms or incorporating shape priors where needed may also expand the method's applicability.

Instance segmentation commonly arise in cellular-level imaging techniques like fluorescence microscopy [69], live-cell imaging, and electron microscopy, as they frequently capture clusters of contiguous cells or organelles. Accurately delineating individual objects that contact or overlap is critical for proper quantification and analysis in these domains. We believe expanding our methods to other imaging modalities, with modality-specific customization of structures as required, could enhance robust segmentation. This is especially important for fields like digital pathology, immunofluorescence, and electron microscopy that demand precise single-instance object recognition in order to strengthen analysis through accurate quantification and characterization of cellular components.

## VI. CONCLUSION

In this work, we presented a novel and efficient formulation of dynamic deformable models, driven by two deep neural networks, to achieve robust and accurate segmentation of crowded objects in histology images. Our framework is specifically designed to properly handle touching and overlapping objects. This ability was demonstrated through superior performance on the PQ and AJI metrics, which independently evaluate both pixel-level segmentation and object-level detection performance.

The proposed dual-branch pipeline is inherently generalizable, as evidenced by successful application to segmenting hepatic steatosis lesions and renal tubules. For the application to other imaging modalities, the internal and external energy terms governing contour evolution may need adaptation depending on object shape, overlap patterns, and imaging characteristics. The synergistic combination of deep learning with interpretable and controllable deformable models shows promise for robust single-instance recognition across diverse biological and biomedical domains.

## ACKNOWLEDGMENT

The authors would moreover like to thank Ing. Luca Costa for the support and useful discussions.

## REFERENCES

[1] A. R. Kherlopian, T. Song, Q. Duan, M. A. Neimark, M. J. Po, J. K. Gohagan, and A. F. Laine, "A review of imaging techniques for systems biology," *BMC Syst. Biol.*, vol. 2, no. 1, pp. 1–18, Dec. 2008.
[2] X. Chen, B. Zheng, and H. Liu, "Optical and digital microscopic imaging techniques and applications in pathology," *Anal. Cellular Pathol.*, vol. 34, nos. 1–2, pp. 5–18, 2011.
[3] L. Xie, J. Qi, L. Pan, and S. Wali, "Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images," *Neurocomputing*, vol. 376, pp. 166–179, Feb. 2020.
[4] W. Yin, Y. Hu, S. Yi, and J. He, "A segmentation method combining probability map and boundary based on multiple fully convolutional networks and repetitive training," *Phys. Med. Biol.*, vol. 64, no. 18, Sep. 2019, Art. no. 185003.
[5] L. Roszkowiak, A. Korzynska, D. Pijanowska, R. Bosch, M. Lejeune, and C. Lopez, "Clustered nuclei splitting based on recurrent distance transform in digital pathology images," *EURASIP J. Image Video Process.*, vol. 2020, no. 1, pp. 1–16, Dec. 2020.
[6] J. R. Gilbertson, J. Ho, L. Anthony, D. M. Jukic, Y. Yagi, and A. V. Parwani, "Primary histologic diagnosis using automated whole slide imaging: A validation study," *BMC Clin. Pathol.*, vol. 6, no. 1, pp. 1–19, Dec. 2006.
[7] A. Vahadane, T. Peng, A. Sethi, S. Albarqouni, L. Wang, M. Baust, K. Steiger, A. M. Schlitter, I. Esposito, and N. Navab, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1962–1971, Aug. 2016.
[8] S. Zhao, J. Wang, X. Wang, Y. Wang, H. Zheng, B. Chen, A. Zeng, F. Wei, S. Al-Kindi, and S. Li, "Attractive deep morphology-aware active contour network for vertebral body contour extraction with extensions to heterogeneous and semi-supervised scenarios," *Med. Image Anal.*, vol. 89, Oct. 2023, Art. no. 102906.
[9] J. Cheng and J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 3, pp. 741–748, Mar. 2009.
[10] H. Irshad, A. Veillard, L. Roux, and D. Racoceanu, "Methods for nuclei detection, segmentation, and classification in digital histopathology: A review—current status and future potential," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97–114, 2014.
[11] M. Veta, J. P. W. Pluim, P. J. van Diest, and M. A. Viergever, "Breast cancer histopathology image analysis: A review," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1400–1411, May 2014.
[12] 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, Mar. 2017.
[13] S. Lal, D. Das, K. Alabhya, A. Kanfade, A. Kumar, and J. Kini, "NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104075.
[14] F. Xing, Y. Xie, and L. Yang, "An automatic learning-based framework for robust nucleus segmentation," *IEEE Trans. Med. Imag.*, vol. 35, no. 2, pp. 550–566, Feb. 2016.
[15] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 448–459, Feb. 2019.
[16] S. Graham and N. M. Rajpoot, "SAMS-NET: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 590–594.
[17] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, "A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images," *Med. Biol. Eng. Comput.*, vol. 57, no. 9, pp. 2027–2043, Sep. 2019.
[18] T. Wan, L. Zhao, H. Feng, D. Li, C. Tong, and Z. Qin, "Robust nuclei segmentation in histopathology using ASPPU-Net and boundary refinement," *Neurocomputing*, vol. 408, pp. 144–156, Sep. 2020.
[19] A. K. Chanchal, A. Kumar, S. Lal, and J. Kini, "Efficient and robust deep learning architecture for segmentation of kidney and breast histopathology images," *Comput. Electr. Eng.*, vol. 92, Jun. 2021, Art. no. 107177.

