Development of Computer Aided Multi-organ Cancer Diagnostic System from Histopathological Images

Harsha Kumari
National Institute of Technology, Rourkela

***Abstract*—This paper describes a modified U-Net architecture with Squeeze-and-Excitation (SE) blocks and Atrous Spatial Pyramid Pooling (ASPP) to achieve better accuracy in nucleus segmentation of H&E histopathology images. SE blocks increase segmentation precision by enhancing channel features, and multi-scale context captured by ASPP improves separation of touching nuclei. The model is tested on the MoNuSeg dataset, where it outperformed other models in boundary delineation and showed high robustness across several tissue types. The model solves the problem of class imbalance for reliable segmentation by incorporating Jaccard distance and Dice loss. This novel method improves the automated analysis of histopathology images, thus decreasing the amount of manual work and increasing diagnostic reliability.**

I. Introduction

For the diagnosis of cancer, histopathological image analysis is crucial, and nucleus segmentation helps to discover anomalies and comprehend tissue structure. However, automated approaches are required since manual segmentation is time-consuming, prone to irregularities, and requires specialist expertise. To promote the creation of models capable of segmenting nuclei across many tissue types, the Multi-Organ Nucleus Segmentation Challenge (MoNuSeg) was established. This effort aims to handle overlapping nuclei efficiently and increase segmentation accuracy.

Numerous investigations have looked into computational methods for segmenting nuclei. To enhance segmentation, the MoNuSeg 2018 Challenge used fully convolutional networks (FCNs) like U-Net and Mask-RCNN in conjunction with color normalization. In order to capture multi-scale information, other research has integrated attention-based architectures by fusing feature enhancement techniques with convolutional algorithms. Hu et al.'s Squeeze-and-Excitation (SE) network enhances feature selection by highlighting crucial characteristics while squelching less pertinent information. Although segmentation has improved with these techniques, class imbalance and complicated nucleus boundaries remain problems.

Models like SAR-U-Net, which combines SE blocks and Atrous Spatial Pyramid Pooling (ASPP) to enhance feature extraction and multi-scale context awareness, have been proposed recently. This has enhanced segmentation performance. Furthermore, research has demonstrated that adding ASPP modules and SE blocks can greatly increase segmentation accuracy in medical imaging applications.

For accurate nucleus segmentation, this work proposes an enhanced U-Net model that incorporates Atrous Spatial Pyramid Pooling (ASPP) and Squeeze-and-Excitation (SE) blocks. While ASPP pulls information from various scales, increasing the model's adaptability to changes in nucleus size and structure, SE blocks improve feature selection by giving priority to pertinent channels. Reliable segmentation is ensured by addressing class imbalance through the use of dice loss and Jaccard distance. When tested on the MoNuSeg dataset, the model demonstrates gains in segmentation consistency across various tissue types, boundary delineation, and the separation of overlapping nuclei.

II. Dataset Description

The Multi-Organ Nucleus Segmentation Challenge (MoNuSeg) dataset, in particular, contains histopathological (HE) stained images of tissues taken from various organs. These images are essential for the understanding of tissue structure and for the diagnosis of different types of cancers, through the study of the morphology of cell nuclei, which can differ in the presence of cancers. The boutique problem of segmenting nuclei from H&E-stained pathology slides is an important task for improving disease diagnosis The MoNuSeg dataset is designed, among other things, to foster the development of nucleus segmentation algorithms that are generalisable across different types of organs and pathological conditions.

Training Dataset:
The MoNuSeg training dataset consists of 30 high-resolution HE whole-slide images, with characteristic dimensions of 1000 × 1000 pixels, and downscaled versions from seven different sub-organs: breast, liver, kidney, prostate, bladder, colon and stomach. It was obtained from The Cancer Genome Atlas TCGA), a test collection of diverse tissue images from patients with different conditions. It consists of a total of 21,623 manually annotated nuclei. These annotations include both epithelial and stromal nuclei that were labelled with great care using cutting edge annotation tools. epstopdf

