IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 27 February 2024, accepted 10 April 2024, date of publication 18 April 2024, date of current version 26 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3390726

RESEARCH ARTICLE

HDA-Net: H&E and RGB Dual Attention Network for Nuclei Instance Segmentation

YU-HAN IM[1], SEO-HYEONG PARK[1], AND SANG-CHUL LEE[1,2], (Member, IEEE)
[1]Department of Computer Science and Engineering, Inha University, Incheon 22211, South Korea
[2]DeepCardio Company Ltd., Yeonsu-gu, Incheon 21984, Republic of Korea

Corresponding author: Sang-Chul Lee (sclee@inha.ac.kr)

This work was supported by Inha University Research Fund.

ABSTRACT H&E-stained images (HSIs) are widely adopted for revealing cellular structures and capturing morphological changes in nuclear instance segmentation. Despite several studies aimed at differentiating pixels of overlapping nuclei, challenges persist due to color inconsistencies introduced by non-uniform manual staining operations. This issue leads to blurred borders and alterations in color representation within the images. This study proposes an H&E and RGB dual attention network (HDA-Net) designed to address these challenges and enhance nuclei instance segmentation accuracy. Specifically, our approach involves decomposing hematoxylin, eosin, and residual (HER) components from HSIs using color deconvolution to extract discriminative information, such as nuclei and cytoplasm, entangled in RGB images. Additionally, we propose an H&E and RGB dual attention module (HDA) and a single-source attention module (SAM) to effectively incorporate information from the decomposed HER to enhance RGB images. The HDA module utilizes cross-attention between RGB features and decomposed HER to enhance RGB representation under the guidance of the HER components. Conversely, SAM applies global channel attention to learn the correlations between channels of each RGB and HER component. In the final decoder stage, the study employs multi-task learning to capture region and centroid information of the label, providing comprehensive supervision for clustered and overlapped nuclei. Our experiments on the CoNSeP, PanNuke, and Kumar datasets demonstrated that the proposed HDA-Net outperformed existing models, showing improvements of 1.2% in AJI and 0.76% in Dice for CoNSeP, 0.7% and 0.5% for PanNuke, and 0.21% and 0.16% for Kumar, respectively.

INDEX TERMS Nuclei instance segmentation, attention mechanism, guided feature attention.

I. INTRODUCTION

H&E-stained images (HSIs) represent high-resolution digital depictions of entire histopathology glass slides, offering detailed insights into tissue and cellular structures. These structural properties are crucial for capturing morphological changes that hold clinical significance, particularly in cancer diagnosis [1]. Thus, the analysis of HSIs using diverse algorithms has gained increasing attention in digital pathology [2]. Nuclei instance segmentation is an essential technique used by computer-assisted systems, contributing to diagnostic processes, such as assessing nuclear pleomorphism and morphology. However, HSIs demonstrate significant inter- and intra-instance variability attributed to factors such as shape, size, surrounding organs, and disease types. Furthermore, the presence of cell nuclei in groups results in clustered and overlapping instances, posing a challenge for precise nuclear structure segmentation.

With the advancement of deep learning in computer vision [3], [4], convolutional neural networks (CNNs) have been adopted for medical image segmentation [5], [6], [7], [8]. However, these methods face challenges in differentiating pixels of overlapping nuclei [9]. To address this, several researchers have focused on detecting the location and size of each nucleus using bounding boxes segmenting pixels within these boxes [10], [11], [12]. However, some approaches train networks to output pixel-wise segmentation maps and boundary maps directly, indicating a separation between instances [13], [14], [15]. Furthermore, a few methods employ multi-task learning to concurrently identify

The associate editor coordinating the review of this manuscript and approving it for publication was Md. Kamrul Hasan.

objects and boundaries, leading to enhanced segmentation outcomes [9], [16]. Traditional image processing techniques, combined with CNNs, have been employed to address touching nuclei by using methods such as watershed or morphological operations [17], [18]. However, these methods focus on object separation between nuclei instances to address the issue of overlapping nuclei without considering the color inconsistency problem occurring in HSIs [19]. The color inconsistency problem is caused by non-uniform manual staining operations, which can result in blurred borders within the nuclei and alterations in the color representation of the image [20]. Considering these factors, using only HSIs represented in RGB complicates both color-based analysis and differentiation between nuclei owing to the inconsistency of staining.

This study proposes HDA-Net, a novel UNet-based model designed to harness the unique characteristics of HSIs for extracting color-invariant features, thereby addressing issues of inconsistency. Specifically, we first decompose the Hematoxylin, Eosin, and Residual (HER) components from RGB images using color deconvolution [21]. Each component contains discriminative and color-invariant information about the cell compared to RGB images. As shown in Figure 1, each cell component contains unique information.

To effectively combine and enhance the representation of HER and RGB components, we propose a HDA and a SAM. In HDA, we enhance the representation of RGB images by applying cross-attention with HER components, guiding the separation of cell elements for accurate segmentation. In each block of the encoder and decoder, SAM enables the network to learn correlations between the channels of each RGB and HER component via global channel attention. The rectified feature vector of the RGB image is subsequently forwarded to the decoder. Utilizing multi-task learning, the model addresses the nuclei instance map task and centroid distance map task, providing comprehensive supervision for the precise localization of overlapped nuclei. The contributions of this study are summarized as follows:

- We propose HDA-Net for nuclei instance segmentation, leveraging the unique characteristic of HSIs to extract color-invariant features against color inconsistency problems and segment each component of nuclei accurately.
- We propose an HDA and SAM module to utilize the information of the decomposed HER while extracting the feature of the RGB image effectively. In HDA, we make RGB representation more discriminative via cross-attention. In SAM, we enable the network to learn correlations between channels of each RGB and HER component through global channel attention.
- Extensive experiments on different benchmark datasets demonstrate that our model achieves state-of-the-art AJI and dice scores. We further justify the generalization ability of our proposed model via cross-domain evaluation.

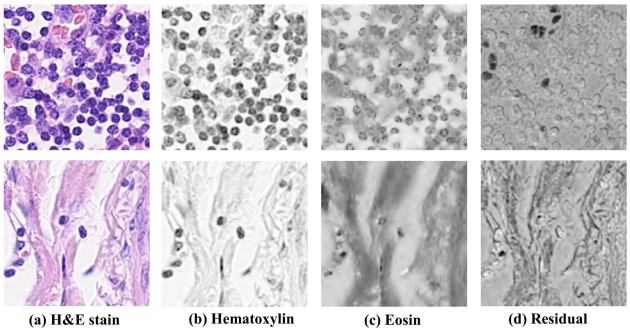

FIGURE 1. Examples of separating H&E stained images into HER components through color deconvolution. The cell nucleus is emphasized in Hematoxylin. The cytoplasm is easily detected in Eosin. However, it is challenging to detect each component of the cell in the original H&E stained image. Therefore, we decompose the H&E stained image into Hematoxylin, Eosin, and Residual, respectively.

