Nucleus Segmentation with Modified Attention Gate UNET using Sobel Edge Detection and Global-Local Encoders

Submitted in partial fulfilment of the requirements of the degree of

Bachelor of Technology

by

Astitva Verma	**207111**
Adarsh Rao	**207102**
Akula Pawan Kalyan	**207105**

Supervisor:

Dr Raju Bhukya

Associate Professor
NIT Warangal

Computer Science and Engineering

DECLARATION

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and can also evoke penal action from the sources which have thus not been properly cited or from where proper permission has not been taken when needed.

Astitva Verma - 207111

Adarsh Rao - 207102

Akula Pawan Kalyan - 207105

Date: 11th May, 2024.

Abstract

Histopathology images play a crucial role in modern medical diagnostics, offering detailed insights into the structural and cellular composition of tissues. These images are obtained through the microscopic examination of thin tissue sections stained with various dyes, allowing pathologists to visualise cellular morphology, identify abnormalities, and diagnose diseases.

In the context of histopathology, the term "nucleus segmentation" refers to the process of delineating and identifying the nuclei within tissue samples. Nuclei serve as vital markers for cellular activity, providing valuable information about cell proliferation, differentiation, and pathology. Accurate nucleus segmentation is essential for quantifying cellular characteristics, measuring biomarkers, and analysing tissue morphology, all of which are critical tasks in histopathological analysis.

Despite its importance, nucleus segmentation in histopathology images presents significant challenges due to the complex and heterogeneous nature of tissues. Factors such as variations in staining intensity, tissue thickness, cellular density, and the presence of artefacts can impede accurate segmentation. Traditional manual segmentation methods are labour-intensive, time-consuming, and prone to subjectivity, making them unsuitable for large-scale analysis or time-sensitive clinical settings.

To tackle these challenges, automated image analysis methods, particularly the U-Net architecture, have gained popularity. U-Net is specifically designed for biomedical image segmentation, featuring a contracting path for context extraction and an expansive path for precise localization. Its ability to retain fine details while aggregating high-level features makes it well-suited for nucleus segmentation in histopathology images. By leveraging skip connections and end-to-end learning, U-Net offers a scalable and efficient solution for automating this critical task.

This study presents a novel methodology aimed at enhancing the segmentation accuracy of nuclei in hematoxylin and eosin (H&E) stained histopathology images. Accurate segmentation of nuclei is crucial for precise cancer diagnosis and prognosis, providing both qualitative and quantitative insights. The report addresses the challenge of accurately segmenting nuclei with variable sizes and ambiguous boundaries, which significantly impacts the reliability of the segmentation process.

This study also introduces key innovations such as the integration of edge information extracted from input data into the widely used UNET model for image segmentation tasks. By modifying

the attention gate mechanism within the UNET model, the focus is shifted towards prioritising edge features during segmentation. Additionally, the project incorporates global features by introducing a transformer self-attention encoder, allowing the segmentation mask prediction to consider both local and global feature dependencies. The fusion of global and local features using a novel fusion module is crucial for improving the delineation of nuclei boundaries, thereby reducing segmentation errors and enhancing overall accuracy.

In summary, the project report presents an innovative approach to enhance nuclei segmentation in H&E stained histopathology images by leveraging edge information and global feature dependencies. The integration of edge-guided attention and global-local encoders offers a promising solution to the challenges associated with variable-sized nuclei and unclear boundaries, ultimately improving the accuracy of cancer diagnosis and prognosis in histopathology image analysis.

Keywords - Attention Mechanisms, Transformer Self-Attention Encoder, Local-Global Feature Fusion, Nucleus Segmentation, U-Net Architecture, Biomedical Image Analysis, Automated Image Segmentation

Contents

List of Figures and Equations

Chapter 1

Introduction

Histopathology images are a fundamental resource in biomedical imaging, offering a microscopic view of tissue structures and cellular compositions. In the context of bio-image sequencing, these images play a critical role in understanding the complex arrangement of cells and tissues within biological samples. By examining these images, pathologists and researchers can uncover detailed information about cellular morphology, tissue organisation, and the presence of disease markers. This in-depth analysis is crucial in diagnosing various conditions, particularly cancers, where subtle changes in tissue architecture can indicate malignancy. Furthermore, histopathology images are used to assess the effectiveness of medical treatments, track disease progression, and guide surgical decisions by providing insights into the spatial relationships between cells and tissues. Through advanced digital imaging techniques, these images can be processed to create bio-sequences that highlight dynamic changes over time or across different tissue regions. This ability to capture and analyse tissue at a cellular level is invaluable for both clinical diagnostics and biomedical research, offering a window into the intricate patterns that underpin health and disease.

1.1 Histopathology Images

Histopathology images are visual representations of tissue samples that have been examined under a microscope to study the microscopic anatomy of cells and tissues, particularly in relation to disease. These images are created during a process known as histopathology, which involves taking a tissue sample from a patient (through a biopsy or surgery), processing it in a laboratory, and then examining it to identify any abnormalities or signs of disease.

Obtaining histopathology images for analysis consists of the following steps:

- **Sample Collection**: Tissue samples are collected from a patient, often during surgery, biopsy, or autopsy. These samples are preserved in a fixative solution, typically formalin, to maintain their structure.
- **Processing and Staining**: The preserved tissue samples are cut into thin sections and mounted onto glass slides. These sections are then stained with various dyes to highlight different components of the tissue, such as the commonly used Hematoxylin and Eosin (H&E) stain, which provides contrast to cellular structures.
- **Microscopic Examination**: The stained tissue sections are examined under a microscope by a histopathologist, a specialist in diagnosing diseases at the microscopic level. During this examination, the pathologist looks for abnormal changes in cells and tissue architecture, such as signs of cancer, inflammation, or other pathological conditions.
- **Analysis and Interpretation**: The pathologist uses histopathology images to identify and diagnose diseases based on the visual characteristics of cells and tissues. These images are used to understand the nature and extent of a disease, which guides medical decision-making, such as treatment options and prognosis.

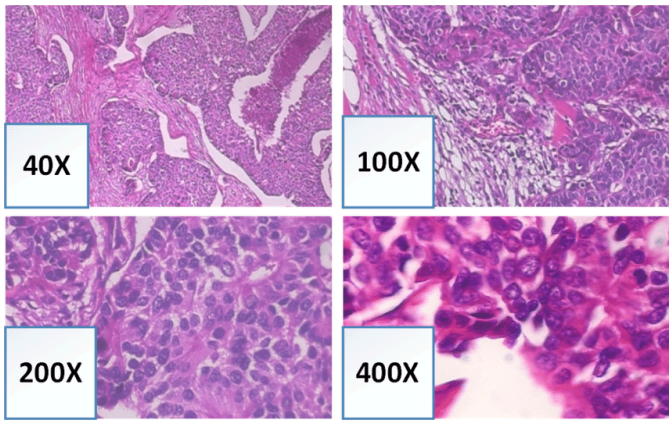


Figure 1.1: Histopathology Images

1.2 Histopathology Images and Segmentation

Nucleus segmentation in histopathology is a specialised technique used to identify and delineate the boundaries of cell nuclei within tissue samples, typically visualised under a microscope. This process is crucial in pathology because nuclei are the command centres of cells, containing genetic material that drives cellular function, division, and differentiation. As such, the accurate

identification and segmentation of nuclei from histopathology images are essential for several reasons.