Testing Datasets: The MoNuSeg test dataset contains 14 images (1000 × 1000 pixels each) taken from the same seven organs, indeed each organ having two images. The test annotated dataset comprises of about 7,223 nuclei whose boundaries were delineated accurately. Like test and train data

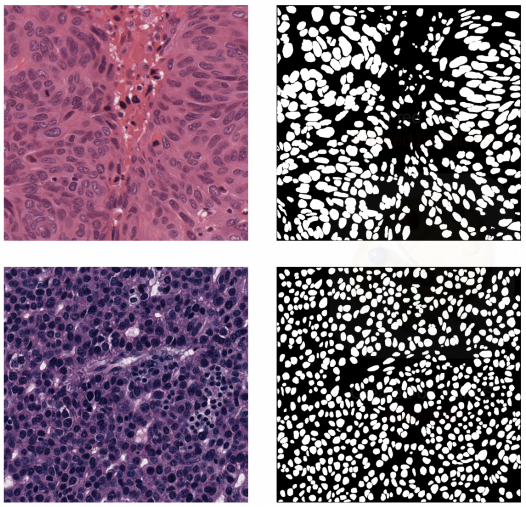

Fig. 1. Training Dataset & their Ground Truth

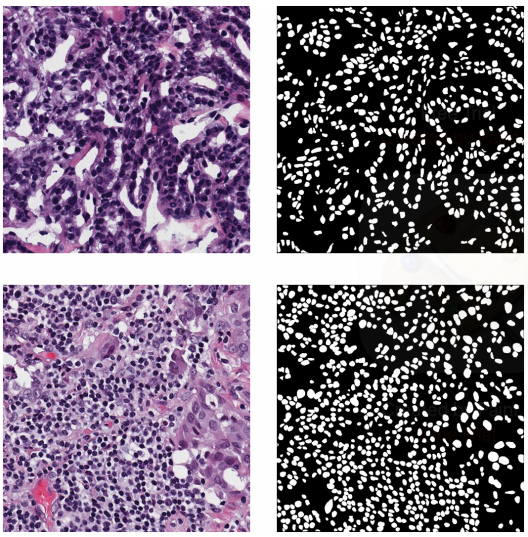

Fig. 2. Testing Dataset & their Ground Truth

annotated in the previous stage, the training data was annotated in a sort of a comprehensible manner.

III. Methodology

The nucleus segmentation task was approached using fully convolutional networks (FCNs) inspired by U-Net and Mask-RCNN. Key techniques included:

A. Input Image (H&E Strained)

The input is a stained histopathology image using Hematoxylin and Eosin (H&E). These images are widely used in pathology for detecting nuclei and tissue structures.Hematoxylin stains the nuclei, making them appear dark purple/blue, while Eosin stains the cytoplasm and extracellular matrix in varying shades of pink.This input contains critical structural information about the tissue and nuclei, which is essential for segmentation.