II. RELATED WORK

A. NUCLEI INSTANCE SEGMENTATION

With advancements in computer vision, several studies have utilized CNNs for medical image segmentation, such as FCN [5], SegNet [6], PSPNet [7], and DeepLab [8]. Despite these methods accurately segment instances in medical images with complex structures, small targets, and overlapping instances remain challenging. UNet, inspired by FCN [5], addresses this by integrating low-level features from the analysis path with deep features in the expansion path using encoder-decoder skip connections, achieving a balanced the tradeoff between local and contextual information [13]. Researchers have proposed the advancements of UNet, such as ResUNet, which enhances skip connections with a dense structure through the addition of a convolution layer.

Nuclei instance segmentation has received significant attention in medical image segmentation. Nuclei instance segmentation has been adopted in various classic segmentation methodologies and algorithms. Early works utilized classical computer vision algorithms, such as thresholding [22] and morphological operations [23]. However, because of the inter- and intra-instance variability displayed by the nuclei, they cannot establish reliable thresholds in a complex background. To address this issue, various methods have adopted CNNs to automatically segment instances, which can be divided into three primary streams.

First, detection-based instance segmentation methods detect instances using a CNN and then segment the instances with each bounding box [10], [11], [12]. Yi et al. detected five predefined points in a cell using keypoint detection and then grouped these points according to a keypoint graph [10]. SpaNet captures the spatial information in a multiscale manner [12]. Secondly, semantic-to-instance segmentation methods train the network to output a pixel-wise segmentation map and a boundary map in end-to-end manner [13], [14], [15]. DCAN [14] leverages a dual architecture to form two separate prediction maps simultaneously: one for nuclear contour

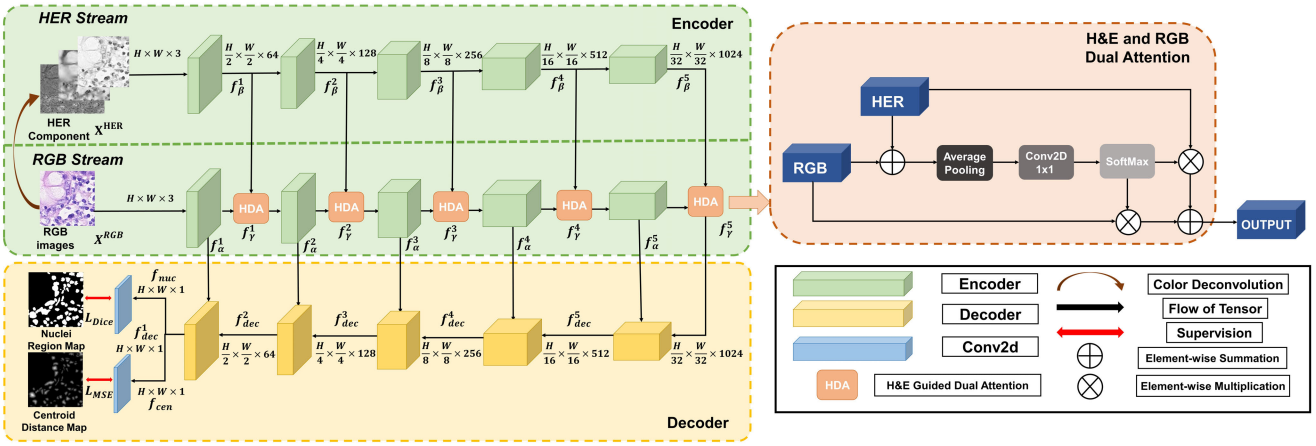

FIGURE 2. The overview of our proposed model. An input RGB image is first decomposed into H, E, and R components by color deconvolution. In the HER stream, the encoder extracts feature f_β from HER components and passes the result to the HDA module. In the RGB stream, encoders extract f_α from RGB images and pass the result to the HDA module. Given the features of each encoder f_β, f_α, generate fused features, utilizing global channel attention. These features are passed to the next encoder in the RGB stream. In decoder, f_α are skip connected to decoder blocks and generate f_{dec}. Finally, at the last decoder block generates f_{nuc}, f_{cen}, and we calculate the multi-task loss (region and centroid loss).

and the other for nuclear prediction. HoVerNet [15] predicts information encoded using horizontal and vertical distances from the center, thereby delineating the nuclei boundaries. Finally, several methods have deployed auxiliary tasks to segment objects and boundaries more accurately [9], [16], [24]. NCNet [16] also utilizes multi-task learning to predict centroids, which is an effective approach to resolving the overlapping nuclei problem. Similarly, CDNet [9] introduced auxiliary task learning using a direction-difference map to recognize the boundaries of nuclei among adjacent pixels. USEnet [24] focuses on predicting the shape markers.

In this study, we propose an HDA-Net for nuclei instance segmentation, which utilizes the unique characteristics of HSIs to extract color-invariant features against the color inconsistency problem that previous methods have often ignored.

B. ATTENTION MECHANISM

The attention mechanism is inspired by human perception and visual cognition and has been commonly applied in machine learning. The attention mechanism was first applied to natural language processing (NLP) [25], which greatly improved capacity. Accordingly, computer vision researchers have applied different strategies using attention mechanisms to the deep neural networks [26], [27], [28]. Residual attention network incorporates both an attention mechanism and residual units [26]. SENet employs squeeze-and-excitation blocks to enable a network to perform dynamic channel-wise feature recalibration [27]. The CBAM combines spatial and channel attention mechanisms to integrate diverse semantic information [28]. These Attention mechanisms can be incorporated into medical image segmentation. The attention UNet integrates the attention module into the skip connection of the UNet [29]. ResUnet deploys multiple Attention Modules that generate attention-aware features for U-net [30]. Deep attentional features (DAF) suppress the background noise of the low-level features of CNN, thereby increasing the target detail information of high-level features [31]. Hou et al. [32] proposed a feature fusion algorithm based on an attention mechanism and used it for aortic segmentation. Fan et al. [33] proposed guiding a network for lung infection segmentation of the boundary information of the target area using an edge attention mechanism. In this study, we propose HDA and SAM to incorporate the information from decomposed HER and RGB images via an attention mechanism.

