An Ensemble Method with Edge Awareness for Abnormally Shaped Nuclei Segmentation

Yue Han* Yang Lei† Viktor Shkolnikov† Daisy Xin†
Alicia Auduong† Steven Barcelo† Jan Allebach* Edward J. Delp*
* Video and Image Processing Lab (VIPER), Purdue University; West Lafayette, Indiana, USA
† HP Inc; Palo Alto, California, USA

Abstract

Abnormalities in biological cell nuclei shapes are correlated with cell cycle stages, disease states, and various external stimuli. There have been many deep learning approaches that are being used for nuclei segmentation and analysis. In recent years, transformers have performed better than CNN methods on many computer vision tasks. One problem with many deep learning nuclei segmentation methods is acquiring large amounts of annotated nuclei data, which is generally expensive to obtain. In this paper, we propose a Transformer and CNN hybrid ensemble processing method with edge awareness for accurately segmenting abnormally shaped nuclei. We call this method Hybrid Edge Mask R-CNN (HER-CNN), which uses Mask R-CNNs with the ResNet and the Swin-Transformer to segment abnormally shaped nuclei. We add an edge awareness loss to the mask prediction step of the Mask R-CNN to better distinguish the edge difference between the abnormally shaped nuclei and typical oval nuclei. We describe an ensemble processing strategy to combine or fuse individual segmentations from the CNN and the Transformer. We introduce the use of synthetic ground truth image generation to supplement the annotated training images due to the limited amount of data. Our proposed method is compared with other segmentation methods for segmenting abnormally shaped nuclei. We also include ablation studies to show the effectiveness of the edge awareness loss and the use of synthetic ground truth images.

1. Introduction

Quantitative analysis of nuclei morphology is important for the understanding of cell architecture. While most nuclei typically have an elliptical shape, deviations from this shape can arise in certain stages of the cell cycle, due to external stress, or in certain disease states. Some cell types also normally have non-elliptical nuclei (e.g. multi-lobed nuclei in neutrophils). Characterization of nuclei shapes, therefore, yields important information for many applications such as determining cell cycle stage, measuring cellular response to environmental stimuli, indicating genetic instability, and cancer diagnostics [13]. Traditional analysis of nuclei shapes requires manual assessment of a large number of microscopy images, which is laborious and time-consuming. Hence, image-based automated nuclei segmentation has been widely used to assist the researcher in the analysis of nuclei morphology.

Both traditional computer vision and Convolutional Neural Networks (CNNs) have been used for automated nuclei segmentation. One of the major challenges in traditional nuclei segmentation methods is that they usually require manual parameter tuning and re-parameterization for new cell types and datasets to achieve adequate performance [23]. CNN-based instance segmentation methods have provided adequate results for general object segmentation and are able to segment nuclei after training on annotated nuclei images [24]. However, they fail to provide precise boundaries for the nuclei when the contrast between foreground and background is low. To improve this, other CNN approaches such as StarDist [39, 44], and Cellpose [41] are specially designed for segmenting the nuclei in microscopy images. These methods have been developed for elliptical nuclei segmentation while approaches for segmenting non-elliptical nuclei are lacking. In recent years, Transformer-based methods have shown performance, which exceeds that of competing methods, across a wide spectrum in the area of computer vision [10, 29]. Transformer architectures are based on a self-attention mechanism that learns the relationships between elements of a sequence, which shows advantages in modeling the long-range relation [26]. However, compared to the CNN-based methods, transformer-based methods sometimes exhibit limitations in extracting detailed localization information. Ensemble processing, where outputs of several segmentation methods are combined or fused, has proven to be useful in improving segmentation performance as well as robustness [16]. Another challenge in nuclei segmentation is the difficulty in obtaining a large number of ground truth annotations. One way to address this problem is by using data augmentation methods to create more training samples [11, 36, 38, 40].

In this paper, we describe a Transformer and CNN hybrid ensemble processing with edge awareness for abnormally shaped nuclei segmentation, known as Hybrid Edge Mask R-CNN (HER-CNN). This method is based on Mask R-CNN models with CNN and Transformers architectures for segmenting abnormally shaped nuclei. We propose to add an edge awareness loss to the mask prediction step of the Mask R-CNN to provide additional awareness of the abnormally shaped nuclei boundary. We describe a modified Non-maximum-Suppression (NMS) to combine or fuse the segmentations from the CNN and the Transformer methods. In addition, we use Generative Adversarial Networks (GANs) to generate synthetic abnormally shaped nuclei images as ground truth to train HER-CNN together with the limited amount of real annotated images. We demonstrate that our method achieves better abnormally shaped nuclei segmentation by comparing it to other widely used nuclei segmentation methods. We also illustrate the effectiveness of the edge awareness loss and the extra synthetic training image by conducting ablation studies.

2. Related Work

2.1. Nuclei Segmentation

Nuclei segmentation is a vital part of cell analysis in areas such as cell biology, drug discovery, functional genomics, and pathology [23]. Traditional computer vision segmentation methods such as Otsu [18, 45], Watershed [31, 42], and active contours [1, 2] have been used in automated nuclei segmentation. Otsu [35] is a threshold-based segmentation method, it automatically selects the appropriate threshold by maximizing the variance between the object and the background. Watershed [3] treats an image as a terrain in which nuclei correspond to valleys in a topographic landscape. Active contours [25] is an energy-based segmentation method to move deformable contours under the influence of forces to minimize an energy function, and therefore locate the object boundaries. One major challenge of traditional nuclei segmentation methods is that they usually require manual parameter tuning and re-parameterization when segmenting nuclei with different staining, scanner, lighting conditions, and magnifications [20].

In recent years, nuclei segmentation is mainly performed using modern machine learning approaches. Mask R-CNN [21] is an instance segmentation network developed using region-based convolutional neural networks. It performs well after training using annotated ground truth images [24]. Since acquiring a large number of annotated training images is not possible in many biomedical applications, U-Net [37] is introduced for segmentation for various biomedical problems with very little training data. Furthermore, methods such as StarDist [39, 44], Cellpose [41], DeepSynth [11], 3D Centroidnet [47], and NISNet3D [46] are specially designed for segmenting the nuclei.