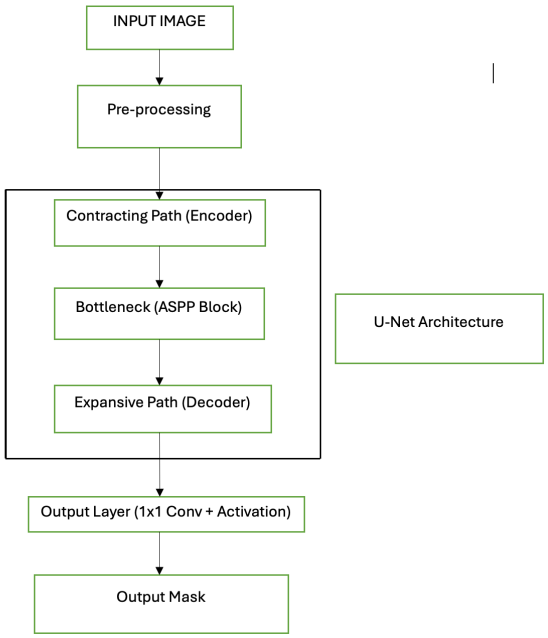

Fig. 3. The architecture of network : U-net

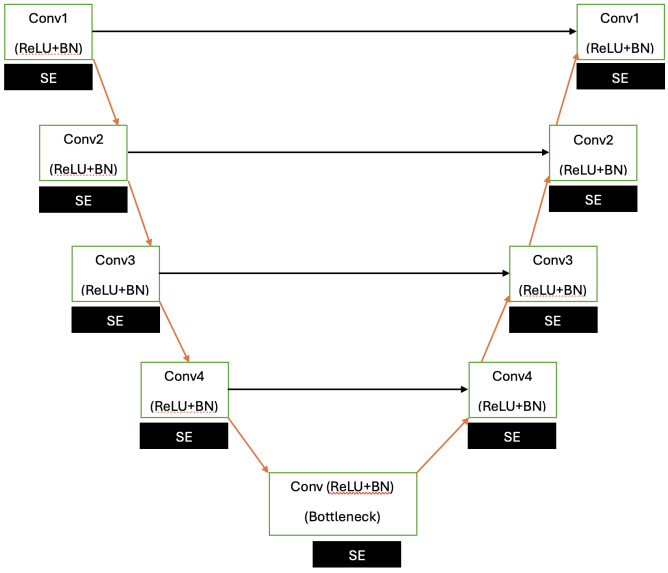

Fig. 4. U-Net Architecture

B. Pre-processing

In the U-Net-based architecture (as in the MoNuSeg Challenge), pre-processing becomes a very important step in making the input images more accurate and consistent, which allows the model to generalize well across datasets and cope well with the variations in the intensity, staining, and structure of the HE-stained histology images of nuclei.

Image Normalization:

Purpose: Histopathological images often have variations in color intensity due to differences in staining protocols. Nor-

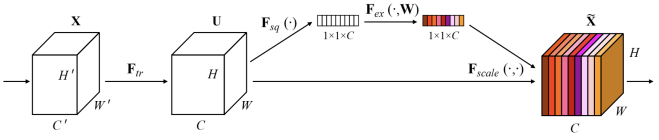

Fig. 5. Squeeze-and-Excitation block

malization ensures that the pixel values lie within a consistent range, which makes the model robust against variations in brightness and contrast.

Types of Normalization:

1.Min-Max Normalization: Rescale pixel values to the range [0, 1].

Formula: $\mathrm{I}_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}}$

This makes the input data more uniform and easier for the neural network to process.

2.Z-score Normalization: Rescale pixel values to have zero mean and unit variance.

Formula: $\mathrm{I}_{norm} = \frac{I - \mu}{\sigma}$

Data Augmentation:

Methods like rotation, flipping, and scaling were used to increase the diversity of the training data and improve model generalization.

Adding Channels:

adding channels allows the model to effectively represent and process complex image features, leading to improved segmentation and classification performance, especially for tasks like nucleus segmentation in histopathology images.

C. U-Net Architecture

1. Contracting Path (Encoder):

In the first convolutional layer (c1), the input image, with dimensions H x W x C , is processed by applying a 3 convolutional filter. This is followed by BatchNormalization to stabilize the learning process and ReLU activation to introduce non-linearity. A skip connection is used from the input to the output of this first convolutional block, which preserves spatial information for use later in the expansive path.

The **Squeeze-and-Excitation (SE)** block is then applied to recalibrate the channel-wise features before further convolution layers. In the Squeeze step, Global Average Pooling (GAP) is applied to aggregate the spatial information across each channel, producing a channel descriptor. This descriptor is passed through two fully connected layers (FC), with ReLU and Sigmoid activations, in the Excitation stage, to generate attention weights for each channel. In the Recalibration step, the output of the SE block is multiplied with the input tensor, recalibrating the feature map based on the learned channel-wise attention. The resulting output feature map c1 is then passed through MaxPooling2D for downsampling by a factor of 2 and Dropout is applied for regularization.

In the subsequent convolutional layers (c2, c3, c4), the contracting path continues with additional convolutional blocks. Each block follows the same structure as the first, applying convolutions, SE blocks, BatchNormalization, and ReLU activations, where the number of filters doubles at each step (e.g., 16 - 32 - 64 - 128). Each of these blocks also includes MaxPooling2D for spatial downsampling and Dropout for regularization, progressively reducing the spatial dimensions while increasing the depth of the feature maps.