III. PRELIMINARY

A. HER COMPONENT EXTRACTION

In H&E staining, hematoxylin stains the cell nuclei blue, and eosin stains the extracellular matrix and cytoplasm pink. Color deconvolution (CD) is used to separate the hematoxylin, eosin, and residual (HER) components from the original RGB image. We aim to use the unique characteristics of H&E staining to guide the segmentation network. We applied a color decomposition technique [21] to the original RGB image. Because of the nonlinearity of the relative intensity in each channel, the intensity values $I \in$ RGB of the image cannot be used directly for stain color decomposition. Thus, we used the optical density of the specimen OD, which is linear with the concentration of the absorbing stain, as follows [34]:

$$OD = -log\left(\frac{I}{I_0}\right), \tag{1}$$

where I_0 is the intensity of light entering the specimen. The deconvolution matrix D is determined from the stain color matrix M as $D = M^{-1}$. Every row of M^{-1} represents a distinct stain, and each column represents OD as detected by each RGB channel. M can be calculated by measuring the relative absorption of the R, G, and B channels on the slides

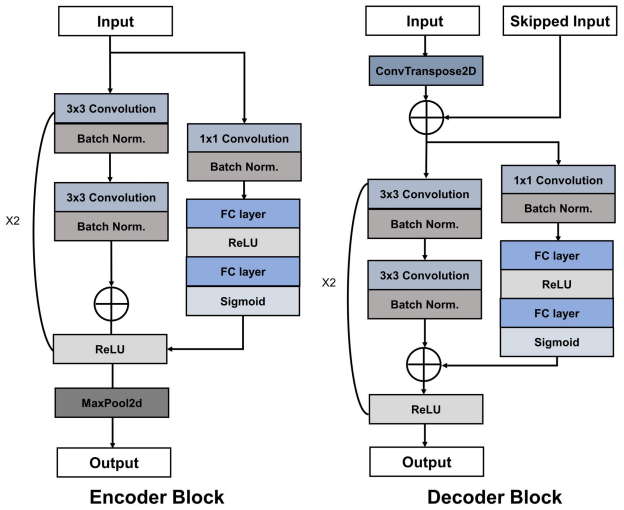

FIGURE 3. Illustration of a detailed encoder & decoder block.

stained with a single stain. An example of the matrix M is shown in Eq (2).

$$D = M^{-1} = \begin{matrix} R & G & B \end{matrix} \begin{bmatrix} 0.65 & 0.70 & 0.29 \\ 0.07 & 0.99 & 0.11 \\ 0.27 & 0.57 & 0.79 \end{bmatrix}^{-1} = \begin{bmatrix} 1.88 & -0.07 & -0.60 \\ -1.02 & 1.13 & -0.48 \\ -0.55 & -0.13 & 1.57 \end{bmatrix} \begin{matrix} Hematoxylin \\ Eosin \\ Residual \end{matrix} \quad (2)$$

Finally, the output staining intensity S can be calculated as

$$S = D * OD, \quad (3)$$

where $S = [h, e, r]^T$ represents the separated intensities of stains, and h, e, and r represent the intensities of hematoxylin, eosin and the residual of stain separation, respectively.

IV. PROPOSED METHOD

Our proposed HDA-Net is a dual-encoder structure consisting of HER stream and RGB stream. In the HER stream, we extracted discriminative and color-invariant features of the HER components and passed them to the RGB stream using SAM. In RGB stream, these features are utilized as guided features to enhance the representation of RGB images using HDA and SAM. At the end of the decoder, multi-task learning is performed for better supervision.

A. H&E AND RGB DUAL ATTENTION

We denote two feature maps $f_\alpha, f_\beta \in R^{H\times W\times C}$, assuming that f_α is the RGB stream feature and f_β is the HER stream H is the height of the feature map, W is the width, and C is the number of channels. To initially integrate these two features, we chose an element-wise summation. $Z = f_\alpha + f_\beta$, where $Z \in R^{H\times W\times C}$ denotes the fused feature. Then, we apply global average pooling to Z to exploit global contextual information with embedded channel-wise statistics. The c-th component is calculated as follows:

$$M_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} Z_c(i, j), \quad (4)$$

where $M_c \in \mathbb{R}^C$ is an attention map that encapsulates the global context information collected via the global average pooling across the spatial dimensions. This attention map emphasizes the correlation between the two features, guiding the RGB features with HER features. We generate the logits L_c by applying a linear transformation to M_c and obtaining channel-wise attention weights A_c by using softmax activation.

$$L_c = W_c M_c + b_c, \quad (5)$$

$$A_c = \frac{\exp(L_c)}{\sum_{c=1}^{N} \exp(L_c)}, \quad (6)$$

where N denotes the total number of channels. R and H are produced using a shortcut connection and elementwise multiplication: $R = A_c \otimes f_\alpha$ and $H = A_c \otimes f_\beta$. Finally, we got the output f_γ is obtained by an element-wise summation: $f_\gamma = R \oplus H$. By combining the two feature maps, the model is trained to consider information from both feature maps, which include cellular structures represented in RGB and detailed cellular information represented in HER.

$$f_\gamma = \mathbf{HDA}\left(f_\alpha, f_\beta\right), \quad (7)$$

where f_γ is extracted from RGB features f_α guided by the HER feature f_β and propagated to the next RGB stream encoder block.

Algorithm 1 Procedure of HDA

Input: RGB feature f_α, HER feature f_β.
Output: fused feature f_γ

1: $Z = f_\alpha \oplus f_\beta$
2: $M_c = \frac{1}{H\times W} \sum_{i=1}^{H} \sum_{j=1}^{W} Z_c(i, j)$
3: $L_c = W_c M_c + b_c$
4: $A_c = \frac{\exp(L_c)}{\sum_{c=1}^{N} \exp(L_c)}$
5: $f_\gamma \leftarrow (A_c \otimes f_\alpha) \oplus \left(A_c \otimes f_\beta\right)$

B. SINGLE-SOURCE ATTENTION MODULE

We propose a single-source attention module (SAM) that is inserted into the encoder and decoder blocks. This allows for the extraction of robust features from both single-source features f_β in the HER stream as well as from the fused features f_γ obtained from the HDA module in the RGB stream. Our encoder blocks use two 3×3 convolution layers and identity mapping to connect the input and output of the convolution layer. Each convolution layer is followed by a Batch Normalization (BN). Additionally, we added a single-source attention module (SAM) to the

residual network. This module operates as a content-aware mechanism that re-weighs each channel to form robust representations of f_γ and f_β. SAM is composed of two stages. First, the feature maps were extracted by 1×1 convolution. These feature maps were then squeezed using global average pooling (GAP) to obtain a global understanding of each channel.

$$f_y = \mathbf{GAP}\left(\mathbf{BN}\left(\mathbf{Conv2D_{1\times1}}\left(f_x\right)\right)\right), \tag{8}$$

where f_y is a feature vector of length n; Second, f_y is fed through two linear layers, where the number of features is first reduced to $\frac{n}{\rho}$ and then expanded to the original size n, where ρ is a dimension-reduction hyperparameter. The n-sized vector represents the weight of the original feature maps used to scale each channel.

$$f_{out} = \sigma\left(W_2\left(\mu\left(W_1\left(f_y\right)\right)\right)\right), \tag{9}$$

where W_1 is a linear layer that reduces the channel and W_2 is a linear layer that expands the channel. In addition, σ is the sigmoid function and μ is the ReLU activation function [35]. In our method, the outputs derived from the two convolution layers are processed through the SAM outputs f_{out} followed by identity mapping. After undergoing this process twice, the Encoder Block is succeeded by MaxPool2D, and an output feature map is generated.

$$f_\beta^{k+1} = E_{\theta_k}\left(f_\beta^k\right), \tag{10}$$

$$f_\alpha^{k+1} = E_{\psi_k}\left(f_\gamma^k\right), \tag{11}$$

where E_{θ_k} is the encoder for the HER stream and E_{ψ_k} is the encoder for RGB stream. where k denotes the number of layers. Our Decoder Block is similar to the encoder block; however, the decoder block is initiated with a 4×4 transpose convolution, effectively doubling the spatial dimension of the incoming feature map and not utilizing MaxPool2D.

$$f_{dec}^5 = D_{\theta_5}\left(f_\gamma^5, f_\alpha^5\right) \tag{12}$$

$$f_{dec}^k = D_{\theta_k}\left(f_{dec}^{k+1}, f_\alpha^k\right) \tag{13}$$

As shown in Eq(12) and Eq(13), the first decoder D_{θ_5} extracts output f_{dec}^5 using f_γ^5 and f_α^5. Then, decoder D_{θ_k} extracts the output features f_{dec}^k using the skipped connected encoder features f_α^k and f_{dec}^{k+1} of the previous decoder.