Nucleus segmentation allows for the quantification of cellular characteristics. By counting nuclei, researchers and pathologists can estimate cell density in a given tissue area, providing insights into processes like cell proliferation, which is a key indicator of tissue growth and, in some cases, tumorigenesis. This information is critical for understanding the pathology of diseases, especially cancers where abnormal proliferation is a hallmark feature.

The segmentation process also aids in measuring biomarkers. Certain diseases, particularly cancers, exhibit specific biomarkers that are often localised within or around nuclei. Accurate segmentation allows for precise measurement of these biomarkers, enabling pathologists to track disease progression, evaluate the aggressiveness of a tumour, and predict patient outcomes. In some cases, the spatial distribution of these biomarkers relative to nuclei can inform treatment strategies, making nucleus segmentation a valuable tool for personalised medicine.

Nucleus segmentation is instrumental in analysing tissue morphology. The shape, size, and distribution of nuclei can reveal significant pathological patterns. For example, irregular or enlarged nuclei might indicate malignancy, while a consistent and uniform appearance could suggest normal tissue. The ability to delineate and categorise these morphological features supports pathologists in their diagnostic assessments and contributes to research studies examining the underlying causes of diseases.

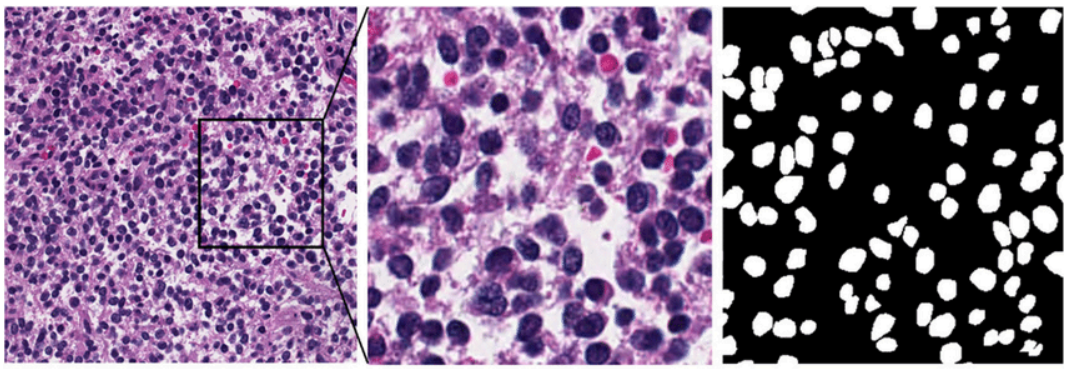


Figure 1.2: a) Histopathology image, (b) Zoomed image, and (c) Ground-truth image for (b). Blue regions in (a,b) and white areas in (c) represent nuclei, which need to be segmented.

1.3 Nuclear Segmentation using the U-Net Architecture

Nuclear segmentation using U-Net is a popular and effective approach in the field of histopathology for identifying and delineating nuclei within tissue samples. U-Net, a deep learning architecture designed for biomedical image segmentation, is well-suited for this task due to its unique structure, which balances the need for high-resolution segmentation with robust contextual information.

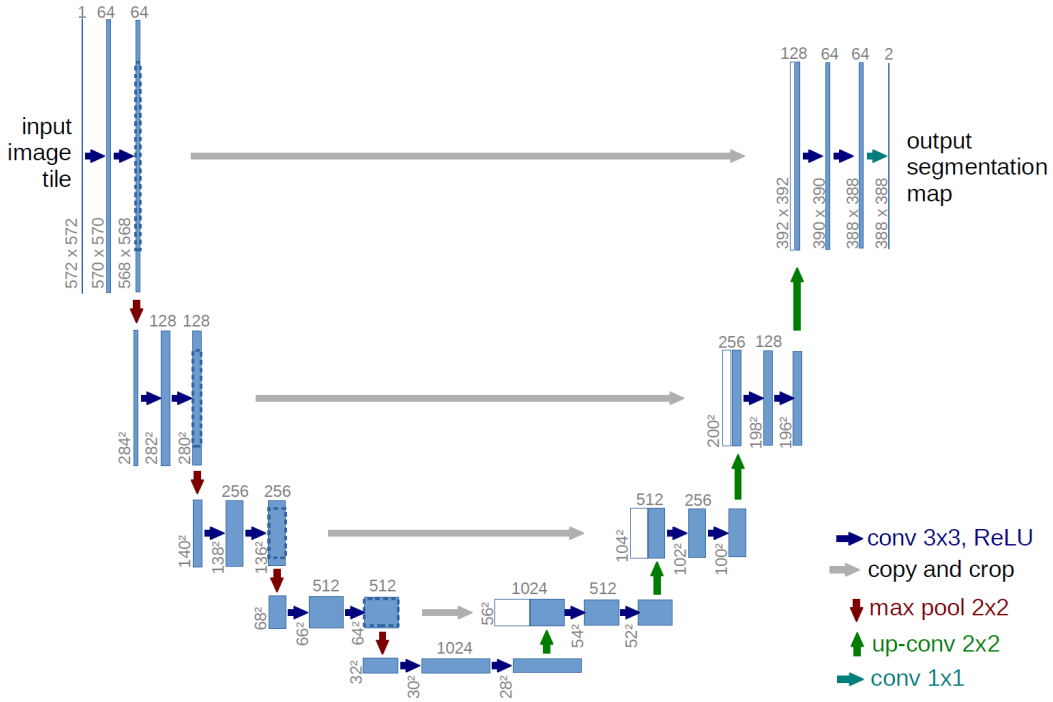


Figure 1.3: Example U-Net Architecture

The U-Net architecture is characterised by its "U" shape, with two main paths: a contracting path and an expansive path. The contracting path, also known as the encoder, involves a series of convolutional layers that progressively reduce the spatial dimensions of the input image while increasing the depth of feature representations. This pathway captures the broader context of the image, allowing the model to understand general patterns and structures.

The expansive path, or decoder, is where the image resolution is gradually restored through a series of deconvolution (up-sampling) layers. This path is designed to recover fine details and achieve precise segmentation. Crucially, U-Net employs skip connections between the

contracting and expansive paths, allowing information to flow directly from early layers to later ones. These connections help preserve spatial information, ensuring that the output segmentation retains high accuracy and detail.