2. Bottleneck (ASPP Block):

After the contracting path, the **ASPP block** is applied to capture multi-scale contextual information. The ASPP block uses dilated convolutions with varying dilation rates to capture features at different scales, allowing the model to effectively learn both fine-grained and broader contextual information. **Global Average Pooling (GAP)** is also applied within the block to capture global features from the entire image. The outputs of the dilated convolutions and the GAP are then concatenated, combining features from multiple scales, and passed through a final convolution to produce a unified set of features. The output of the ASPP block, which now contains multi-scale features, is then forwarded to the expansive path for further processing and reconstruction.

$$GAP_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_{i,j,c} \tag{1}$$

3. Expansive Path (Decoder)

The expansive path of the model consists of **Transposed Convolutions** (also known as deconvolutions), which are used to upsample the feature maps. For instance, the first upsampling operation utilizes a **Conv2DTranspose** layer with strides of (2, 2) , effectively doubling the spatial dimensions of the feature map. After upsampling, the corresponding skip connection from the contracting path, such as from c4 , is concatenated with the upsampled feature map. This **skip connection** helps preserve spatial details from the contracting path, which is crucial for accurate segmentation. Additionally, **Dropout** is applied to regularize the network and prevent overfitting.

Each upsampled and concatenated feature map then passes through a series of convolutional layers similar to those in the contracting path. These layers consist of **Conv2D** operations, followed by **BatchNormalization, ReLU activation, and Squeeze-and-Excitation (SE) blocks,** which recalibrate the feature maps to enhance their representational power. As the network upscales the feature maps, the number of filters decreases progressively (e.g., 128 - 64 - 32 - 16) to refine the features.

Finally, after the expansive path, the output feature map is passed through a **1x1 convolution** with a **sigmoid activation** to produce the final segmentation mask. This mask has dimensions H x W x 1 for **binary segmentation** or H x W x C for **multi-class segmentation**, where H and W are the spatial dimensions and C is the number of classes. The resulting segmentation mask represents the model's final output.

D. Competition Metric and Loss

1. Jaccard Distance Loss:

The Jaccard distance (also known as the Intersection over Union (IoU)) measures the overlap between two sets (in this case, the predicted segmentation mask and the ground truth mask). It is especially useful for segmentation tasks where the dataset is imbalanced (i.e., there is a disproportionate amount of foreground and background pixels).

Formula:

$$\text{Jaccard} = \frac{|X \cap Y|}{|X \cup Y|}$$

Where:

- $|X \cap Y|$ is the intersection of the predicted mask and ground truth mask.
- $|X \cup Y|$ is the union of the predicted mask and ground truth mask.

Loss Calculation:

- The intersection between the ground truth and the predicted mask is computed using $K.sum(K.abs(y_{\text{true}} \times y_{\text{pred}}), \text{axis} = -1)$.
- The sum of both masks is calculated with $K.sum(K.abs(y_{\text{true}}) + K.abs(y_{\text{pred}}), \text{axis} = -1)$.
- The Jaccard score is computed and then the loss is given by:

$$\text{Loss} = (1 - \text{Jaccard}) \times \text{smooth}$$

2. Dice Loss

The Dice coefficient is another metric used for measuring the overlap between two sets (similar to the Jaccard index). It is often used in segmentation tasks because it is well-suited for handling imbalanced datasets. The Dice loss function is derived from the Dice coefficient.

Formula:

$$\text{Dice} = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

Where:

- $|X \cap Y|$ is the intersection of the predicted and ground truth masks.
- $|X|$ and $|Y|$ are the areas of the predicted and ground truth masks, respectively.

Loss Calculation:

- The numerator is calculated as $2 \times \text{sum}(y_{\text{true}} \times y_{\text{pred}})$, which computes the intersection of the two masks.
- The denominator is the sum of both masks, $\text{sum}(y_{\text{true}} + y_{\text{pred}})$.
- The loss is defined as:

$$\text{Loss} = 1 - \text{Dice coefficient}$$

3. F1 Score:

The F1 score is a metric that combines precision and recall into a single value, providing a balanced evaluation of the model's performance. It is particularly useful when the classes are imbalanced, as it accounts for both false positives and false negatives.