C. LOSS FUNCTION

We incorporate multi-task learning in our proposed model to utilize extended supervision for the localization of overlapped nuclei. In the last decoder, the region and centroid predictions are derived from two individual 1×1 convolutional layers. And the outputs are defined as $f_{nuc} \in R^{H\times W\times 1}$ and $f_{cen} \in R^{H\times W\times 1}$.

$$f_{nuc} = \mathbf{Conv2D_{1\times1}}\left(f_{dec}^1\right) \tag{14}$$

$$f_{cen} = \mathbf{Conv2D_{1\times1}}\left(f_{dec}^1\right) \tag{15}$$

Algorithm 2 HDA-Net Forward Process

Input: RGB HSIs image X^{RGB}.
Output: Region prediction f_{nuc}, Centroid prediction f_{cen}

```
Apply Color Deconvolution X^{HER} ← D * OD(X^{RGB})
for i = 1 to 5 do
    if i = 1 then
        f_β^0 = X^{HER}
    end if
    f_β^i ← E_{θ_i}(f_β^{i-1})
end for                                   ▷ HER stream
for j = 1 to 5 do
    if j = 1 then
        f_α^0 = X^{RGB}
        f_α^1 ← E_{ψ_1}(f_α^0)
        f_γ^1 ← HDA(f_α^1, f_β^1)
    end if
    f_α^j ← E_{ψ_j}(f_γ^{j-1})
    f_γ^j ← HDA(f_α^j, f_β^j)
end for                                   ▷ RGB stream
for k = 1 to 5 do
    if k = 1 then
        f_dec^6 = f_γ^5
    end if
    f_dec^{6-k} ← D_{θ_k}(f_α^{6-k}, f_dec^{7-k})
end for
f_nuc = Conv2D_1x1(f_dec^1)
f_cen = Conv2D_1x1(f_dec^1)
```

Our model is trained for two tasks: Nuclei Region Map and Centroid Distance Map; therefore, we need a loss function to minimize the errors for both tasks. The losses for these two tasks are minimized using the loss function $\boldsymbol{L_{total}}$ which is expressed as follows:

$$\boldsymbol{L_{total}} = \lambda_1 \boldsymbol{L_{Dice}} + \lambda_2 \boldsymbol{L_{MSE}} \tag{16}$$

In our experiment, we set $\lambda_1 = \lambda_2 = 1$ such that the segmentation ability of our network is not significantly affected by λ_1 and λ_2. $\boldsymbol{L_{Dice}}$ is the Dice loss [36]. Dice loss reduces background imbalance by optimizing the network based on the Dice overlap coefficient between the predicted segmentation result and correct answer annotation. Dice loss is used to calculate the difference between the actual Nuclei Region Map compared to region prediction $\boldsymbol{f_{nuc}} \in \mathbb{R}^{H\times W\times 1}$ for region loss. $\boldsymbol{L_{MSE}}$ is Mean Squared Error (MSE) loss. MSE loss was used to calculate the difference between the actual Centroid Distance Map and the centroid prediction $\boldsymbol{f_{cen}} \in \mathbb{R}^{H\times W\times 1}$ using the segmentation model.

V. EXPERIMENT

A. DATASET

We evaluated our model on three nuclei segmentation datasets: Colorectal Nuclear Segmentation and Phenotypes (CoNSeP) [15], PanNuke [37], and Kumar [38].

TABLE 1. Comparative experiments on the CoNSeP, PanNuke, and Kumar. Red and blue are the first and second-best performance results.

Method	CoNSeP				PanNuke				Kumar			
	AJI	Dice	Recall	Precision	AJI	Dice	Recall	Precision	AJI	Dice	Recall	Precision
SegNet	46.80(0.47)	63.66(0.44)	65.57(0.46)	62.29(1.16)	56.74(0.11)	72.35(0.09)	72.01(0.39)	72.91(0.30)	47.84(0.34)	64.66(0.31)	66.28(0.91)	63.69(1.45)
DeeplabV3+	61.15(0.39)	75.84(0.30)	80.08(0.48)	73.05(0.40)	65.54(0.37)	79.04(0.29)	79.33(0.56)	79.11(1.09)	62.18(0.38)	76.50(0.29)	75.11(0.44)	78.66(0.79)
Unet	64.97(1.16)	79.14(0.92)	81.27(0.12)	76.19(1.52)	66.97(1.03)	79.88(0.77)	80.32(0.95)	80.01(2.10)	62.13(0.50)	76.59(0.38)	76.69(0.53)	76.89(1.20)
Att-Unet	65.37(2.47)	79.98(1.83)	81.49(1.31)	76.97(2.21)	69.01(0.26)	81.60(0.18)	81.74(0.41)	81.66(0.75)	65.73(0.15)	79.18(0.11)	80.84(0.47)	77.87(0.57)
Unet++	66.56(0.13)	79.89(0.09)	82.64(0.27)	77.61(0.37)	68.13(0.61)	80.96(0.44)	82.63(0.39)	79.58(1.20)	64.95(0.80)	78.57(0.63)	**81.64(1.07)**	76.12(1.98)
ResUnet	67.24(0.22)	80.37(0.15)	80.27(0.57)	**80.78(0.88)**	69.26(0.32)	81.78(0.22)	81.76(0.70)	**82.04(1.16)**	66.04(0.14)	79.41(0.10)	79.64(0.30)	**79.54(0.29)**
USEnet	66.70(0.34)	79.98(0.25)	80.81(0.92)	79.59(1.32)	**69.46(0.18)**	**81.91(0.12)**	82.20(0.46)	81.83(0.22)	65.05(0.29)	78.68(0.22)	80.69(0.40)	77.18(0.52)
HoVerNet	**68.19(0.16)**	**81.13(0.11)**	**83.75(0.29)**	78.85(0.56)	69.02(0.16)	81.61(0.11)	**82.99(0.31)**	80.46(0.44)	**66.22(0.39)**	**79.52(0.29)**	81.57(1.27)	77.94(1.64)
CDNet	66.91(0.34)	80.13(0.24)	82.65(0.39)	78.05(0.80)	68.52(0.35)	81.25(0.25)	**83.28(0.34)**	79.50(0.77)	66.11(0.20)	79.44(0.15)	81.32(0.98)	77.97(0.70)
NCNet	66.92(0.48)	80.15(0.35)	**83.22(0.39)**	77.61(0.88)	69.39(0.09)	81.87(0.06)	82.53(0.53)	81.38(0.63)	65.11(0.17)	78.67(0.12)	**82.31(0.87)**	75.73(0.51)
Ours	**69.39(0.14)**	**81.89(0.09)**	82.84(0.65)	**81.31(0.44)**	**70.16(0.13)**	**82.41(0.09)**	81.66(0.58)	**83.37(0.42)**	**66.43(0.37)**	**79.68(0.27)**	80.80(0.79)	**78.90(0.89)**