Vision transformers have demonstrated good performance in general computer vision areas such as image classification [10], detection [4], and segmentation [50]. They have also been used for biomedical image analysis and nuclei segmentation. In [7], a Swin-Transformer [29] is combined with a U-Net [37] and it outperforms the CNN methods for segmenting tumors in CT scans. In [17], a ViT architecture [10] is used to classify and segment the nuclei in the histopathological images.

While many methods are available for segmenting normal elliptical nuclei, only a few studies [6, 19, 49] focused on segmenting the abnormally shaped nuclei.

2.2. Ensemble Methods

Ensemble methods, where outputs from several segmentation methods are combined or fused have been used for improving segmentation performance [16]. In [32, 48], the stacking of CNN architectures is used to improve the performance of the classification of cancer tumors and nuclei segmentation. Since CNN architectures are better at extracting detailed localization information while Transformer architectures are better at modeling explicit long-range relations, ensemble methods are frequently used to benefit from both detailed high-resolution spatial information from CNNs and the global context encoded by Transformers. In [30], a cross-teaching between CNN and Transformer is introduced for medical image segmentation.

2.3. Synthetic Ground Truth Image Generation

Lacking ground truth labeled images is a common challenge of deep learning-based methods. One way to address this problem is by using data augmentation methods to create more training samples. Traditional data augmentation methods utilize linear and non-linear transformations including flipping, random cropping, color space transformations, and elastic deformations [5, 33]. However, these augmentation methods do not work well when the training data is limited and cannot solve the data imbalance problem within the dataset. With the development of Generative Adversarial Networks (GANs), generating synthetic images from GANs for data augmentation is widely used for various deep learning methods. GANs have been used to generate synthetic CT scan images for training to improve the classification of liver lesions [14]. An auxiliary classifier based on a GAN is used for data augmentation of chest X-ray images for improving Covid-19 detection [43]. The Sp-

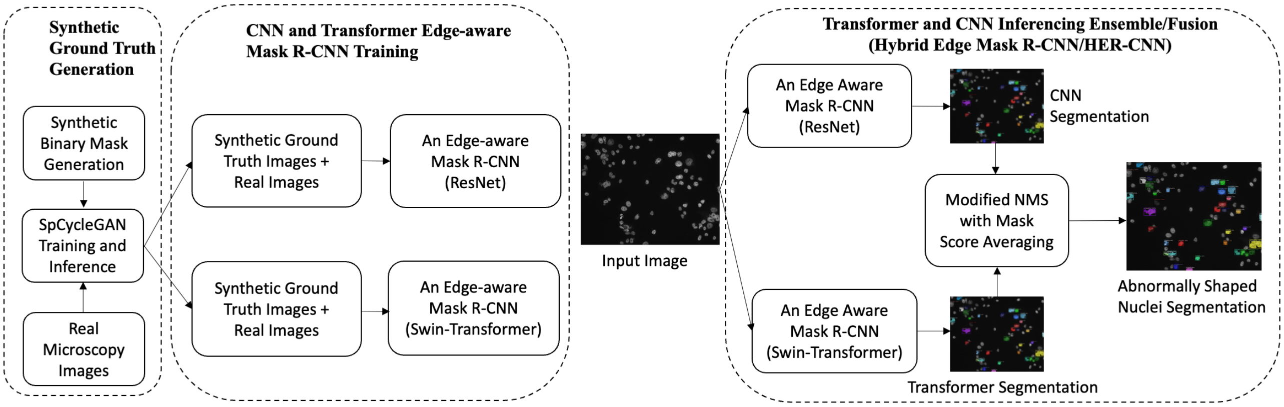

Synthetic Ground Truth Generation
Synthetic Binary Mask Generation
SpCycleGAN Training and Inference
Real Microscopy Images
CNN and Transformer Edge-aware Mask R-CNN Training
Synthetic Ground Truth Images + Real Images
An Edge-aware Mask R-CNN (ResNet)
Synthetic Ground Truth Images + Real Images
An Edge-aware Mask R-CNN (Swin-Transformer)
Input Image
Transformer and CNN Inferencing Ensemble/Fusion (Hybrid Edge Mask R-CNN/HER-CNN)
An Edge Aware Mask R-CNN (ResNet)
CNN Segmentation
Modified NMS with Mask Score Averaging
Abnormally Shaped Nuclei Segmentation
An Edge Aware Mask R-CNN (Swin-Transformer)
Transformer Segmentation

Figure 1. The block diagram of the proposed Hybrid Edge Mask R-CNN (HER-CNN) for abnormally shaped nuclei segmentation. The training system is on the left and the inferencing system is on the right.

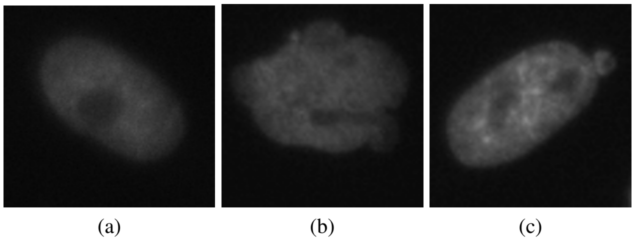


Figure 2. The images of (a): a typical elliptical nucleus, (b) an abnormally shaped nucleus with large differences in shape, (c) an abnormally shaped nucleus which looks similar to elliptical shape but with a small bump on its surface.

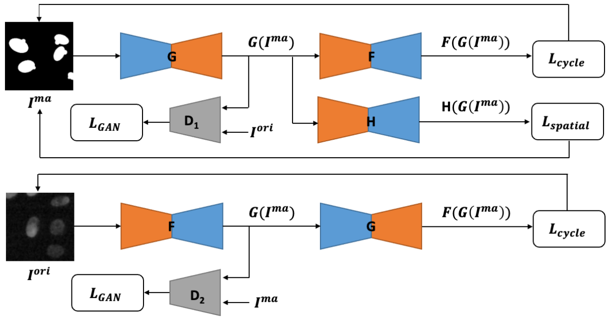


Figure 3. The architecture of the SpCycleGAN [15].

CycleGAN [11,15,47] is proposed to generate the synthetic microscopy images with spatial constraints relative to the nuclei location.