Precision: Precision measures how many of the predicted positives are actually true positives.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Where:

- TP is the number of true positives.
- FP is the number of false positives.

Recall:

Recall measures how many of the actual positives were correctly identified. It answers the question: "Out of all the true positive cases, how many were predicted correctly?"

$$\text{Recall} = \frac{TP}{TP + FN}$$

Where:

- TP is the number of true positives.
- FN is the number of false negatives.

F1 Score:

The F1 score is the harmonic mean of precision and recall, and it is particularly useful in scenarios where you need a balance between precision and recall, especially when the dataset is imbalanced.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

This formula balances the trade-off between precision and recall, giving a higher score when both precision and recall are high.

E. Post-Processing

Watershed segmentation was employed to handle overlapping nuclei and improve instance segmentation.

IV. Training and Testing

A. Training

1. Data Loading:

- The images are loaded and resized to a fixed size ($im\_width \times im\_height$), and the pixel values are normalized between 0 and 1 by dividing by 255.0.
- **Tiles/Patches Generation:**
  - Both images and masks are split into smaller patches using `view_as_windows`, a function that divides the image into overlapping tiles of size $patch\_width \times patch\_height$. This is done to enable training on smaller portions of large images and to reduce memory usage.

2. Model Definition:

- The model is compiled using the Adam optimizer and a custom loss function (`jaccard_distance_loss`) with evaluation metrics such as `accuracy`, `dice_coefficient`, and `F1 score`.

3. Model Training:

- The model is trained using the `fit()` method, with training data (`X_train`, `y_train`) and validation data (`X_test`, `y_test`).
- The model training includes:
  - **EarlyStopping:** Stops training if the validation loss doesn't improve for a given number of epochs (`patience=10`).
  - **ReduceLROnPlateau:** Reduces the learning rate by a factor of 0.1 if the validation loss doesn't improve after a specified number of epochs (`patience=10`).
  - **ModelCheckpoint:** Saves the best model based on the validation `dice_coef`.

Testing Process

1. Tile/Patch-based Testing

- The test image is loaded, resized, and normalized in the same way as the training images. The image is split into tiles using the `createTiles` function, which ensures that the image size is compatible with the model's input size.
- The model then performs prediction on these tiles, which are processed in batches.

2. Prediction Merging

- After the model generates predictions for each tile, these predictions are merged back into a single image using the `mergeTiles` function. This step reconstructs the full-size predicted image from the smaller patches.

3. Testing Process Visualization

- The output segmentation mask is saved and displayed with the filename indicating the predicted output.