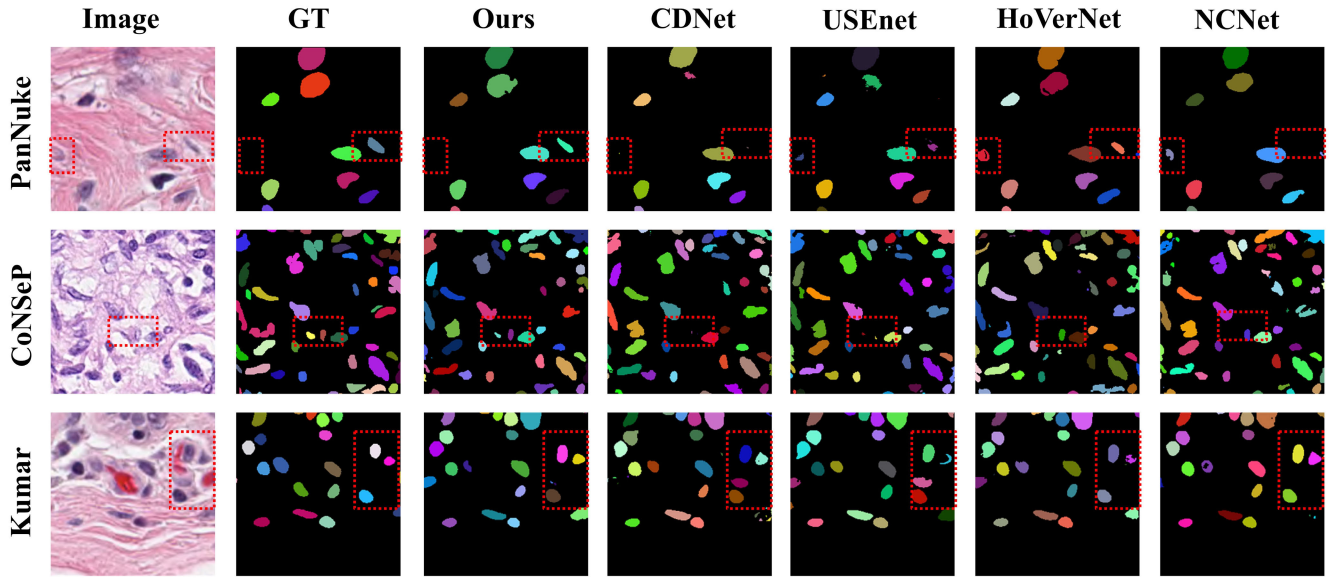

FIGURE 4. Visualization of segmentation on PannNuke, CoNSeP, and Kumar datasets. From left to right: Orginal image; Ground Truth; Ours; CDNet [9]; USEnet [24]; HoVerNet [15]; NCNet [16]; Different colors indicate different instances in the images. The red dotted rectangles are drawn for a clear comparison.

CoNSeP involves 41 H&E stained images, each of which is 1000 × 1000 pixels at 40x magnification. Images were extracted from 16 HSIs of different colorectal adenocarcinoma (CRA) HSIs, each belonging to an individual patient group. We used the same image split as in existing methods [15], where the training set contained 27 images and the test set contained 14 images.

PanNuke dataset contains 7,904 images with 256 × 256 patches from a total of 19 different tissue types and 189,744 labeled nuclei. We used the same image split as in the existing methods [16], where the total images were divided into three folds, each having 2,657, 2,525, and 2,723.

Kumar dataset contains 30 H&E stained images from seven organs (6 breasts, 6 liver, 6 kidneys, 6 prostate, 2 bladder, 2 colon, and 2 stomach). Each image had 1000 × 1000 pixels at 40x magnification. We used the same image split as in existing methods [38], where the training set contained 16 images and the test set contained 14 images. We conducted training and evaluation using the training and test sets individually for each of the three datasets.

B. EVALUATION METRICS

To compare our model with others, we utilized the Aggregated Jaccard Index (AJI) and Dice coefficient (Dice) as evaluation metrics [9]. In addition, we used precision and recall to evaluate the three datasets. The AJI is a crucial evaluation metric that quantifies the overlap and accuracy between predicted segmentation and the ground truth for nuclei instance segmentation [38]. Dice measures the overlap between the ground truth and segmentation results. Precision represents the ratio of segmented nuclei in the labeled images, which is also called the positive predictive value. Recall represents the percentage of the total number of nuclear pixels in a label image correctly segmented using the proposed method, which is also called the sensitivity.

C. IMPLEMENTATION AND TRAINING DETAILS

We used PyTorch to implement the HDA-Net on an NVIDIA RTX 3090 with CUDA 11.2. Adam [39] was used as an optimizer. The initial learning rate was set to 0.0005, the batch size was four, and the training epoch was set to

TABLE 2. Cross-domain evaluation experiments on the CoNSeP, PanNuke, and Kumar. → denotes that the left is the train dataset and the right is the test dataset. **Red** and **blue** are the first and second-best performance results.

Tasks	Method	AJI	Dice	Recall	Precision	Tasks	Method	AJI	Dice	Recall	Precision
Kumar→PanNuke	USEnet	53.13	69.18	**89.50**	56.68	Kumar→CoNSeP	USEnet	58.08	73.41	79.35	68.73
Kumar→PanNuke	HoVerNet	51.28	67.45	**90.02**	54.37	Kumar→CoNSeP	HoVerNet	58.13	73.41	**85.51**	64.55
Kumar→PanNuke	CDNet	**56.83**	**72.26**	86.31	**62.47**	Kumar→CoNSeP	CDNet	58.56	73.73	**83.84**	66.05
Kumar→PanNuke	NCNet	52.55	68.63	88.67	56.37	Kumar→CoNSeP	NCNet	**60.23**	**75.08**	82.95	**68.84**
Kumar→PanNuke	Ours	**57.96**	**73.25**	85.42	**64.32**	Kumar→CoNSeP	Ours	**62.56**	**76.89**	80.50	**73.85**