[20] X. Liu, Z. Guo, J. Cao, and J. Tang, "MDC-Net: A new convolutional neural network for nucleus segmentation in histopathology images with distance maps and contour information," *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104543.

[21] S. Wazir and M. M. Fraz, "HistoSeg: Quick attention with multi-loss function for multi-structure segmentation in digital histology images," in *Proc. 12th Int. Conf. Pattern Recognit. Syst. (ICPRS)*, Jun. 2022, pp. 1–7.

[22] V. Thi-Tuong Vo, S.-H. Kim, and T. Lee, "MF-hovernet: An extension of hovernet for colon nuclei identification and counting (CoNiC) challenge," 2022, *arXiv:2203.02161*.

[23] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101563.

[24] K. Saednia, W. T. Tran, and A. Sadeghi-Naini, "A cascaded deep learning framework for segmentation of nuclei in digital histology images," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 4764–4767.

[25] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: Deep contour-aware networks for accurate gland segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2487–2496.

[26] A. BenTaieb, J. Kawahara, and G. Hamarneh, "Multi-loss convolutional networks for gland analysis in microscopy," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 642–645.

[27] H. Qu, Z. Yan, G. M. Riedlinger, S. De, and D. N. Metaxas, "Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Springer, 2019, pp. 378–386.

[28] Y. Xu, Y. Li, Y. Wang, M. Liu, Y. Fan, M. Lai, and E. I. Chang, "Gland instance segmentation using deep multichannel neural networks," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 12, pp. 2901–2912, Dec. 2017.

[29] H. Zhang, D. Jiang, and R. Li, "SGDANet: Gland instance segmentation based on spatial and geometric dual-path attention modules," in *Proc. 5th Int. Conf. Pattern Recognit. Artif. Intell. (PRAI)*, Aug. 2022, pp. 1033–1039.

[30] C. Chen, Y. Huang, P. Fang, C. Liang, and R. Chang, "A computer-aided diagnosis system for differentiation and delineation of malignant regions on whole-slide prostate histopathology image using spatial statistics and multidimensional DenseNet," *Med. Phys.*, vol. 47, no. 3, pp. 1021–1033, Mar. 2020.

[31] F. Ciompi, O. Geessink, B. E. Bejnordi, G. S. de Souza, A. Baidoshvili, G. Litjens, B. van Ginneken, I. Nagtegaal, and J. van der Laak, "The importance of stain normalization in colorectal tissue classification with convolutional networks," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 160–163.

[32] M. Salvi, M. Bosco, L. Molinaro, A. Gambella, M. Papotti, U. R. Acharya, and F. Molinari, "A hybrid deep learning approach for gland segmentation in prostate histopathological images," *Artif. Intell. Med.*, vol. 115, May 2021, Art. no. 102076.

[33] A. Anghel, M. Stanisavljevic, S. Andani, N. Papandreou, J. H. Rüschoff, P. Wild, M. Gabrani, and H. Pozidis, "A high-performance system for robust stain normalization of whole-slide images in histopathology," *Frontiers Med.*, vol. 6, p. 193, Sep. 2019.

[34] F. G. Zanjani, S. Zinger, B. E. Bejnordi, J. A. W. M. van der Laak, and P. H. N. de With, "Stain normalization of histopathology images using generative adversarial networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 573–577.