3. Proposed Method

The block diagram of our proposed abnormally shaped nuclei segmentation system is shown in Figure 1. The system includes (1) Synthetic ground truth image generation for training, (2) Edge-aware Mask R-CNNs with ResNet and Swin-Transformer architectures, trained with mixed synthetic and real images, (3) Transformer and CNN segmentation masks fusion done by a modified version of Non-Maximum Suppression (NMS) with mask score averaging, and (4) Final fused segmentation.

Our goal is to locate a specific cell condition, which we are calling an unhealthy cell. Unhealthy cells include cells that may be dying or stressed from external stimuli during experiments. We desire to avoid fluorescent labeling of the cell membrane, thus only the nuclei are being labeled. Additionally, some cells are out of focus due to natural limitations in cell positioning in the imaging system, as well as limitations in the microscope. For these reasons, we choose to segment unhealthy cells' nuclei for counting and localization. There is a wide range of differences between typical elliptical nuclei and abnormally shaped nuclei of the unhealthy cells in our experimental dataset. While most of the abnormally shaped nuclei have significant differences from the typical elliptical nuclei, the most difficult abnormally shaped nuclei are the ones that are elliptical in shape with additional bumps and protrusions. This specific irregularity in nucleus shape can be caused by cell stress which can lead to nuclei fragmentation and blebbing. Blebbing of the plasma membrane is a morphological feature of cells undergoing late stage apoptosis (cell death). A bleb is an irregular bulge in the plasma membrane of a cell. Figure 2 shows the images of a typical elliptical nucleus and two abnormally shaped nuclei. Our goal is to segment the abnormally shaped nuclei out of the typical elliptical nuclei.

3.1. Synthetic Ground Truth Image Generation

Since a specific shape of abnormal nuclei is challenging to segment and ground truth images for this specific shape of abnormal nuclei are hard to obtain, using SpCycleGAN [15], we generate synthetic images with annotations and use them as the ground truth images to add to our training dataset. The synthetic ground truth image generation consists of synthetic binary map generation (used as annotated segmentation mask), SpCycleGAN training, and inferences (left side of Figure 1). To generate the synthetic binary maps of this specific shape of abnormal nuclei, we first generate elliptical nuclei, then generate a "circular bump" on the generated nuclei, where the center of the circular bump is located on the contour of nuclei. The sizes and the numbers of nuclei and bumps are randomly chosen from ranges based on observing actual microscopy images. We add constraints to make sure the generated nuclei are not overlapping. We denote these synthetic binary maps as I^{ma}, the original microscopy images as I^{ori}, and the generated synthetic microscopy images of nuclei as I^{syn}.

SpCycleGAN [15] is an extension of CycleGAN [51] with spatial constraints added to the loss function. CycleGAN combines two generators, where G translates a binary segmentation mask to a microscopy image, F is for reverse translation of G, and two adversarial discriminators D_1 and D_2 are learned to make the two domain translations indistinguishable. SpCycleGAN adds one more network H for maintaining the spatial location between I^{ma} and $F(G(I^{ma}))$. The architecture of the SpCycleGAN is shown in Figure 3, and the loss function of SpCycleGAN is shown in Equation 1.

$$\begin{aligned}\mathcal{L}(G,F,H,D_1,D_2) = \mathcal{L}_{GAN}(G,D_1,I^{ma},I^{ori})\\ +\mathcal{L}_{GAN}(F,D_2,I^{ori},I^{ma})\\ +\lambda_1\mathcal{L}_{cycle}(G,F,I^{ori},I^{ma})\\ +\lambda_2\mathcal{L}_{spatial}(G,H,I^{ori},I^{ma})\end{aligned} \tag{1}$$

Here $\mathcal{L}_{GAN}$ is the adversarial loss, λ_1 and λ_2 are the weight coefficients to control the loss balance between the cycle consistency loss $\mathcal{L}_{cycle}$ and the spatial loss $L_{spatial}$ proposed in SpCycleGAN. The spatial loss $L_{spatial}$ is defined in Equation 2.

$$\mathcal{L}_{spatial}(G,H,I^{ori},I^{ma}) = \mathbb{E}_{I^{ma}}[||H(G(I^{ma}))-I^{ma}||_2] \tag{2}$$

Here $||\cdot||_2$ is L_2 is the norm. The addition of the spatial loss is to ensure that the generated microscopy image has the nuclei in the correct position according to the binary segmentation map.

The SpCycleGAN is trained with unpaired real microscopy images I^{ori} and synthetic binary segmentation maps I^{ma}. It then generates synthetic microscopy images of nuclei I^{syn} corresponding to the I^{ma}. Therefore, I^{syn} can be used as annotated ground truth images for training the abnormally shaped nuclei segmentation network.

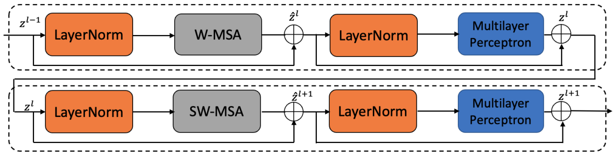


Figure 4. The Swin-Transformer block [29].

3.2. CNN and Transformer Mask R-CNN

Mask R-CNN [21] is a two-stage instance segmentation network. The first stage of Mask R-CNN uses a region proposal network (RPN) to generate bounding boxes indicating the potential objects, or we call them the regions of interest (ROIs). Then, the second stage is used to classify and further refine the bounding box of the object, finally providing the segmentation mask of the object inside the detected bounding box. The architecture of the Mask R-CNN is the feature pyramid network (FPN) [27]. FPN is designed to have lateral connections between each layer of the bottom-up and top-down convolutional layers to provide the predictions at multiple levels of the feature maps, thus maintaining strong semantically meaningful features at various resolution scales.

While the ResNet [22] is frequently used with Mask R-CNN in numerous applications, we also combine the Swin-Transformer [29] with Mask R-CNN to better capture global information and improve performance. The Swin-Transformer contains a patch partition module at the top, which splits an input image into non-overlapping patches, then a linear embedding layer is used to project the patches into the designed dimension of the transformer. Multiple stages are connected together after the linear embedding layer, each with a patch merging module and a Swin-Transformer block to perform hierarchical learning similar to FPN. The Swin-Transformer block is illustrated in Figure 4, it consists of a LayerNorm (LN) layer before the regular or shifted window attention module (W-MSA/SW-MSA) and each multi-layer perceptron module (MLP). There is a residual connection after each module. The regular window attention module (W-MSA) is designed to compute the self-attention within local non-overlapped windows, therefore reducing the computational complexity. The shifted window attention module (SW-MSA) is designed to compute the self-attention across the windows, thus maintaining the strong modeling power of the global information.