Tasks	Method	AJI	Dice	Recall	Precision	Tasks	Method	AJI	Dice	Recall	Precision
CoNSeP→PanNuke	USEnet	**58.49**	**73.59**	83.92	**65.91**	CoNSeP→Kumar	USEnet	56.44	72.04	**74.07**	70.63
CoNSeP→PanNuke	HoVerNet	57.46	72.70	**84.96**	64.08	CoNSeP→Kumar	HoVerNet	**59.71**	**74.68**	71.64	**78.40**
CoNSeP→PanNuke	CDNet	57.37	72.80	84.22	64.31	CoNSeP→Kumar	CDNet	54.29	70.20	**76.10**	65.57
CoNSeP→PanNuke	NCNet	55.73	71.42	**85.05**	61.81	CoNSeP→Kumar	NCNet	56.76	72.30	72.53	72.82
CoNSeP→PanNuke	Ours	**59.75**	**74.72**	82.59	**68.48**	CoNSeP→Kumar	Ours	**59.25**	**74.31**	69.71	**79.93**

Tasks	Method	AJI	Dice	Recall	Precision	Tasks	Method	AJI	Dice	Recall	Precision
PanNuke→CoNSeP	USEnet	41.55	58.54	**79.20**	46.92	PanNuke→Kumar	USEnet	43.70	60.56	71.07	53.94
PanNuke→CoNSeP	HoVerNet	46.05	62.88	77.34	53.46	PanNuke→Kumar	HoVerNet	43.26	60.09	**68.78**	54.73
PanNuke→CoNSeP	CDNet	**46.90**	**63.74**	77.50	**54.58**	PanNuke→Kumar	CDNet	**47.34**	**64.09**	68.57	**61.32**
PanNuke→CoNSeP	NCNet	39.48	56.43	**78.65**	44.44	PanNuke→Kumar	NCNet	39.70	56.47	**69.71**	48.82
PanNuke→CoNSeP	Ours	**47.14**	**63.91**	74.69	**56.64**	PanNuke→Kumar	Ours	**49.56**	**66.07**	68.44	**64.86**

TABLE 3. Ablation study on the effectiveness of utilizing her components: Hematoxylin, Eosin, and Residual Analysis.

Hematoxylin	Eosin	Residual	AJI	Dice
			68.41	81.19
✓			68.73	81.43
	✓		68.45	81.23
		✓	68.21	81.06
✓	✓		68.94	81.57
✓	✓	✓	69.60	82.03

200. For training, we randomly augmented our data using standard data augmentation techniques, such as horizontal flip, rotation, and cropping. The total number of parameters in our model is 68.35M, and the inference speed is 42.89ms. For ConSep and Kumar, each image in the datasets was divided from the original size of 1000 × 1000 to 250 × 250, resulting in 16 patches per image. The patches were then resized to 256 × 256. For the ConSep dataset, we used 432 images for training and 224 images for testing. For the Kumar dataset, we used 256 training and 224 test images in our experiments. By conducting three trials, we aimed to enhance the robustness and reliability of our performance measurements for models trained on ConSep and Kumar datasets. For PanNuke, we trained the model three times, each with approximately 2500 images. In the first fold, we consider fold1 and fold2 as the training set and fold3 as the test set. In the second fold, we consider fold1 and fold3 as the training set and fold2 as the test set. In the third fold, we consider fold2 and fold3 as the training set, and fold1 as the test set.

D. COMPARATIVE RESULTS

We compared our proposed model to recent segmentation approaches used in computer vision (SegNet [6], DeeplabV3+ [8]), medical imaging (Unet [13], Att-Unet [29], Unet++ [40], ResUnet [30]) and methods for nuclei instance segmentation (USEnet [24], HoVerNet [15], CDNet [9], NCNet [16]). We conducted experiments on three public datasets using the above models and the proposed model. From Table 1, we can see that our proposed model outperforms all state-of-the-art nuclei instance segmentation methods listed in AJI and Dice. Specifically, the proposed HDA-Net improves AJI by 1.2% and Dice by 0.76% compared with the second-best performing model on the CoNSeP dataset. For PanNuke, HDA-Net achieved a 0.7% improvement in AJI and a 0.5% improvement in Dice compared with the second-best model. Finally, in the Kumar dataset, HDA-Net improves AJI by 0.21% and Dice by 0.16% compared with the second-best model. As expected, our model outperformed the compared models by utilizing the decomposed HER to localize each component of the cell. Furthermore, we propose HDA and SAM to learn the hidden correlations between RGB and HER components. However, our method was inferior to the other models in terms of recall. After analyzing Figure 5, it becomes apparent that our model displays less noise in comparison to others. The occurrence of noisy predictions has the potential to increase the prediction of positive cases, thereby possibly enhancing the recall rate. While recall value is reduced, our model has shown higher precision. This precision is vital for accurately predicting nuclei shapes to capture morphological changes that hold clinical relevance [41]. The increased precision in our model suggests that despite the challenges posed by noisy data on recall, our model effectively categorizes positive instances, reducing the likelihood of false positives and ensuring the trustworthiness of predictions in clinical settings.

E. QUALITATIVE RESULTS

In this experiment, we conducted a qualitative comparison using the PanNuke, CoNSeP, and Kumar datasets. In Figure 4,

TABLE 4. Ablation study results on the model components: Encoder, Decoder, HDA module, and multi-task experimentation.

Encoder	Decoder	HDA	multi-task	AJI	Dice
				64.58	79.04
	✓		✓	66.92	80.12
✓			✓	67.88	80.82
✓	✓		✓	68.41	81.19
✓	✓	✓		69.21	81.76
✓	✓	✓	✓	69.60	82.03

TABLE 5. Performance comparison of our proposed HDA method with other attention mechanism methods.

Method	AJI	Dice
CBAM [28]	68.41	81.21
SENet [26]	68.89	81.53
GCNet [42]	66.06	79.41
ECA-Net [43]	69.29	81.82
AFF [44]	68.38	81.18
Ours	69.60	82.03

we compare the results obtained using the methods used for nuclei instance segmentation (CDNet, USEnet, HoVerNet, and NCNet). As shown in the first row of Figure 4, our proposed model, in comparison to other models, accurately predicted the nuclei and did not make predictions for ambiguous nuclei shapes that resembled nuclei. Similarly, as shown in the second row, our model effectively identified a small nucleus that the other models failed to predict. In the third row, our model does not make mispredictions from the others. Our HDA-Net can effectively detect even small nuclei and ambiguous nuclei shapes, which other models struggle to predict accurately. This is largely due to the fact that other models focus mainly on separating the nuclei using only RGB images. In contrast, our model is enhanced by the use of HER components, which provide detailed information about the nuclei for the prediction process. As demonstrated in Table 3, the gradual integration of HER components has led to improved segmentation results in comparison to using only RGB images.

VI. ABLATION STUDY

A. ALGORITHM ABLATION

Here, we discuss the impact of the components in the framework: the Encoder, Decoder, HDA, and multi-task. Specifically, we began with experimental verification of the Encoder and Decoder blocks by removing certain components and replacing them with a convolution layer similar to Unet. As shown in Table 4, when encoder blocks are not used, a performance drop of 1.49% is observed in AJI when encoder blocks are not utilized, and a 0.53% drop is observed in performance when decoder blocks are not in use. Subsequently, we discuss our proposed HDA module. When the HDA, the AJI improved by 1.59% AJI.