[35] M. Salvi, F. Branciforti, F. Molinari, and K. M. Meiburger, "Generative models for color normalization in digital pathology and dermatology: Advancing the learning paradigm," *Expert Syst. Appl.*, vol. 245, Jul. 2024, Art. no. 123105.

[36] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2813–2821.

[37] X. Wu, D. Sahoo, H. Doyen, and H. Steven, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.

[38] H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 650–657.

[39] A. Neubeck and L. Van Gool, "Efficient non-maximum suppression," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, 2006, pp. 850–855.

[40] M. Salvi, A. Mogetta, A. Gambella, L. Molinaro, A. Barreca, M. Papotti, and F. Molinari, "Automated assessment of glomerulosclerosis and tubular atrophy using deep learning," *Computerized Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101930.

[41] M. Salvi, A. Mogetta, K. M. Meiburger, A. Gambella, L. Molinaro, A. Barreca, M. Papotti, and F. Molinari, "Karpinski score under digital investigation: A fully automated segmentation algorithm to identify vascular and stromal injury of Donors' kidneys," *Electronics*, vol. 9, no. 10, p. 1644, Oct. 2020, doi: 10.3390/electronics9101644.

[42] T. F. Cootes and C. J. Taylor, "Active shape models—'Smart snakes,'" in *Proc. Brit. Mach. Vis. Conf., Organised British Mach. Vis. Assoc.*, pp. 266–275, 1992.

[43] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollár, "Panoptic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9396–9405.

[44] N. Kumar, R. Verma, S. Sharma, S. Bhargava, A. Vahadane, and A. Sethi, "A dataset and a technique for generalized nuclear segmentation for computational pathology," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, pp. 1550–1560, Jul. 2017.

[45] N. Kumar et al., "A multi-organ nucleus segmentation challenge," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1380–1391, May 2020.

[46] S. Graham, M. Jahanifar, Q. D. Vu, G. Hadjigeorghiou, T. Leech, D. Snead, S. E. A. Raza, F. Minhas, and N. Rajpoot, "CoNIC: Colon nuclei identification and counting challenge 2022," 2021, *arXiv:2111.14485*.

[47] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. Lawrence Zitnick, and P. Dollár, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, Springer, 2014, pp. 740–755.

[48] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[49] K. Parvati, B. S. P. Rao, and M. M. Das, "Image segmentation using gray-scale morphology and marker-controlled watershed transformation," *Discrete Dyn. Nature Soc.*, vol. 2008, pp. 1–8, 2008.

[50] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[51] W. Zhang, J. Pang, K. Chen, and C. C. Loy, "K-Net: Towards unified image segmentation," in *Proc. NIPS*, 2021, pp. 10326–10338.

[52] X. Chu, Z. Tian, Y. Wang, B. Zhang, H. Ren, X. Wei, H. Xia, and C. Shen, "Twins: Revisiting the design of spatial attention in vision transformers," in *Proc. 35th Conf. Neural Inf. Process. Syst.*, 2021, pp. 9355–9366.

[53] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A ConvNet for the 2020s," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11966–11976.

[54] Z M.-H. Guo, C.-Z. Lu, Q. Hou, Z. Liu, M.-M. Cheng, and S.-M. Hu, "Segnext: Rethinking convolutional attention design for semantic segmentation," in *Proc. Adv. Neural Inf. Process Syst.*, vol. 35, 2022, pp. 1140–1156.

[55] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 1280–1289.

[56] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, 2015, pp. 234–241.

[57] N K. He and G. Gkioxari, "P. Doll ar, and R. Girshick,'Mask R-CNN,"' in *Proc. IEEE Int. Conf. Comput. Vis*, 2017, pp. 2980–2988.

[58] H. Chen, X. Qi, L. Yu, Q. Dou, J. Qin, and P.-A. Heng, "DCAN: Deep contour-aware networks for object instance segmentation from histology images," *Med. Image Anal.*, vol. 36, pp. 135–146, Feb. 2017.

[59] S. E. A. Raza, L. Cheung, M. Shaban, S. Graham, D. Epstein, S. Pelengaris, M. Khan, and N. M. Rajpoot, "Micro-net: A unified model for segmentation of various objects in microscopy images," *Med. Image Anal.*, vol. 52, pp. 160–173, Feb. 2019.