Assume that a 2D microscopy image I^{ori} with dimension $H \times W$ with D color channels is input into the Swin-Transformer. The patch partition module will split the image into $P \times P$ patches, where P is the patch size, and the number of patches is $N = \frac{H}{P} \times \frac{W}{P}$. The collection of these patches will go through the linear embedding to project into a 2D matrix $\mathbf{Z} \in \mathbb{R}^{N \times C}$, where C is the designed dimension

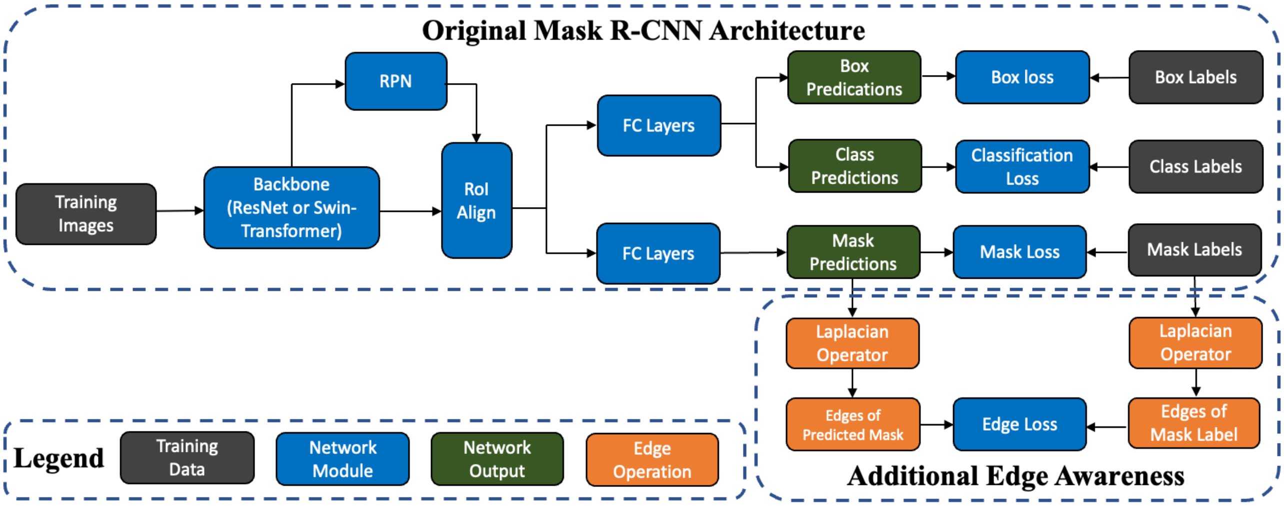


Figure 5. The block diagram of our proposed Edge-aware Mask R-CNN. The color indicates different operations, as shown in the legend.

of the transformer, and each row of $\mathbf{Z}$ is a patch embedding vector. Then the 2D matrix $\mathbf{Z}$ is used as input into the Swin-Transformer blocks. The computation is shown in Equation 3.

$$\begin{aligned}\hat{\mathbf{Z}}^l &= W-MSA(LN(\mathbf{Z}^{l-1}))+\mathbf{Z}^{l-1}\\ \mathbf{Z}^l &= MLP(LN(\hat{\mathbf{Z}}^l))+\hat{\mathbf{Z}}^l\\ \hat{\mathbf{Z}}^{l+1} &= SW-MSA(LN(\mathbf{Z}^l))+\mathbf{Z}^l\\ \mathbf{Z}^{l+1} &= MLP(LN(\hat{\mathbf{Z}}^{l+1}))+\hat{\mathbf{Z}}^{l+1}\end{aligned} \tag{3}$$

Here $\hat{\mathbf{Z}}^l$ is the output of the (S)W-MSA module of block l and $\mathbf{Z}^l$ is the output of the MLP module of block l.

While Transformer architectures are better at modeling explicit long-range relations, CNN architectures are better at extracting detailed localization information. Therefore, we use a Mask R-CNN with ResNet and a Mask R-CNN with Swin-Transformer for abnormally shaped nuclei segmentation and fuse the results to achieve better performance.

3.3. Edge-aware Mask R-CNN

The main difference between abnormal nuclei and typical elliptical nuclei is the shape of the edge, which sometimes are hard to distinguish (described in Section 3). We propose combining an additional edge loss L_{edge} with the existing Mask R-CNN to address this problem. The block diagram of our proposed Edge-aware Mask R-CNN is shown in Figure 5.

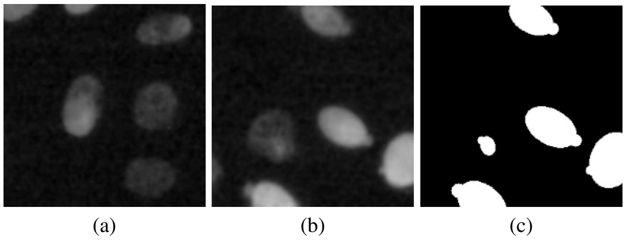
To estimate this additional edge loss L_{edge}, we threshold the mask predictions of the Mask R-CNN y_{mask} and the corresponding ground truth mask labels gt_{mask} by 0 and use an edge operator on the masks to obtain the corresponding edge images y_{edge} and gt_{edge}. We use the Laplacian operator as our edge operator to extract the edges. Since this edge loss L_{edge} is an additional edge awareness term in the original mask loss L_{mask} of the Mask R-CNN, we use the same category-specific binary cross-entropy (BCE) loss used for the mask loss L_{mask} to find the L_{edge} between the gt_{edge} and y_{edge}. The final multi-task loss L is shown in Equation 4.

$$L = L_{cls} + L_{box} + L_{mask} + \lambda L_{edge} \tag{4}$$

Here the classification loss L_{cls}, bounding box regression loss L_{box}, and mask loss L_{mask} are identical to Mask R-CNN. We propose appending the edge loss L_{edge} as additional edge awareness to the multi-task loss; and λ is the weight coefficient.