For nuclear segmentation in histopathology, U-Net has proven particularly effective because of its ability to:

- **Handle Complex Structures**: Tissue samples often exhibit complex structures with overlapping cells, varying staining intensities, and ambiguous boundaries. U-Net's contracting path effectively captures broader contextual information, while the expansive path provides the detailed localization needed to delineate individual nuclei.
- **Retain Spatial Information**: The use of skip connections allows U-Net to maintain a balance between global context and fine details, ensuring that nuclei boundaries are accurately identified and segmented.
- **Scale with Data Size**: U-Net can be applied to various resolutions and scales, making it adaptable to different types of histopathology images, from small tissue sections to larger whole-slide images.

To train a U-Net model for nuclear segmentation, a dataset of histopathology images with annotated nuclei is used. The model learns to segment nuclei by analysing these examples, adjusting its parameters to minimise the error between predicted segmentation masks and the ground truth. Once trained, U-Net can automatically segment nuclei from new histopathology images, providing a fast and consistent method for analysis.

Chapter 2

2.1 Problem Statement

The research paper proposes the development and evaluation of a novel approach for nucleus segmentation in histopathology images using edge-guided attention and global-local encoders. The study aims to enhance the accuracy of nucleus segmentation by leveraging edge information and global-local encoders within the UNET architecture. The research explores the integration of edge maps extracted from input images to improve attention allocation and delineation of nucleus boundaries, particularly focusing on addressing segmentation errors along edges.

Objective

The primary objective of the research is to investigate and implement an innovative methodology for nucleus segmentation in histopathology images. The specific objectives include:

1. Introducing edge-guided attention and global-local encoders to the UNET architecture for improved nucleus segmentation accuracy.
2. Enhancing the attention gate mechanism within the UNET model to prioritize edge features during segmentation.
3. Incorporating a transformer self-attention encoder to consider global feature dependencies for more precise segmentation mask prediction.
4. Developing a fusion module to combine global and local features, aiming to reduce segmentation errors and enhance overall accuracy in nucleus boundary delineation.
5. Experimenting with benchmark datasets, such as the KMC Liver dataset and the Monuseg dataset, to evaluate the performance of the proposed approach and compare it with existing models.

By addressing these objectives, the research aims to advance the field of histopathology image analysis by introducing a novel methodology for nucleus segmentation. The study seeks to contribute to the automation of nucleus segmentation tasks, improve the accuracy of cellular analysis in histopathology images, and provide insights for potential applications in medical diagnostics and research.

Chapter 3

Related Work

3.1 NucleiSegNet: Tailored Deep Learning Architecture for Nucleus Segmentation

The study by Lal et al. [1], introducing NucleiSegNet, presents a tailored deep learning architecture designed for segmenting nuclei in HE stained liver cancer histopathology images. By incorporating robust residual blocks and attention mechanisms, the model addresses challenges such as segmenting touching nuclei and detecting variable shapes, ultimately outperforming existing models both quantitatively and qualitatively. This study not only contributes a new dataset of annotated liver nuclei images but also suggests potential future applications and extensions in histopathology image analysis. The innovative approach of NucleiSegNet underscores the importance of attention mechanisms in improving segmentation accuracy, providing valuable insights for the development of advanced segmentation models in medical imaging.

3.2 Attention Gate Model: Enhancing Sensitivity and Accuracy in Medical Imaging

In the work by Oktay et al. [2], the introduction of an attention gate (AG) model for medical imaging showcases a significant advancement in focusing on target structures of varying shapes and sizes to enhance sensitivity and accuracy. By automatically learning to highlight salient features and suppress irrelevant regions in input images, the AG model eliminates the need for external localization modules, offering a streamlined approach to image analysis. The seamless integration of AGs into standard CNN architectures like U-Net with minimal computational overhead demonstrates notable enhancements in tissue/organ identification and localization, particularly benefiting the analysis of small organs like the pancreas. This study highlights the potential of attention mechanisms in improving the efficiency and effectiveness of image segmentation tasks, paving the way for further advancements in medical image analysis and diagnosis.

3.3 Deep Residual U-Net for Road Extraction in Aerial Images

The study by Zhang et al. [3] introduces a novel method for road extraction from aerial images using a Deep Residual U-Net model, combining residual learning and U-Net architecture to enhance road area segmentation accuracy in remote sensing images. By surpassing existing state-of-the-art techniques in terms of precision and recall, the proposed approach demonstrates its effectiveness in remote sensing applications. The integration of residual learning with U-Net architecture showcases the potential of combining different deep learning concepts to improve segmentation tasks, setting a benchmark for road extraction accuracy in aerial imagery analysis. This study not only advances the field of remote sensing image analysis but also underscores the importance of model architecture design in achieving superior segmentation results in complex image datasets.

3.4 Effectiveness of U-Net in Cell Segmentation Challenges

Ronneberger et al. [4] showcase the effectiveness of U-Net across segmentation challenges, particularly in cell segmentation in light microscopy images. By outperforming previous methods in accuracy and speed, especially on challenging 2D transmitted light datasets, U-Net proves to be a versatile and robust architecture for biomedical image segmentation tasks. The authors provide implementation details, trained networks, and supplementary materials, facilitating further research and application in the field of biomedical image analysis. The success of U-Net in retaining fine details while aggregating high-level features highlights its suitability for tasks requiring precise localization and segmentation, emphasising its significance in advancing automated image analysis techniques for medical diagnostics and research.

3.5 Transformer Model: Revolutionising Sequence Transduction Tasks

Vaswani et al. [5] introduce the Transformer model, a neural architecture that relies solely on attention mechanisms for sequence transduction tasks, showcasing state-of-the-art results in machine translation. The Transformer model's key innovation lies in its self - attention mechanism, enabling the model to focus on different parts of the input sequence during output sequence generation, thereby enhancing interpretability and performance. By outperforming traditional models in machine translation tasks, the Transformer model demonstrates the power of attention mechanisms in capturing long-range dependencies and improving the quality and efficiency of sequence transduction tasks. This study marks a significant advancement in natural language processing and sets a new standard for attention-based neural architectures in various domains beyond translation tasks.

Chapter 4

Proposed Solution

4.1 Overview

Our work explores an innovative approach to enhance nucleus segmentation in histopathology images. The study addresses the challenges associated with accurately segmenting nuclei with variable sizes and ambiguous boundaries, crucial for precise cancer diagnosis and prognosis. By integrating edge information extracted from input data into the UNET model and modifying the attention gate mechanism, the research aims to prioritise edge features during segmentation. Additionally, the study incorporates global features through a transformer self-attention encoder to consider both local and global feature dependencies for segmentation mask prediction. The fusion of global and local features using a novel fusion module is proposed to improve the delineation of nuclei boundaries, reduce segmentation errors, and enhance overall accuracy. Through this methodology, the research aims to contribute to the advancement of automated detection of nuclei in histopathology images, offering a comprehensive solution to the challenges in nucleus segmentation for medical diagnostics and analysis.