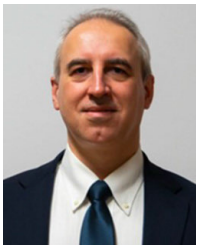
[60] Y. Zhou, O. F. Onder, Q. Dou, E. Tsougenis, H. Chen, and P.-A. Heng, "CIA-Net: Robust nuclei instance segmentation with contour-aware information aggregation," in *Proc. Int. Conf. Inf. Process. Med. Imag.* Switzerland: Springer, 2019, pp. 682–693.

[61] S. Chen, C. Ding, and D. Tao, "Boundary-assisted region proposal networks for nucleus segmentation," in *Proc. Med. Image Comput. Comput. Assist. Intervent.-MICCAI 2020, 23rd Int. Conf., Lima, Peru*. Springer, 2020, pp. 279–288.

[62] D. Liu, D. Zhang, Y. Song, H. Huang, and W. Cai, "Panoptic feature fusion net: A novel instance segmentation paradigm for biomedical and biological images," *IEEE Trans. Image Process.*, vol. 30, pp. 2045–2059, 2021.
[63] H. He, Z. Huang, Y. Ding, G. Song, L. Wang, Q. Ren, P. Wei, Z. Gao, and J. Chen, "CDNet: Centripetal direction network for nuclear instance segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4006–4015.
[64] D. A. H. Neil, M. Minervini, M. L. Smith, S. G. Hubscher, E. M. Brunt, and A. J. Demetris, "Banff consensus recommendations for steatosis assessment in donor livers," *Hepatology*, vol. 75, no. 4, pp. 1014–1025, 2022.
[65] M. Salvi, L. Molinaro, J. Metovic, D. Patrono, R. Romagnoli, M. Papotti, and F. Molinari, "Fully automated quantitative assessment of hepatic steatosis in liver transplants," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103836.
[66] R. Ding, P. Prasanna, G. Corredor, C. Barrera, P. Zens, C. Lu, P. Velu, P. Leo, N. Beig, H. Li, P. Toro, S. Berezowska, V. Baxi, D. Balli, M. Belete, D. L. Rimm, V. Velcheti, K. Schalper, and A. Madabhushi, "Image analysis reveals molecularly distinct patterns of TILs in NSCLC associated with treatment outcome," *npj Precis. Oncol.*, vol. 6, no. 1, pp. 1–15, Jun. 2022.
[67] H. Mi, T. J. Bivalacqua, M. Kates, R. Seiler, P. C. Black, A. S. Popel, and A. S. Baras, "Predictive models of response to neoadjuvant chemotherapy in muscle-invasive bladder cancer using nuclear morphology and tissue architecture," *Cell Rep. Med.*, vol. 2, no. 9, Sep. 2021, Art. no. 100382.
[68] M. Bilal, S. E. A. Raza, A. Azam, S. Graham, M. Ilyas, I. A. Cree, D. Snead, F. Minhas, and N. M. Rajpoot, "Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: A retrospective study," *Lancet Digit. Health*, vol. 3, no. 12, pp. e763–e772, Dec. 2021.
[69] G. Garoffolo et al., "Reduction of cardiac fibrosis by interference with YAP-dependent transactivation," *Circulat. Res.*, vol. 131, no. 3, pp. 239–257, Jul. 2022.

**MASSIMO SALVI** is currently an Assistant Professor with Politecnico di Torino. In addition to his work in biomedical signal and image processing. He is also actively engaged in the field of artificial intelligence and deep learning. He has experience in the development of hybrid techniques that combine AI-based methods and mathematical-statistical techniques to address complex problems in biomedical research. His primary research interests include the development of automated solutions for fluorescence and optical microscopy, with a particular emphasis on the analysis and interpretation of cellular and tissue images.

**KRISTEN M. MEIBURGER** (Member, IEEE) is currently active in the field of biomedical image processing, with a focus on vascular network analysis using biomedical optical imaging methods (i.e., optical coherence tomography angiography, photoacoustics) and radio mics and deep learning applications. She is also active in the field of biomedical signal processing and, particularly, in ultrasound image beamforming methods.



**FILIPPO MOLINARI** (Senior Member, IEEE) is currently a Full Professor with Politecnico di Torino, where he leads research in several areas of biomedical engineering. His primary research interests include biomedical signal processing, image processing, and ultrasound technology, with a focus on developing advanced diagnostic tools for a range of medical applications. In addition to his work in these areas, his research activities also include the non-invasive characterization of tumor vascularization, neuroimaging for the advanced assessment of neurodegenerative disorders, and neurovascular and metabolic assessment of cerebral autoregulation.