3.4. CNN and Transformer Ensemble Processing

To fuse the outputs from the CNN and the Transformer Mask R-CNNs, we use a modified version of Non-Maximum Suppression (NMS) [34] with confidence score averaging to combine the output segmentations from the individual Mask R-CNNs. We have M Mask R-CNNs, and the $i-th$ Mask R-CNN's output is denoted as $Det_i^n = \{Seg_i^n, Prob_i^n\}$, where Seg_i^n is a segmentation mask and $Prob_i^n$ is its corresponding confidence score, n is the index of a detected nucleus and $i \in \{1, ..., M\}$. Our goal is to generate a refined final segmentation Det^n based on Det_i^n. First, we find the Intersection over Union (IoU) of the segmentation masks from each Mask R-CNN Seg_i^n. We then use Non-Maximum Suppression with a threshold of τ to construct the final segmentation mask Seg^n. Unlike the original NMS which uses the highest confidence score to find the final confidence score of the mask, we use the average confidence score to find the final confidence score



(a) (b) (c)

Figure 6. Examples of (a): the cropped original microscopy image, (b): the generated microscopy image of abnormally shaped nuclei, and (c): its corresponding synthetic segmentation mask.

$Prob^n$. This ensures that each Mask R-CNN's output is contributing to the final segmentation. Equation 5 shows our modified NMS with mask score averaging.

$$\begin{aligned} Seg^n &= NMS(Seg_i^n, \tau) \\ Prob^n &= \frac{1}{M}\sum_{i=0}^{M} Prob_i^n \end{aligned} \tag{5}$$

Here $NMS(\cdot, \tau)$ is the Non-Maximum Suppression [34] with threshold IoU= τ.

4. Experimental Results

4.1. Datasets

We manually annotated abnormally shaped nuclei on 50 fluorescence microscopy images (Figure 7a). All the nuclei are from CHO-K1 cells stained with ”Hoechst 33342 stain” (Note only the nuclei are stained and not the cell membranes). The images were captured at 200ms exposure with a DAPI filter cube on the microscope. The size of each image is 2758×2208 pixels. The images were divided into 30, 5, and 15 images for training, validation, and testing. For training HER-CNN, we used the SpCycleGAN to generate synthetic ground truth images for the challenging case of abnormally shaped nuclei (described in Section 3) and combined them with the 30 real training images. Each synthetic ground truth image is 256×256 pixels. The training dataset of the Edge-aware Mask R-CNN with ResNet consists of 300 synthetic ground truth images with 30 real images, and the training dataset for the Edge-aware Mask R-CNN with Swin-Transformer consists of 400 synthetic ground truth images with the same 30 real images.

4.2. Experiments

SpCycleGAN training and inference. The training of the SpCycleGAN requires unpaired original microscopy images I^{ori} and synthetic binary maps I^{mask}. We randomly select 5 images with high nuclei density from the training dataset and crop 200 256×256 image patches as I^{ori} for training the SpCycleGAN. We then generate 200 256×256 synthetic binary maps as I^{mask}, using the method described in Section 3.1. The generated 200 binary maps I^{mask} and 200 cropped image patches are used to train SpCycleGAN. We then generate extra 500 256×256 synthetic binary maps and feed them into trained SpcyCleGAN to generate synthetic abnormally shaped nuclei images. The examples of the generated abnormally shaped nuclei images are shown in Figure 6.

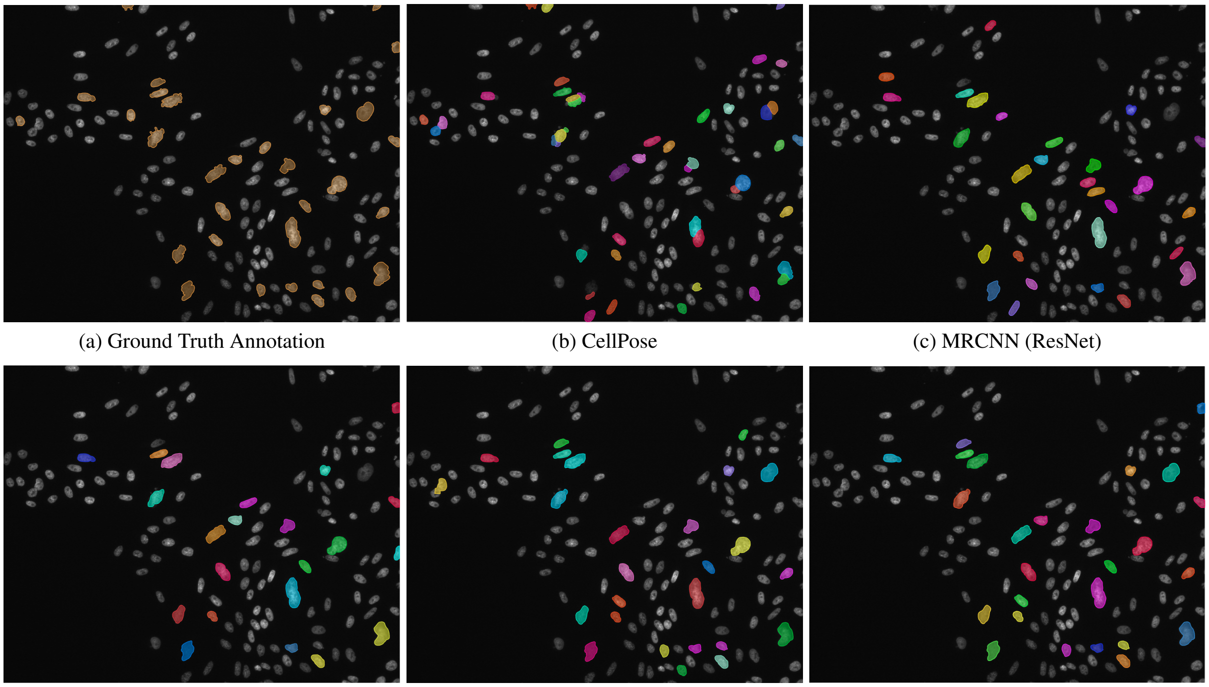
Method	AP50	AP75	mAP
U-Net [37]	47.17	36.22	29.21
StarDist [44]	47.64	41.34	32.42
CellPose [41]	63.33	55.97	44.67
MRCNN (ResNet) [21]	65.27	63.41	51.38
Edge MRCNN (ResNet)+syn	73.74	72.87	58.77
Edge MRCNN (Swin)+syn	73.83	71.36	57.42
HER-CNN	**76.07**	**75.04**	**61.35**

Table 1. This table shows the performance metric of the instance-based abnormally shaped nuclei segmentation using AP50, AP75, and mAP as described in Section 4.3. For simplicity, Mask R-CNN is written as ”MRCNN”, Edge-aware Mask R-CNN is written as ”Edge MRCNN”, and methods trained with the extra synthetic ground truth images are denoted as ”+syn”. The higher number indicates better performance.