Our methodology includes the following steps:

- **Convolutional Encoder**: The encoders in a U-Net architecture are responsible for extracting hierarchical features from input images. Typically composed of convolutional layers followed by downsampling operations, they reduce spatial resolution while increasing feature depth. These features capture both low-level details and high-level semantics, crucial for accurate segmentation. Leveraging pretrained CNNs and skip connections, U-Net encoders efficiently extract meaningful features to enhance segmentation performance.
- **Decoder:** In U-Net, decoders upsample features and fuse them with skip connections for precise segmentation. Using transposed convolutions, they recover spatial details lost during encoding. Decoders integrate features from multiple scales, improving

segmentation accuracy. They play a vital role in reconstructing high-resolution segmentation masks.

- **Preprocessing with Sobel Edge Detection**: The input histopathology images undergo preprocessing using Sobel edge detection to extract edge maps. This step aims to enhance the visibility of edges in the images, which is crucial for accurate nucleus segmentation.

- **Modification of Attention Gate Mechanism**: The attention gate within the UNET architecture is modified to prioritise edge features during segmentation. By incorporating edge information into the attention mechanism, the model can focus on important edge details for improved segmentation accuracy.

- **Introduction of Transformer Encoder**: In addition to the convolutional encoder traditionally used in UNET, a parallel encoder based on the transformer architecture is introduced. This global-local encoder captures global dependencies in the input data, allowing the model to consider not only local features but also global information for more accurate nucleus boundary delineation.

- **Fusion of Global and Local Features**: A novel fusion module is employed to combine the global features extracted by the transformer encoder with the local features obtained from the convolutional encoder. This fusion of global and local information aims to enhance the model's ability to accurately delineate nucleus boundaries, reducing segmentation errors and improving overall segmentation accuracy.

4.2 Methodology

4.2.1 Convolutional Encoder

The encoder of the UNET model, adapted from the NucleiSegNet[1] paper, employs 4 residual blocks and incorporates 3 downsampling operations. Each residual block comprises a 3×3 kernel convolution layer followed by a separable convolution layer, then another 3×3 convolutional layer. Batch Normalisation and ReLU activation layers follow all convolutions. These residual blocks are iterated four times, with three downsampling layers interspersed between them, halving the output image size each time. Post each residual block, the channel count is doubled to retain vital information. The encoder meticulously captures intricate image details crucial for subsequent segmentation tasks in the decoder.

4.2.2 Decoder

The features extracted in the encoder are used to reconstruct the segmentation map. We pass the compressed features from the bottleneck layer to four consecutive Convolution blocks, each containing two Convolution layers followed by Batch Normalisation. Each Convolutional block is followed by an attention block. After each convolution, the output image size is doubled and the number of channels are halved.

4.2.3 Preprocessing with Sobel Edge Detection

The Sobel operator, sometimes known as the Sobel-Feldman operator or Sobel filter, is a fundamental tool in image processing and computer vision, widely used in edge detection. This operator calculates an approximation of the gradient of the image intensity function, helping to identify areas in the image with rapid intensity changes, which typically signify edges. The Sobel operator achieves this by applying discrete differentiation, typically using 3x3 convolution kernels to estimate gradients in both the horizontal (x) and vertical (y) directions.

Definition and Operation

The Sobel operator employs two small, integer-valued, separable 3x3 kernels to convolve with an input image, yielding two derivative approximations: one for horizontal changes (G_x), and one for vertical changes (G_y). Given an image A, these derivative approximations are defined as follows:

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \qquad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

$$I_x = I * G_x \,,\; I_y = I * G_y$$

Equation 4.1: Sobel Edge Detection Filter

Here, the * symbol denotes the 2D convolution operation. The first kernel calculates the horizontal gradient, while the second calculates the vertical gradient. These gradient images are

then combined to compute the gradient magnitude M at each pixel location using the following equation:

$$M = \sqrt{I_x^2 + I_y^2}$$

Equation 4.2: Computation of Pixel-Wise Gradient Magnitude

Next, we use a threshold on the gradient magnitude image to convert it into a binary format. Pixels with gradient magnitudes exceeding a certain threshold are classified as edge pixels, while those with lower magnitudes are treated as non-edge pixels.

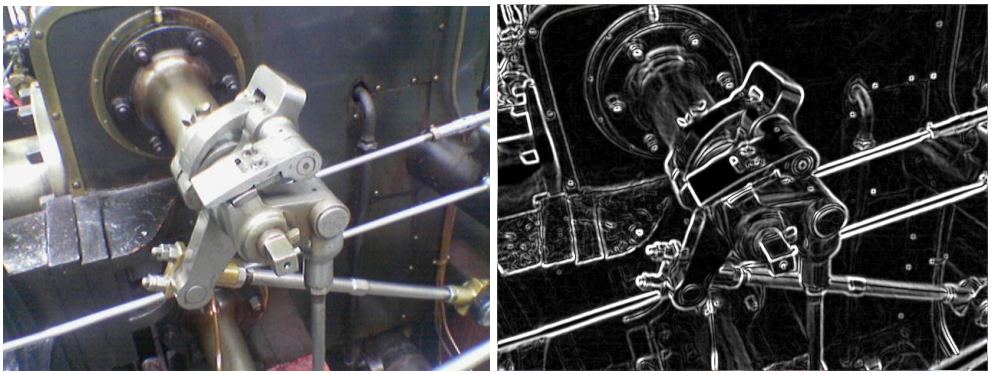


Figure 4.3: Coloured image and a sobel operator applied to that image

4.2.4 Modification of Attention Gate Mechanism

We propose an attention mechanism for image segmentation that integrates low-level features from the encoder block with high-level features from the bottleneck block. This fusion guides the model's attention toward task-specific regions. However, segmentation errors frequently occur along the boundaries of segmented areas. To mitigate these errors, we introduce a novel approach that extracts edge maps from the input data to guide the model's focus toward these critical boundary regions.

Edge Map Generation

We extract edge maps from the input data and process them through a sequence of four residual blocks, progressively enriching the informative features within these maps. This process aims to train the model to identify relevant edges for enhanced attention allocation. The resulting edge maps are scaled to a range of 0 to 1, yielding E_c.

Attention Mechanism

The gating signal g, derived from the subsequent layer with high-level features, undergoes convolution and upsampling. This processed signal is then combined with the skip connection x from the corresponding encoder. After passing through a convolutional layer and applying a sigmoid activation, the result is X_c. X_c is multiplied by E_c, yielding the final scaling coefficient X'_c.

$$X_c = \sigma(U\,(H_{1x1}\,(H_{3x3}(g))) + H_{3x3}(x))$$

$$X'_c = X_c E_c + X_c$$

Equation 4.4: Computation of final scaling coefficient for attention gate

This operation assigns higher importance to pixels with greater values in E_c, which typically represent boundaries. This effectively directs attention toward nuclei edges, aiding in precise segmentation while reducing focus on irrelevant areas.