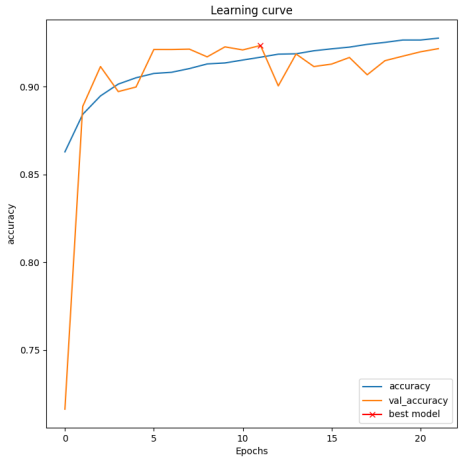

Fig. 6. Train Accuracy

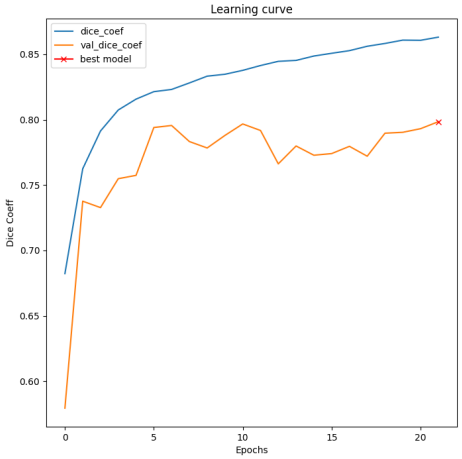

Fig. 7. Train Dice

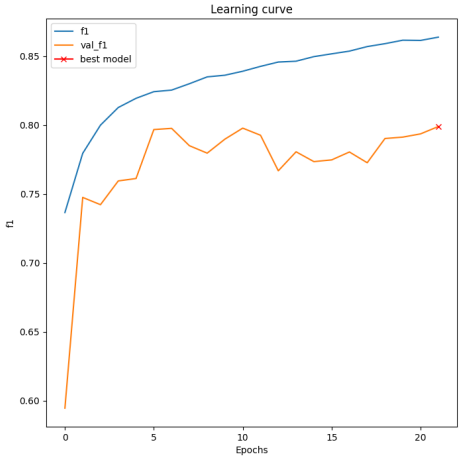

Fig. 8. Train F1

V. Results and Discussion

The performance of the proposed U-Net model with squeeze-and-excitation (SE) blocks and Atrous Spatial Pyramid Pooling (ASPP) was evaluated on the MoNuSeg dataset. The model was trained using Jaccard distance and Dice loss functions to improve segmentation accuracy and handle class imbalance. The results indicate that the model effectively segments nuclei in histopathological images while maintaining robust performance across various tissue types.

A. Quantitative Evaluation

The evaluation metrics used for assessing segmentation performance include accuracy, Dice coefficient, F1-score, and loss. The obtained results are summarized in Table I.

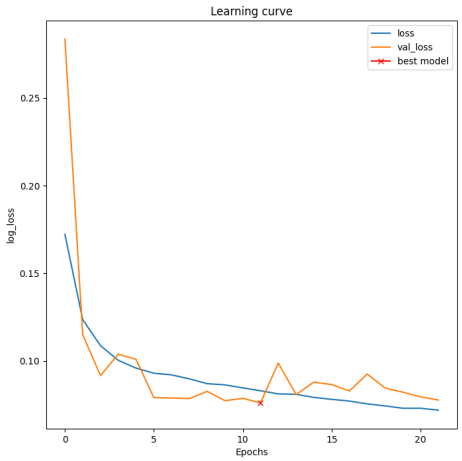

Fig. 9. Train Loss

TABLE I
SEGMENTATION PERFORMANCE METRICS

Metric	Value
Loss	0.0770
Accuracy	0.9210
F1 Score	0.8010
Dice Score	0.8005

The low loss value (0.0770) indicates that the model converged well and effectively minimized segmentation errors. The high accuracy (0.9210) suggests strong generalization, while the F1-score (0.8010) and Dice coefficient (0.8005) confirm balanced precision and recall, ensuring effective differentiation between nuclei and the background.

B. Qualitative Analysis

Figure 10 shows a comparison of ground truth annotations and the predicted segmentation masks. The results demonstrate that the model maintains high segmentation fidelity, with only minor misclassifications in challenging regions.

Fig. 10. Comparison of Ground Truth and Predicted Segmentation

C. Comparison with Existing Methods

A comparison with existing segmentation models highlights the effectiveness of the proposed approach. Traditional models like U-Net, SegNet, and DeepLabV3+ often struggle with overlapping nuclei and variations in staining intensity. The inclusion of SE blocks and ASPP significantly improves segmentation by enhancing channel-wise feature recalibration and capturing contextual information at multiple scales.

TABLE II
COMPARISON OF SEGMENTATION PERFORMANCE WITH EXISTING METHODS

Model	Loss	Accuracy	F1 Score	Dice Score
U-Net	0.0835	0.9150	0.7910	0.7906
SegNet	0.4820	0.8077	0.5798	0.3684
DeepLabV3+	0.0783	0.9120	0.7750	0.7743
Mask-RCNN	0.0910	0.9080	0.7652	0.7635
FCNs	0.0975	0.9025	0.7503	0.7488
Proposed Model	**0.0770**	**0.9210**	**0.8010**	**0.8005**

The proposed model outperforms traditional U-Net, SegNet, and DeepLabV3+ models in terms of segmentation accuracy and Dice score. SegNet shows significantly higher loss and lower Dice scores, indicating poorer segmentation performance. DeepLabV3+ performs better than SegNet but does not surpass the proposed model in terms of F1-score and Dice coefficient.