Hybrid Edge Mask R-CNN (HER-CNN) training and inference. The block diagram of the training and inference of Hybrid Edge Mask R-CNN is shown in Figure 1. In our experiments, we train an Edge-aware Mask R-CNN with ResNet-50 as the backbone architecture and an Edge-aware Mask R-CNN with Swin-S as the backbone architecture. Both architecture networks have been pre-trained on the ImageNet [9] dataset and then trained with our data which consists of real microscopy images and generated synthetic ground truth images. Both the Edge-aware Mask R-CNNs are trained on an NVIDIA RTX A5000 GPU with a batch size of 8, the learning rate of the Edge-aware Mask R-CNN with ResNet-50 is $5e^{-4}$ and it is trained for 200 epochs, the learning rate of the Edge-aware Mask R-CNN with Swin-S is $5e^{-7}$ and it is trained for 2000 epochs. The weight λ for the Edge-aware loss L_{edge} is set to 0.2 for both models. The IoU matching threshold for our modified NMS is set to 0.7.

4.3. Evaluation

To evaluate HER-CNN, we define that a segmentation mask is considered as a true positive (TP) if the ground truth mask and it have an Intersection over Union (IoU) score greater than a given threshold. If a segmentation mask does not have an IoU score greater than a given threshold with any ground truth mask, it will be considered as a false positive (FP). A ground truth mask not being segmented will be considered as a false negative (FN). Precision and Recall are



(a) Ground Truth Annotation

(b) CellPose

(c) MRCNN (ResNet)

(d) Edge MRCNN (ResNet)+syn

(e) Edge MRCNN (Swin)+syn

(f) HER-CNN

Figure 7. Examples of abnormally shaped nuclei segmentation for HER-CNN and comparison methods with ground truth annotation. The confidence threshold for each method is set differently to perform better segmentation. The different nuclei instances are shown in different colors.

defined as $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$. In order to have a generalized evaluation over a wide range of confidence score thresholds and IoU thresholds, we adopt the widely used Average Precision and the mean Average Precision metrics. The Average Precision (AP) is defined as the area under the Precision-Recall curve [8], which is drawn by calculating Precision and Recall with a set of confidence score thresholds from 0 to 1 with a 0.1 increment. The mean Average Precision (mAP) is defined as the average of AP scores for a set of IoU thresholds. We choose three commonly used thresholds for AP and mAP, which are AP with IoU threshold at 0.5 (AP50), AP with IoU threshold at 0.75 (AP75), and mAP with a set of IoU thresholds from 0.5 to 0.95 with a 0.05 increment (mAP). AP50 is the evaluation metric for the PASCAL VOC Object Detection Challenges [12], AP75 is a stricter metric that is used for the MS COCO Challenge [28], and mAP is the most generalized metric among these three and is used as the main evaluation metric for the MS COCO Challenge [28]. The evaluation results of our proposed method with other comparison methods are shown in Table 1. The comparison methods are only trained with real microscopy images from our training dataset. For the evaluation of U-Net segmentation, since it is a semantic segmentation method, each closed contour on the U-Net segmentation mask will be considered as an instance. The abnormally shaped nuclei segmentation results of HER-CNN and the comparison methods with ground truth annotation are shown in Figure 7.

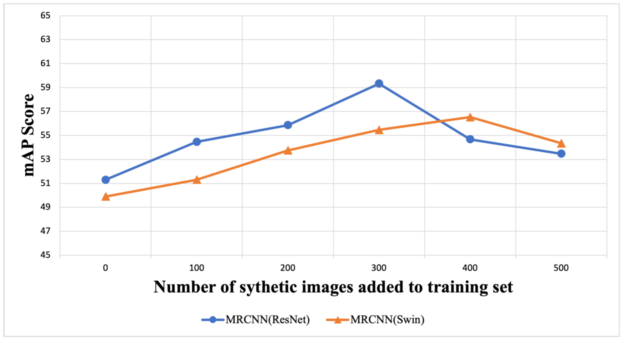


Figure 8. Mean Average Precision (mAP) as a function of the numbers of generated microscopy images in the training set.

4.4. Ablation Study

The use of synthetic ground truth images in training . In our proposed HER-CNN, in order to improve the abnormally shaped nuclei segmentation, we synthetically generate microscopy images with abnormal nuclei that are difficult to segment as described in Section 3.1. These generated images will be combined with real images to train the Edge-aware Mask R-CNNs. Since the synthetically generated

Method	AP50	AP75	mAP
MRCNN	64.73	63.11	51.46
Edge MRCNN (λ=0.1)	67.10	64.69	52.97
Edge MRCNN (λ=0.2)	**69.37**	**66.61**	**54.48**
Edge MRCNN (λ=0.3)	67.03	66.43	52.94

Table 2. This table shows abnormally shaped segmentation performance with MRCNN and Edge MRCNN with different weights on edge loss using AP50, AP75, and mAP as described in Section 4.3.

images are not guaranteed to be an accurate representation of the real images, the segmentation networks may learn the errors in the synthetic images. Also, our generated microscopy images are only for one type of abnormal nuclei, too many synthetic images in the training set may cause a class imbalance issue between various shapes of the abnormal nuclei. Therefore, we conducted an ablation study to determine how many synthetic images should be combined with the real images in the training data. We randomly split the 500 generated synthetic abnormally shaped nuclei images (described in Section 4.2) into 5 batches each having 100, 200 300, 400, and 500 images. We then combine our 30 real training images with these 5 batches to create 5 training sets. We train Mask R-CNNs with ResNet-50 and with Swin-S and evaluate them using the validation dataset (described in Section 4.1), the results are shown in Figure 8.