Finally, the gating signal g is scaled with X'_c,which further refines the attention to essential features, ultimately contributing to more accurate segmentation of nuclei boundaries and minimising focus on irrelevant regions.

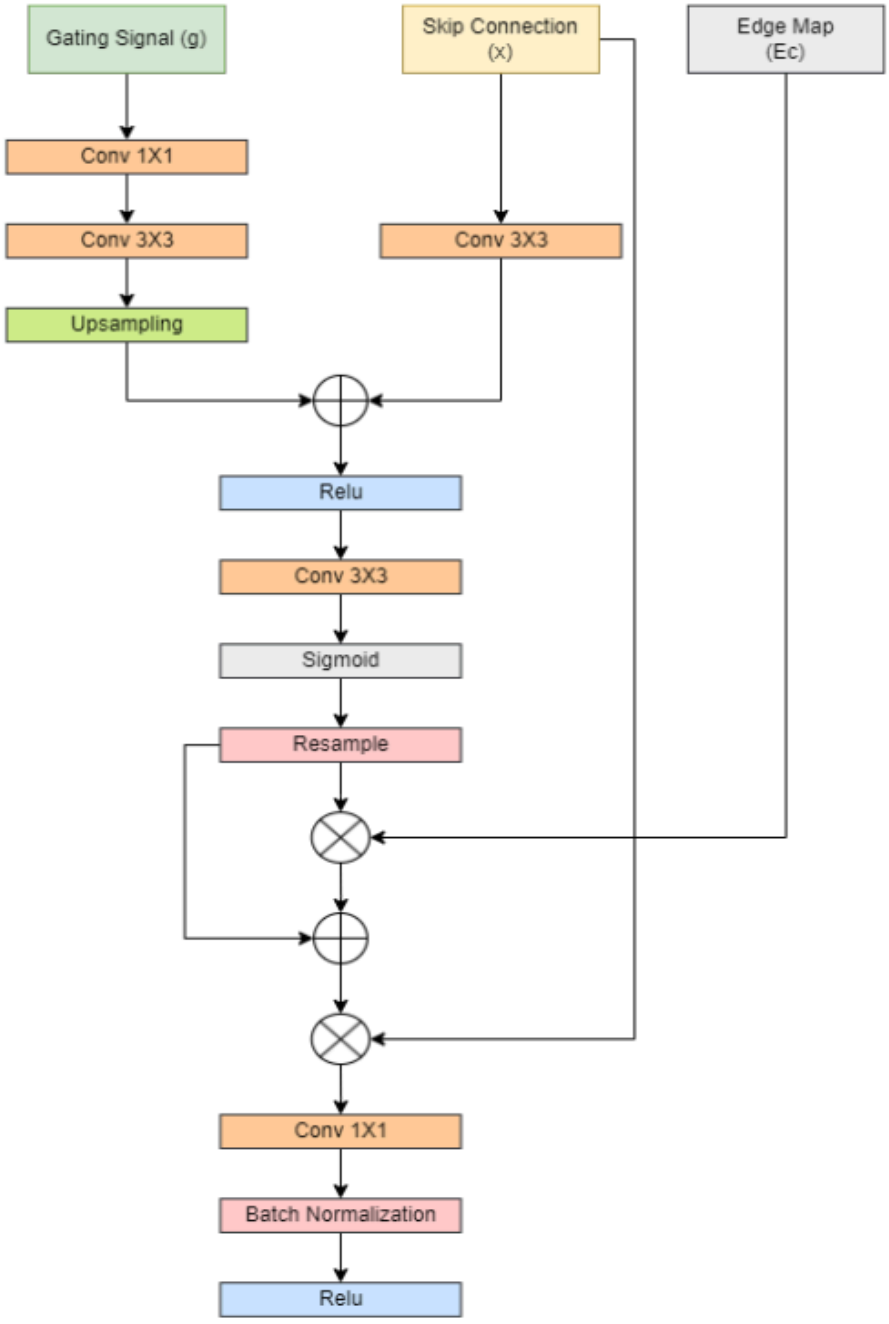


Figure 4.5: Implementation of the Novel Attention Gate Mechanism

4.2.5 Transformer Encoders

Transformer encoders revolutionise computer vision by efficiently capturing global features within images through their self-attention mechanism. Unlike traditional CNNs limited by local receptive fields, Transformer encoders weigh the relevance of each patch across the entire image simultaneously. This holistic view enables them to capture long-range dependencies and semantic relationships, enhancing their effectiveness in tasks requiring comprehensive understanding of visual data.

Moreover, the patch-based processing approach of Transformer encoders treats images as sequences of patches, allowing for the extraction of spatial relationships and structural information across the entire image. By incorporating positional encodings, the model gains spatial context, enabling it to understand the relative positions of patches and capture spatially coherent global features effectively. With these capabilities, Transformer encoders offer a powerful alternative to traditional CNNs, providing a versatile framework for visual representation learning in computer vision tasks.

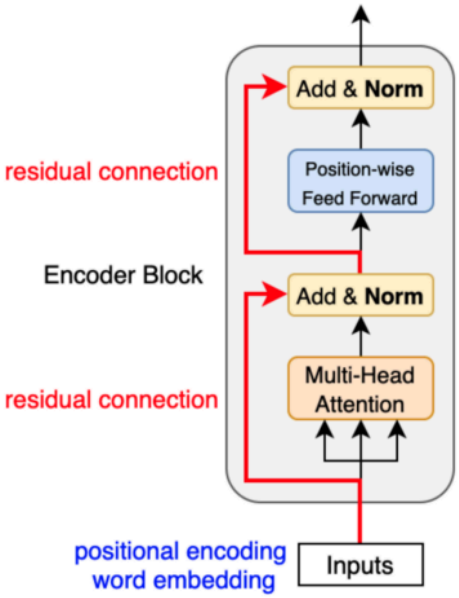


Figure 4.6: Transformer Encoder Block

Overall, the ability of Transformer encoders to capture global features in images offers a promising alternative to traditional CNNs, particularly in tasks requiring holistic understanding and context-aware processing. By leveraging self-attention mechanisms and patch-based processing, Transformer encoders demonstrate remarkable efficacy in capturing long-range dependencies and semantic relationships within images, opening new avenues for visual representation learning and computer vision applications.