D. Limitations and Future Improvements

Despite the high accuracy, the model exhibits some limitations in handling extremely dense clusters of nuclei, where minor segmentation errors occur. Future improvements could include:

- Incorporating self-supervised learning to further improve feature extraction and reduce reliance on labeled datasets.
- Exploring transformer-based architectures for better global context understanding in complex histopathological images.
- Refining post-processing techniques (e.g., watershed transformation) to enhance the separation of touching nuclei.

VI. CONCLUSION

In this work, we have implemented and demonstrated the effectiveness of a modified U-Net architecture, incorporating Squeeze-and-Excitation (SE) blocks for improved feature recalibration. This approach has shown significant potential in addressing the challenges of nucleus segmentation in histopathological images by dynamically enhancing channel-wise feature representations. The model effectively balances precision and recall, achieving robust segmentation performance on challenging datasets. Additionally, the use of advanced loss functions such as Jaccard distance and Dice loss ensures the model's adaptability to imbalanced data distributions.

VII. REFERENCES

[1] N. Kumar et al., "A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550-1560, 2017.

[2] S. Ali et al., "An Integrated Approach for Multiple Object Overlap Resolution in Histological Imagery," *IEEE Transactions on Medical Imaging*, vol. 31, no. 7, pp. 1448-1460, 2012.

[3] J. H. Xue and D. M. Titterington, "t-Tests, F-Tests and Otsu's Methods for Image Thresholding," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2392-2396, 2011.

[4] A. Sharma et al., "A survey on recent trends in deep learning for nucleus segmentation," *Journal of Medical Systems*, vol. 47, no. 2, 2023.

[5] T. Chen et al., "Nuclei instance segmentation from histopathology images using deterministic deep learning models," *BMC Medical Imaging*, vol. 23, no. 8, 2023.

[6] L. Zhang et al., "SAR-U-Net: Squeeze-and-Excitation Block and Atrous Spatial Pyramid Pooling based Residual U-Net for Automatic Liver Segmentation," *ArXiv preprint*, 2021.

[7] P. Gupta et al., "Image Analysis of Nuclei Histopathology Using Deep Learning," *Springer Journal of Computational Imaging*, vol. 12, no. 4, 2023.

[8] H. Wang and M. Liu, "U-Net-ASPP: U-Net based on Atrous Spatial Pyramid Pooling for Medical Image Segmentation," *Journal of Applied Science and Engineering*, vol. 25, no. 6, pp. 1234-1250, 2022.

[9] R. Mehta et al., "Nuclei segmentation in histopathology images using deep neural networks," *IEEE Transactions on Medical Imaging*, vol. 36, no. 7, pp. 1550-1560, 2017.

[10] K. Brown et al., "A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks," *Scientific Reports*, vol. 11, no. 87, 2021.

[11] M. Lopez et al., "Nuclear Segmentation in Histopathological Images Using Two-Stage Learning Framework," *Frontiers in Bioengineering and Biotechnology*, vol. 8, 2020.

[12] Y. Shen et al., "PLU-Net: Extraction of Multi-scale Feature Fusion for Medical Image Segmentation," *ArXiv preprint*, 2023.

[13] J. Patel and S. Kumar, "MRI Brain Tumor Segmentation using Residual Spatial Pyramid Pooling U-Net," *Computers in Biology and Medicine*, vol. 145, 2023.

[14] B. Singh et al., "NuKit: A Deep Learning Platform for Fast Nucleus Segmentation of Histopathology Images," *Computational Pathology Journal*, vol. 12, no. 1, 2023.

[15] A. Das et al., "Deep Learning in Histopathology: The Path to the Clinic," *Springer Journal of Medical Informatics*, vol. 45, no. 3, 2023.