As can be seen in Figure 8, both the CNN and Transformer Mask R-CNNs have performance gains after synthetic ground truth microscopy images combined with the real training dataset, which indicates the effectiveness of the use of the synthetic training images. However, the performance of both Mask R-CNNs starts to decrease as the number of generated microscopy images in the training set increases. This number is 300 for the Mask R-CNN with ResNet-50, and 400 for the Mask R-CNN with Swin-S. Therefore, in our experiments, we use 300 synthetic ground truth images combined with real images to train the Edge-aware Mask R-CNN with ResNet-50. We use 400 synthetic ground truth images combined with real images to train the Edge-aware Mask R-CNN with Swin-S.

The edge loss L_{edge} in the Edge-aware Mask R-CNN. The edge loss L_{edge} in our proposed method is designed to add additional edge awareness to the original Mask R-CNN, therefore it can better distinguish between the abnormally shaped nuclei and the typical elliptical nuclei. However, if the weight λ of the edge loss L_{edge} is too big, then this loss will overwhelm the multi-task loss function described in Section 3.3, resulting in inaccurate segmentation. For this reason, we conducted an additional ablation study to determine the proper weight λ of the edge loss L_{edge} used in our HER-CNN. We use real images to train the Edge-aware Mask R-CNNs with different weights λ and evaluate the result using the validation data. For this experiment, we only use ResNet-50 as the backbone architecture, since both the CNN and Transformer Edge-aware Mask R-CNN share the same multi-task prediction structure. The results are shown in Table 2. With the proper weights for the additional edge loss to the Mask R-CNN, the Edge-aware Mask R-CNN performs better than the original Mask R-CNN for abnormally shaped nuclei. Furthermore, we can see when $\lambda = 0.2$, the Edge-aware Mask R-CNN performs the best among other weight settings. Therefore, in our experiments, we use $\lambda = 0.2$ for HER-CNN.

4.5. Discussion

In Table 1, the original Mask R-CNN performs the best among the four comparison methods, for this reason, we use Mask R-CNN for HER-CNN. Although the CNN and the Transformer Edge-aware Mask R-CNNs show similar performance in Table 1, there are many differences in the actual segmentation masks. Comparing Figure 7d and Figure 7e, we can see that the false positives and the false negatives in the two masks are different although they are trained with similar data. Therefore, the ensemble processing in HER-CNN improves the performance for abnormally shaped nuclei segmentation by fusing these two masks from CNN and Transformer. HER-CNN was compared with other methods used for nuclei segmentation including U-Net, StarDist, CellPose, and Mask R-CNN. The evaluation (Table 1) based on the AP50, AP75, and mAP scores shows HER-CNN outperforms other methods.

5. Conclusion and Future Work

In this paper, we described Hybrid Edge Mask R-CNN (HER-CNN) that uses ensemble processing to fuse the results from the Transformer and CNN Edge-aware Mask R-CNNs for abnormally shaped nuclei segmentation. The evaluations demonstrate HER-CNN outperforms other CNN approaches for abnormally shaped nuclei segmentation. The ablation studies indicate the effectiveness of the use of synthetic training images and the Edge-aware Mask R-CNN. Note that using synthetic ground truth images, HER-CNN can perform very well for abnormally shaped nuclei segmentation with only 30 annotated images.

In the future, we will focus on improving the quality of the synthetically generated ground truth images, which could reduce the need for real annotated images. We will also investigate semi-supervised and weakly-supervised learning approaches, which would allow one to learn from unlabelled images.

6. Acknowledgments

This work was supported by HP Inc. Address all correspondence to Edward J. Delp at ace@ecn.purdue.edu.

References

[1] Khamael Al-Dulaimi, Inmaculada Tomeo-Reyes, Jasmine Banks, and Vinod Chandran. White blood cell nuclei segmentation using level set methods and geometric active contours. *Proceedings of the International Conference on Digital Image Computing: Techniques and Applications*, pages 1–7, November 2016. Gold Coast, Australia. 2

[2] Pascal Bamford and Brian Lovell. Unsupervised cell nucleus segmentation with active contours. *Signal processing*, 71(2):203–213, December 1998. 2

[3] Serge Beucher. The watershed transformation applied to image segmentation. *Scanning Microscopy*, 1992(6):28, 1992. 2

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. *Proceedings of the European Conference on Computer Vision*, pages 213–229, August 2020. Glasgow, United Kingdom (Virtual). 2

[5] Eduardo Castro, Jaime S Cardoso, and Jose Costa Pereira. Elastic deformations for data augmentation in breast cancer mass detection. *Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics*, pages 230–234, March 2018. Las Vegas, NV. 2

[6] Chin-Wen Chang, Ming-Yu Lin, Horng-Jyh Harn, Yen-Chern Harn, Chien-Hung Chen, Kun-His Tsai, and Chi-Hung Hwang. Automatic segmentation of abnormal cell nuclei from microscopic image analysis for cervical cancer screening. *Proceedings of the IEEE 3rd International Conference on Nano/Molecular Medicine and Engineering*, pages 77–80, October 2009. Tainan, Taiwan. 2

[7] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, February 2021. 2

[8] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. *Proceedings of the 23rd International Conference on Machine learning*, pages 233–240, June 2006. Pittsburgh, PA. 7

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. Miami, FL. 6

[10] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, June 2020. 1, 2

[11] Kenneth W Dunn, Chichen Fu, David Joon Ho, Soonam Lee, Shuo Han, Paul Salama, and Edward J Delp. Deepsynth: Three-dimensional nuclear segmentation of biological images using neural networks trained with synthetic data. *Scientific Reports*, 9(1):1–15, December 2019. 2, 3

[12] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111:98–136, 2015. 7

[13] Edgar G Fischer. Nuclear morphology and the biology of cancer cells. *Acta Cytologica*, 64(6):511–519, June 2020. 1

[14] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. *Proceedings of the IEEE 15th International Symposium on Biomedical Imaging*, pages 289–293, April 2018. Washington, DC. 2

[15] Chichen Fu, Soonam Lee, David Joon Ho, Shuo Han, Paul Salama, Kenneth W Dunn, and Edward J Delp. Three dimensional fluorescence microscopy image synthesis and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2221–2229, June 2018. Salt Lake City, UT. 3, 4

[16] Mudasir A Ganaie, Minghui Hu, et al. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence*, 115:105151, October 2022. 2