4.2.6 Fusion of Global and Local Features

To merge the local features extracted from the convolutional encoder with the global features W obtained from the transformer encoder, we propose a fusion block. We combine the global features W_g and the local features W_l to obtain b. This is then passed through two convolutional layers to determine the significance of each pixel in both the local and global features. The resulting b' is then multiplied with W_g and W_l to yield W_g' and W_l' respectively. These are then added together to produce the final fused encoder output W, which is passed to the attention gate of the corresponding decoder.

$$b' = H_{3x3}(H_{3x3}(W_l + W_g))$$

$$W = W_g b' + W_x b'$$

Equation 4.7: Fusion of Global and Local features

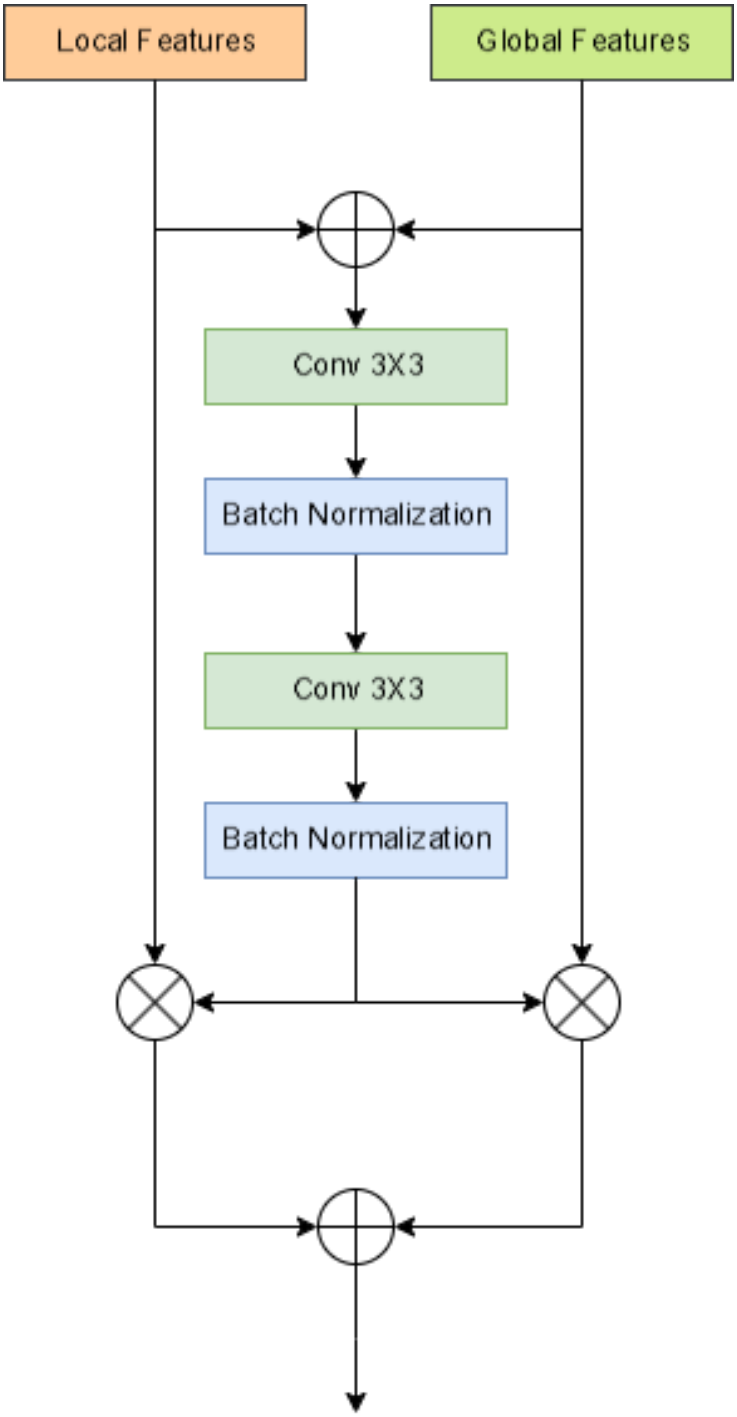


Figure 4.8: Global Local Fusion Module

4.2.7 Proposed Architecture

In our architectural design, we employ a multi-stage approach to process H&E stained images for the task of nucleus segmentation. The input images are simultaneously fed into two distinct encoder branches: the convolutional encoder and the transformer encoder.

The convolutional encoder utilises conventional convolutional neural network (CNN) layers to extract hierarchical features from the input images. This branch employs convolutional filters to capture local patterns and structures, enabling the learning of representations sensitive to spatial variations and texture information. Additionally, in parallel with the convolutional encoder, we incorporate a Sobel filter to compute edge maps. These edge maps offer valuable boundary information that aids in nucleus segmentation. By leveraging the proposed attention mechanism, we effectively integrate the boundary information extracted from the input images into the segmentation process.

On the other hand, the transformer encoder operates on the pixel-wise representation of the input images, treating them as sequences of patches. Leveraging self-attention mechanisms, the transformer encoder captures global dependencies and semantic relationships across the entire image simultaneously. By attending to each patch in relation to all others, the transformer encoder can effectively discern intricate patterns and long-range dependencies, complementing the local features extracted by the convolutional encoder.

These two encoder branches work in tandem, capturing both local and global features of the input images. The outputs of the convolutional encoder, Sobel filter, and transformer encoder are then fused to generate a comprehensive representation of the input images. Finally, the fused representation is processed through additional layers to produce the segmentation mask, wherein each pixel is classified as either belonging to a nucleus or not. This multi-stage approach enables our architecture to leverage both traditional CNN-based feature extraction and the transformative power of transformer encoders, ultimately enhancing the accuracy and robustness of nucleus segmentation in H&E stained images.