TABLE 6. Comparison of performance metrics between Ours and HER-only, a method that calculates weights using only HER and multiplies it with RGB.

	AJI	Dice	Recall	Precision
Ours	69.39	81.89	82.84	81.31
HER-only	68.53	81.38	81.79	80.98

Finally, we justified the use of multi-task learning with a centroid distance map in our proposed model by removing the convolution layer that predicted the distance map and observed a performance drop of 0.39% AJI. The improvement in multi-task performance seems small, but the distance map is an important element for localizing the positions between overlapped nuclei. Overall, our model achieved high performance by utilizing our proposed components and generated better segmentation results in the nuclei instance segmentation datasets.

B. CROSS-DOMAIN EVALUATION

To justify the generalizability of our model, we conducted a cross-domain evaluation. More specifically, we test the performance of Ours and other nuclei instance segmentation methods on the Kumar→PanNuke, Kumar→CoNSeP, CoNSeP→PanNuke, CoNSeP→Kumar, PanNuke→CoNSeP, and PanNuke→Kumar. Each dataset on the left was used as a training set, whereas the dataset on the right was used for testing. The experimental results of the cross-dataset evaluation are listed in Table 2. For the five tasks, our method consistently outperformed the others except for the task CoNSeP→Kumar. These experimental results demonstrate that HDA-Net achieves high generalizability by utilizing the decomposed HER, which is color invariant.

C. EFFECTIVENESS OF H&E AND RGB DUAL ATTENTION

In this section, we compare the proposed H&E and RGB dual attention(HDA) mechanism with other attention mechanisms, including CBAM [28], SENet [26], GCNet [42], and ECA-Net [43]. For each of these attention mechanisms, we performed an element-wise sum operation on the RGB and HER tensors and subsequently applied the respective CBAM, SE-block, Global Context, and ECA modules. Additionally, for the AAF [44], we applied the AAF module to the RGB and HER tensors as inputs. As shown in Table 5, the proposed method achieved the best performance for the various attention mechanisms.

D. EFFECTIVENESS OF RGB AND HER COMPONENT INTEGRATION ON MODEL PERFORMANCE

In this section, we conduct an ablation study to evaluate the performance of our proposed HDA-Net method and the HER-only method, as displayed in Table 6. Our approach, HDA-Net, which computes weights following the integration

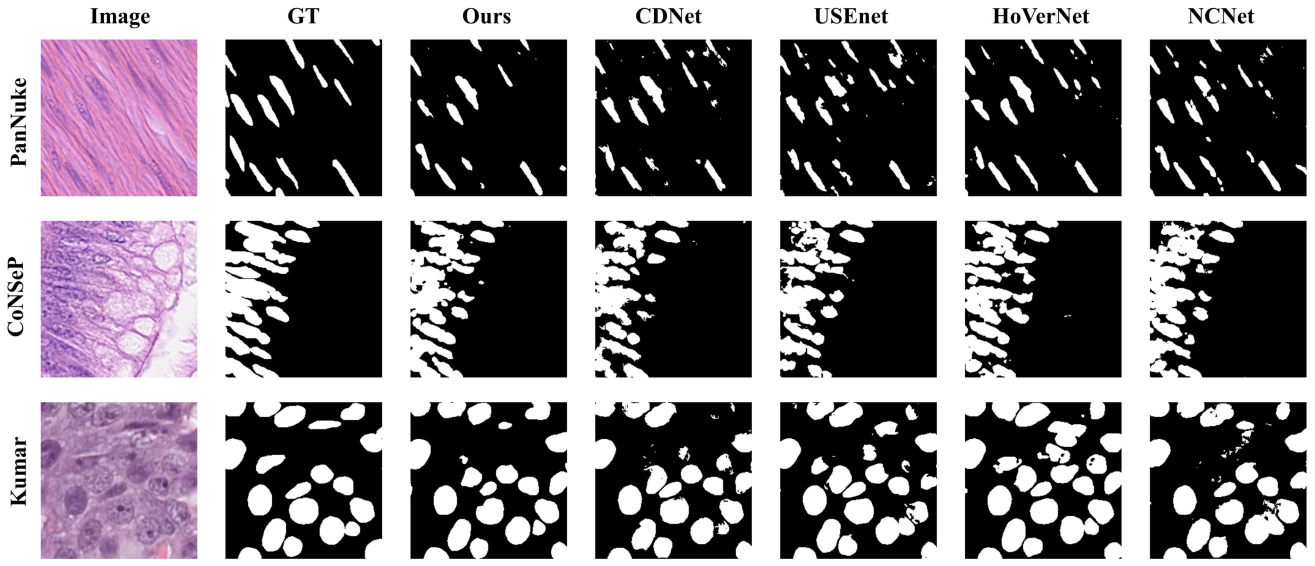

FIGURE 5. Sample cropped regions were extracted from each of the three instance segmentation datasets used in our experiments. We present our prediction results with Ground Truth for images. These results are failure cases, occurring from the prediction of images with blurred boundaries of the cell nucleus.

of RGB and HER, is compared with the HER-only method that solely calculates weights using HER prior to their multiplication with RGB. Evidently, HDA-Net demonstrates an improvement of 0.86% in AJI and 0.51% in Dice scores over the HER-only method. This outcome underscores the benefits of utilizing both RGB and HER rather than relying solely on HER.

VII. DISCUSSION

We propose a nuclei instance segmentation method to address challenges in HSIs, such as color inconsistencies and overlapping nuclei boundaries. Our HDA-Net incorporates the SAM into the encoder and decoder block to enhance feature extraction. Illustrated in Figure 4, HDA-Net outperforms other models by avoiding incorrect shape predictions resembling nuclei and effectively detecting nuclei missed by alternative methods. This indicates that HDA extracts features from RGB images guided by HER components, providing detailed information about cells (e.g., nuclei and cytoplasm) via an attention mechanism. HDA-Net primarily focuses on the HER components and centroid for localization, leading to suboptimal performance when applied to complex textures and ambiguous boundaries. Nevertheless, our model manages to generate outputs with reduced noise compared to other models. In light of these observations, our future plan involves developing a more robust model for nuclei texture by incorporating frequency information, which contains substantial texture details from the image, in addition to the HER component. Furthermore, HDA-Net demonstrates comparable performance in cross-domain evaluation owing to the color-invariant characteristics of the HER components. However, the proposed method does not fully utilize decomposed HER because of the indirect nature of the cross-attention mechanism for information injection.

VIII. CONCLUSION

HER components were extracted from H&E-stained RGB images using color deconvolution. To leverage these components, we proposed a novel architecture for nuclei instance segmentation. We evaluated the method on three nuclei instance segmentation datasets. The results reveal the proposed method effectively aggregates correlations between the two images, achieving superior segmentation than those of previous studies. Generalizability was demonstrated through cross-domain evaluation. However, the method used in this study has certain limitations. Owing to the attention mechanism's complexity, this model is computationally expensive and time-consuming. In our future work, we aim to focus on reducing the time and computing limitations and developing improved nuclei instance segmentation models. Additionally, we plan to develop a model that is robust to noise by also taking into account the texture of nuclei and utilizing frequency information.