[17] Zeyu Gao, Bangyang Hong, Xianli Zhang, Yang Li, Chang Jia, Jialun Wu, Chunbao Wang, Deyu Meng, and Chen Li. Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image. *Proceedings of the Medical Image Computing and Computer Assisted Intervention*, pages 299–308, September 2021. Strasbourg, France. 2

[18] Yasmeen M George, Bassant M Bagoury, Hala H Zayed, and Mohamed I Roushdy. Automated cell nuclei segmentation for breast fine needle aspiration cytology. *Signal Processing*, 93(10):2804–2816, October 2013. 2

[19] Yue Han, Yang Lei, Viktor Shkolnikov, Daisy Xin, Alicia Auduong, Steven Barcelo, and Edward J Delp. Ensemble processing and synthetic image generation for abnormally shaped nuclei segmentation. *bioRxiv*, pages 2023–01, January 2023. 2

[20] Tomohiro Hayakawa, VB Surya Prasath, Hiroharu Kawanaka, Bruce J Aronow, and Shinji Tsuruoka. Computational nuclei segmentation methods in digital pathology: A survey. *Archives of Computational Methods in Engineering*, 28:1–13, January 2021. 2

[21] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, December 2017. Venice, Italy. 2, 4, 6

[22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, June 2016. Las Vegas, NV. 4

[23] Reka Hollandi, Nikita Moshkov, Lassi Paavolainen, Ervin Tasnadi, Filippo Piccinini, and Peter Horvath. Nucleus segmentation: Towards automated solutions. *Trends in Cell Biology*, 32:295–310, April 2022. 1, 2

[24] Jeremiah W Johnson. Automatic nucleus segmentation with mask-rcnn. *Proceedings of the Computer Vision Conference*, pages 399–407, April 2019. Las Vegas, NV. 1, 2

[25] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988. 2

[26] Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–41, September 2022. 1

[27] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2117–2125, July 2017. Honolulu, HI. 4

[28] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. *Proceedings of the European Conference on Computer Vision*, pages 740–755, September 2014. Zurich, Switzerland. 7

[29] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, October 2021. Montreal, Canada. 1, 2, 4

[30] Xiangde Luo, Minhao Hu, Tao Song, Guotai Wang, and Shaoting Zhang. Semi-supervised medical image segmentation via cross teaching between cnn and transformer. *Proceedings of the International Conference on Medical Imaging with Deep Learning*, pages 820–833, July 2022. Zürich, Switzerland. 2

[31] Norberto Malpica, Carlos Ortiz De Solórzano, Juan José Vaquero, Andrés Santos, Isabel Vallcorba, José Miguel García-Sagredo, and Francisco Del Pozo. Applying watershed algorithms to the segmentation of clustered nuclei. *Cytometry: The Journal of the International Society for Analytical Cytology*, 28(4):289–297, August 1997. 2

[32] Mohanad Mohammed, Henry Mwambi, Innocent B Mboya, Murtada K Elbashir, and Bernard Omolo. A stacking ensemble deep learning approach to cancer type classification based on tcga data. *Scientific Reports*, 11(1):1–22, August 2021. 2

[33] Daniel Mas Montserrat, Qian Lin, Jan Allebach, and Edward J Delp. Training object detection and recognition cnn models using data augmentation. *Proceedings of the IS&T International Symposium on Electronic Imaging*, 2017(10):27–36, January 2017. Burlingame, CA. 2

[34] Alexander Neubeck and Luc Van Gool. Efficient non-maximum suppression. *Proceedings of the 18th International Conference on Pattern Recognition*, 3:850–855, August 2006. Hong Kong, China. 5, 6

[35] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, January 1979. 2

[36] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, December 2017. 2

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, October 2015. Munich, Germany. 2, 6

[38] Sajith Kecheril Sadanandan, Petter Ranefall, Sylvie Le Guyader, and Carolina Wählby. Automated training of deep convolutional neural networks for cell segmentation. *Scientific reports*, 7(1):7860, August 2017. 2

[39] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 265–273, September 2018. Granada, Spain. 1, 2

[40] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, July 2019. 2

[41] Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: A generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, January 2021. 1, 2, 6

[42] Mitko Veta, A Huisman, Max A Viergever, Paul J van Diest, and Josien PW Pluim. Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images. *Proceedings of the IEEE International Symposium on Biomedical Imaging*, pages 618–621, March 2011. Chicago, IL. 2

[43] Abdul Waheed, Muskan Goyal, Deepak Gupta, Ashish Khanna, Fadi Al-Turjman, and Plácido Rogerio Pinheiro. Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. *IEEE Access*, 8:91916–91923, May 2020. 2

[44] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers. Star-convex polyhedra for 3d object detection and segmentation in microscopy. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3666–3673, March 2020. Snowmass, CO. 1, 2, 6

[45] Khin Yadanar Win and Somsak Choomchuay. Automated segmentation of cell nuclei in cytology pleural fluid images using otsu thresholding. *Proceedings of International Conference on Digital Arts, Media and Technology*, pages 14–18, March 2017. Chiang Mai, Thailand. 2

[46] Liming Wu, Alain Chen, Paul Salama, Kenneth Dunn, and Edward Delp. NISNet3D: Three-Dimensional Nuclear Synthesis and Instance Segmentation for Fluorescence Microscopy Images. *bioRxiv*, June 2022. 2

[47] Liming Wu, Alain Chen, Paul Salama, Kenneth W Dunn, and Edward J Delp. 3D Centroidnet: Nuclei Centroid Detection with Vector Flow Voting. *bioRxiv*, July 2022. 2, 3

[48] Liming Wu, Alain Chen, Paul Salama, Kenneth W Dunn, and Edward J Delp. An ensemble learning and slice fusion strategy for three-dimensional nuclei instance segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1884–1894, June 2022. New Orleans, LA. 2

[49] Ling Zhang, Hui Kong, Chien Ting Chin, Shaoxiong Liu, Zhi Chen, Tianfu Wang, and Siping Chen. Segmentation of cytoplasm and nuclei of abnormal cells in cervical cytology

using global and local graph cuts. *Computerized Medical Imaging and Graphics*, 38(5):369–380, July 2014. 2

[50] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6881–6890, June 2021. Nashville, TN. 2

[51] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232, October 2017. Venice, Italy. 4