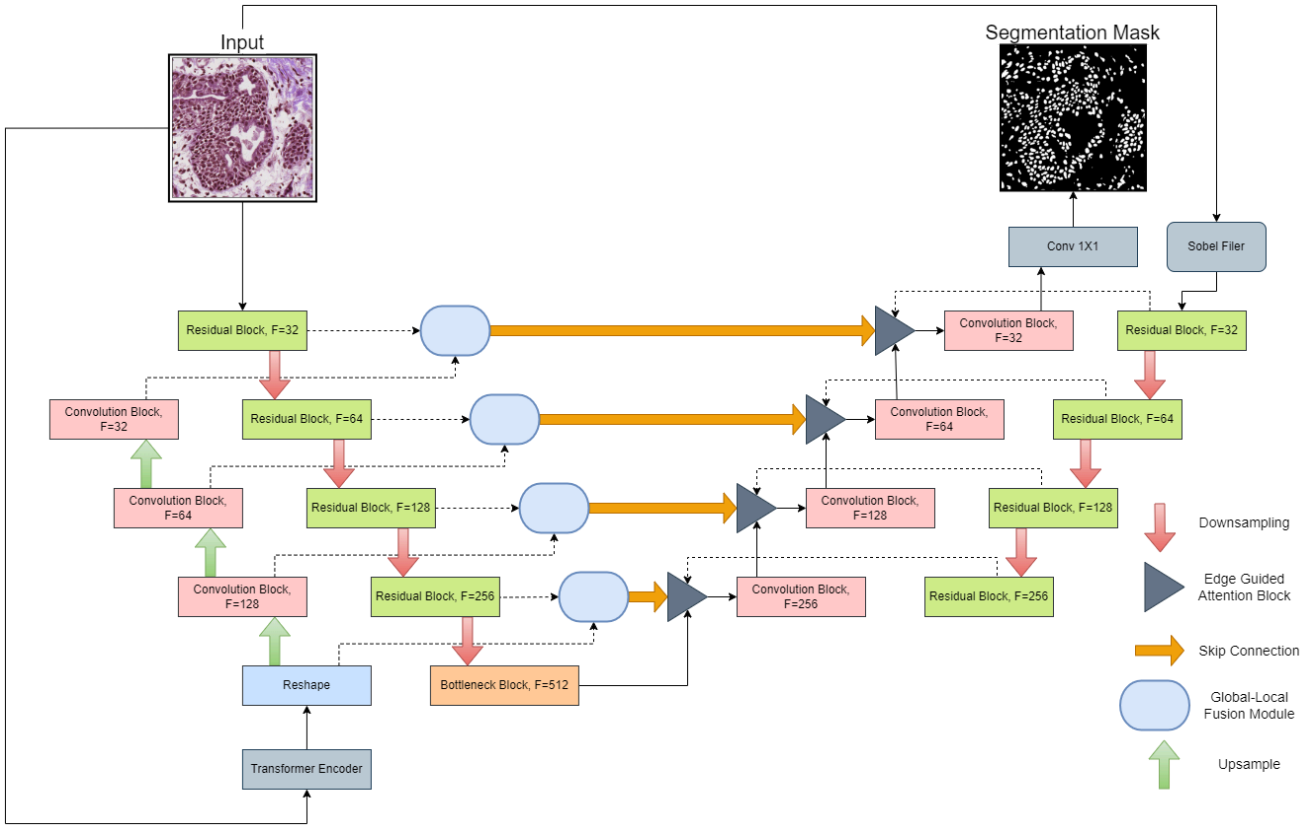


Figure 4.9: Architecture of the Proposed U-Net Model with edge guided attention and Global-Local features

Chapter 5

Result Analysis

5.1 Datasets

Our study utilised two datasets for training the model:

1. **KMC Liver Dataset**: This dataset consists of 194 training images and 4 test images, all of size 512x512 pixels. The KMC Liver dataset was used to train the model for nucleus segmentation tasks.

2. **Monuseg Dataset**: The Monuseg dataset comprises 30 training images with 22,000 nuclear boundary annotations and 7 test images with 7000 nuclear boundaries. All images in the Monuseg dataset are of size 1000x1000 pixels.

5.2 Evaluation Metrics

Our research employs several evaluation metrics to assess the performance of the deep learning models trained on the KMC Liver dataset and the Monuseg dataset. Here t_p, f_p, t_n and f_n represent true-positive, false-positive, true-negative, and false-negative w.r.t to the pixels of the predicted segmentation mask.

The evaluation metrics used in the study include:

1. **Dice Score**: The Dice score is a metric that measures the spatial overlap between the predicted segmentation mask and the ground truth mask. It is calculated as the intersection of the predicted and ground truth masks divided by the average of their areas, multiplied by 2.

$$Dice\ Score\ = \frac{2t_p}{2t_p+f_p+f_n}$$

2. **Intersection over Union (IoU)**: The Intersection over Union metric calculates the ratio of the intersection area between the predicted and ground truth masks to the union area of the two masks. It provides a measure of how well the predicted mask aligns with the ground truth mask.

$$IoU\ Score\ = \frac{t_p}{t_p+f_p+f_n}$$

3. **Accuracy**: The accuracy metric measures the overall correctness of the predicted segmentation compared to the ground truth. It calculates the ratio of correctly classified pixels (true positives and true negatives) to the total number of pixels.

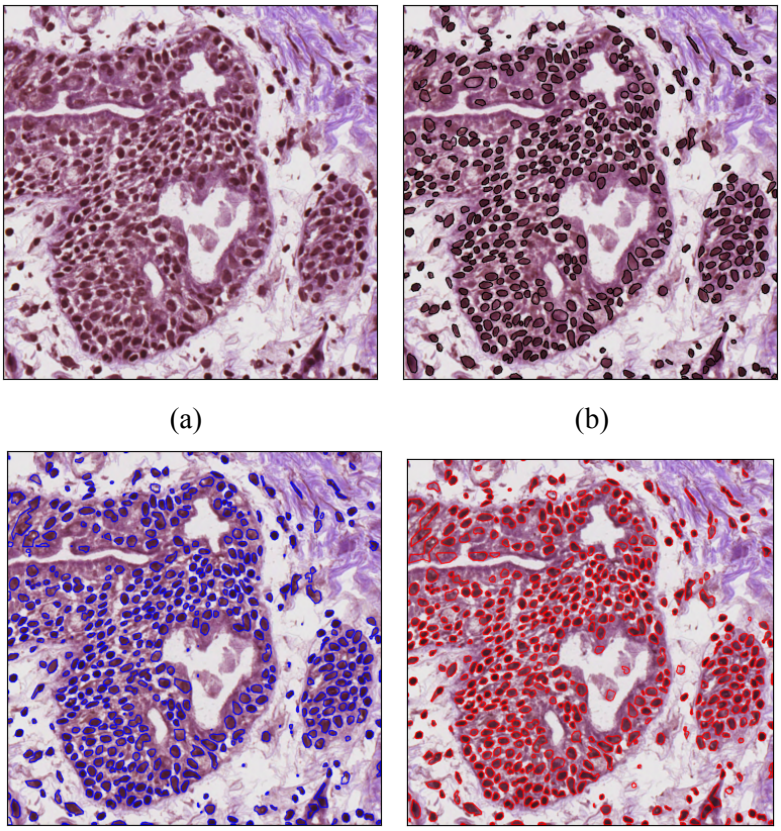
$$Accuracy\ = \frac{t_p+t_n}{t_p+t_n+f_p+f_n}$$

4. **Precision**: Precision is a metric that evaluates the proportion of true positive predictions among all positive predictions made by the model. It is calculated as the ratio of true positive predictions to the sum of true positive and false positive predictions.

$$Precision\ = \frac{t_p}{t_p+f_p}$$

5.3 Segmentation Mask

Segmentation Mask is the output generated by the deep learning model for delineating the boundaries of nuclei in histopathology images. The segmentation mask is a binary image that highlights the regions identified by the model as nuclei within the input image. The resulting segmentation masks are compared with the ground truth masks, which represent the manually annotated or expert-verified boundaries of nuclei in the histopathology images. The segmentation mask serves as a critical output of the deep learning model, enabling precise identification and analysis of nuclei for various medical applications.