REFERENCES

[1] L. Pantanowitz, "Digital images and the future of digital pathology," *J. Pathol. Informat.*, vol. 1, no. 1, p. 15, Jan. 2010.
[2] M. K. K. Niazi, A. V. Parwani, and M. N. Gurcan, "Digital pathology and artificial intelligence," *Lancet Oncol.*, vol. 20, no. 5, pp. e253–e261, 2019.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[6] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[7] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[8] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder–decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 801–818.

[9] 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.

[10] J. Yi, P. Wu, Q. Huang, H. Qu, B. Liu, D. J. Hoeppner, and D. N. Metaxas, "Multi-scale cell instance segmentation with keypoint graph based bounding boxes," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, Shenzhen, China. Cham, Switzerland: Springer, Oct. 2019, pp. 369–377.

[11] L. Liu, C. Hong, A. I. Aviles-Rivero, and C.-B. Schönlieb, "Simultaneous semantic and instance segmentation for colon nuclei identification and counting," in *Proc. Annu. Conf. Med. Image Understand. Anal.* New York, NY, USA: Springer, 2022, pp. 130–138.

[12] N. A. Koohbanani, M. Jahanifar, A. Gooya, and N. Rajpoot, "Nuclear instance segmentation using a proposal-free spatially aware deep learning framework," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Shenzhen, China. New York, NY, USA: Springer, Oct. 2019, pp. 622–630.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[14] 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.

[15] 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.

[16] S. N. Rashid and M. M. Fraz, "Nuclei probability and centroid map network for nuclei instance segmentation in histology images," *Neural Comput. Appl.*, vol. 35, no. 21, pp. 15447–15460, Jul. 2023.

[17] L. Yang, R. P. Ghosh, J. M. Franklin, S. Chen, C. You, R. R. Narayan, M. L. Melcher, and J. T. Liphardt, "NuSeT: A deep learning tool for reliably separating and analyzing crowded cells," *PLOS Comput. Biol.*, vol. 16, no. 9, Sep. 2020, Art. no. e1008193.

[18] D. Bannon, E. Moen, E. Borba, A. Ho, I. Camplisson, B. Chang, E. Osterman, W. Graf, and D. Van Valen, "DeepCell 2.0: Automated cloud deployment of deep learning models for large-scale cellular image analysis," *BioRxiv*, vol. 12, Dec. 2018, Art. no. 505032.

[19] H. Irshad, L. Montaser-Kouhsari, G. Waltz, O. Bucur, J. A. Nowak, F. Dong, N. W. Knoblauch, and A. H. Beck, "Crowdsourcing image annotation for nucleus detection and segmentation in computational pathology: Evaluating experts, automated methods, and the crowd," in *Proc. Pacific Symp. Biocomput. Co-Chairs*. Singapore: World Scientific, Nov. 2014, pp. 294–305.

[20] S. Vijh, M. Saraswat, and S. Kumar, "A new complete color normalization method for H&E stained histopatholgical images," *Appl. Intell.*, vol. 51, no. 11, pp. 7735–7748, 2021.

[21] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quantum Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.

[22] C. Lin, Y. Chan, and C. Chen, "Detection and segmentation of cervical cell cytoplast and nucleus," *Int. J. Imag. Syst. Technol.*, vol. 19, no. 3, pp. 260–270, Sep. 2009.

[23] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 11, pp. 2405–2414, Nov. 2006.

[24] G. Rahmon, I. E. Toubal, and K. Palaniappan, "Extending U-Net network for improved nuclei instance segmentation accuracy in histopathology images," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2021, pp. 1–7.

[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[26] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6450–6458.

[27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[28] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.

[29] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[30] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *Proc. 9th Int. Conf. Inf. Technol. Med. Educ. (ITME)*, Oct. 2018, pp. 327–331.

[31] Y. Wang, Z. Deng, X. Hu, L. Zhu, X. Yang, X. Xu, P.-A. Heng, and D. Ni, "Deep attentional features for prostate segmentation in ultrasound," in *Proc. 21st Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Granada, Spain. New York, NY, USA: Springer, Sep. 2018, pp. 523–530.

[32] G. Hou, J. Qin, X. Xiang, Y. Tan, and N. N. Xiong, "AF-Net: A medical image segmentation network based on attention mechanism and feature fusion," *Comput., Mater. Continua*, vol. 69, no. 2, pp. 1877–1891, 2021.

[33] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-Net: Automatic COVID-19 lung infection segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[34] K. Prasad and G. K. Prabhu, "Image analysis tools for evaluation of microscopic views of immunohistochemically stained specimen in medical research—A review," *J. Med. Syst.*, vol. 36, no. 4, pp. 2621–2631, Aug. 2012.

[35] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[36] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Proc. Int. Workshop Deep Learn. Med. Image Anal., 7th Int. Workshop Multimodal Learn. Clin. Decis. Support*, Québec City, QC, Canada. Cham, Switzerland: Springer, Sep. 2017, pp. 240–248.

[37] J. Gamper, N. A. Koohbanani, K. Benet, A. Khuram, and N. Rajpoot, "PanNuke: An open pan-cancer histology dataset for nuclei instance segmentation and classification," in *Proc. 15th Eur. Congr. Digit. Pathol.*, Warwick, U.K. New York, NY, USA: Springer, 2019, pp. 11–19.

[38] 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.

[39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[40] Z. Zhou, M. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. 4th Int. Workshop Deep Learn. Med. Image Anal.*, vol. 11045, Granada, Spain. New York, NY, USA: Springer, Sep. 2018, pp. 3–11.

[41] P. Dey, "Cancer nucleus: Morphology and beyond," *Diagnostic Cytopathol.*, vol. 38, no. 5, pp. 382–390, May 2010.

[42] Y. Cao, J. Xu, S. Lin, F. Wei, and H. Hu, "GCNet: Non-local networks meet squeeze-excitation networks and beyond," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1971–1980.

[43] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[44] Y. Dai, F. Gieseke, S. Oehmcke, Y. Wu, and K. Barnard, "Attentional feature fusion," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 3559–3568.

YU-HAN IM received the bachelor's degree in computer science from Inha University, South Korea, in 2022, where he is currently pursuing the M.S. degree in computer engineering. His research interests include artificial intelligence and computer vision.

SEO-HYEONG PARK received the bachelor's degree in computer science from Inha University, South Korea, in 2021, where he is currently pursuing the M.S. degree in software (computer engineering). His research interests include artificial intelligence and computer vision.

SANG-CHUL LEE (Member, IEEE) received the B.E. degree in computer science and engineering from Inha University, South Korea, in 1998, and the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, in 2006. He was a Research Programmer with the National Center for Supercomputing Applications, from 2006 to 2008. He was a Guest Researcher with the Information Technology Laboratory, National Institute of Standards and Technology (NIST), in 2012. He is currently a Professor with the Department of Computer Engineering, Inha University. His research interests include computer vision, robotics, and medical imaging.