(a) (b)

(c) (d)

Figure 5.1: (a) H&E stained tissue image. (b) Ground Truth. (c) Mask predicted by Existing model. (d) Mask predicted by Proposed model.

5.4 Training and Validation

The training process involves feeding the input images into the model, computing the loss, and updating the model parameters through backpropagation to minimise the loss function.

The dataset with ground truth annotations was divided into training, validation, and testing sets. The KMC dataset was split in a 24:6:1 ratio, while the Monuseg dataset was divided into a 5:1:1 ratio. Because the Monuseg dataset is smaller than the KMC dataset, the validation set's proportion had to be reduced.

We used the AdamW optimizer for model training. Our experiments found that a learning rate of 0.001 served as an effective initial setting. To minimise the model's parameter count, we

partitioned the input images into patches measuring 256 x 256 pixels, which helped reduce the number of weights in the model. During testing, we generated masks for each patch and then stitched them together to assess overall performance metrics.

5.5 Results

In Table 5.1 and Table 5.2, we present a comparison of performance metrics for various UNET models trained on both the KMC liver dataset and the Monuseg dataset. The evaluation metrics include Dice score, Intersection over Union (IOU), accuracy, and precision. We assess the performance of the following UNET variants: the base UNET, UNET with edge guided attention mechanism, UNET with global-local encoder, and UNET with both edge-guided attention and global-local encoder.

Model	Dice	IOU	Accuracy	Precision
Existing UNET	85.26	74.38	84.00	98.27
Proposed UNET	**87.27**	**77.51**	**85.69**	**98.50**

Table 5.1 Performance metrics for deep learning models on KMC liver dataset

Model	Dice	IOU	Accuracy	Precision
Existing UNET	81.65	69.05	**82.60**	93.24
Proposed UNET	**83.60**	**71.88**	80.79	**93.66**

Table 5.2 Performance metrics for deep learning models on Monuseg dataset

Fig. 5.2 shows a comparison between the training loss of our proposed model and an existing model using the KMC liver dataset. Initially, our model exhibits a lower training loss compared to the existing one. Over time, our model's loss continues to decrease and eventually converges to 0.00242, which is notably lower than the existing model's convergence point of 0.00366 as shown in Table 5.3

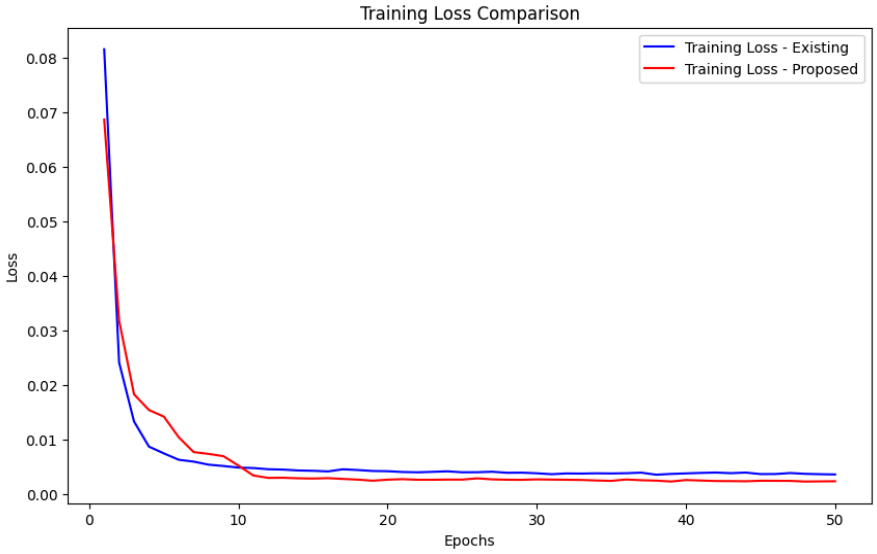


Figure 5.2 Training Loss progression for KMC Liver dataset

Iteration	Existing	Proposed
5	0.007516	0.014248
10	0.004923	0.005286
15	0.004336	0.002919
20	0.004256	0.002706
25	0.004060	0.002720
30	0.003881	0.002766
35	0.003847	0.002491
40	0.003858	0.002644
45	0.003732	0.002513
50	0.003665	0.002421

Table 5.3 Training Loss progression for KMC Liver dataset

Fig. 5.3 shows a comparison between the validation loss of our proposed model and an existing model using the KMC liver dataset. In the initial epochs of the training phase, our model's validation loss is higher compared to that of the existing model. However, gradually, we observe lower values, signifying improved results on new input over time, as shown in Table 5.4

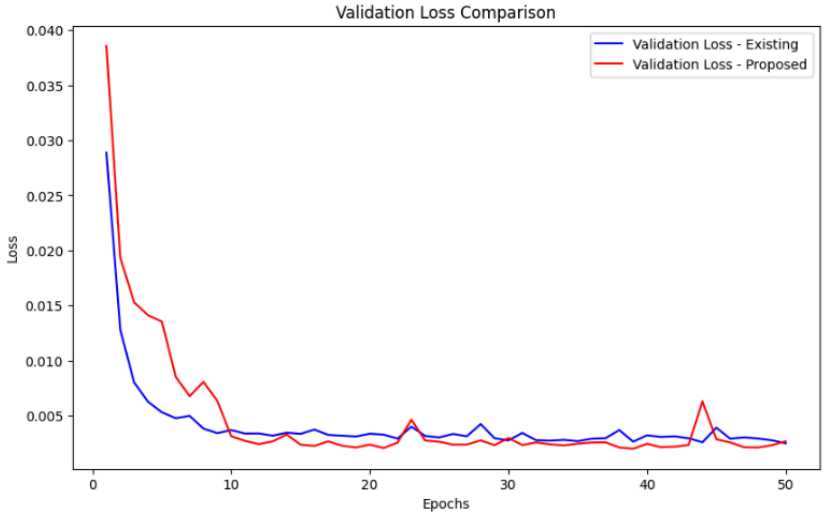


Figure 5.3: Validation Loss progression for KMC Liver dataset

Iteration	Existing	Proposed
5	0.017645	0.013528
10	0.003673	0.003124
15	0.003332	0.002356
20	0.003347	0.002359
25	0.003009	0.002622
30	0.002744	0.002960
35	0.002670	0.002445
40	0.003203	0.002434
45	0.003900	0.002848
50	0.002480	0.002661

Table 5.4: Validation Loss progression for KMC Liver dataset

In Figure 5.4 we present a comparison of the F1 Scores achieved by our proposed model with those of an existing model, utilizing the KMC liver dataset. Given the higher training loss observed at the outset of training our model, correspondingly lower F1 Scores are evident in the initial epochs. However, as training progresses, our model demonstrates an improvement in F1 Scores. Ultimately, our model achieves a final F1 score of 0.852116, outperforming the existing model which attained a score of 0.844240, as detailed in Table 5.5.

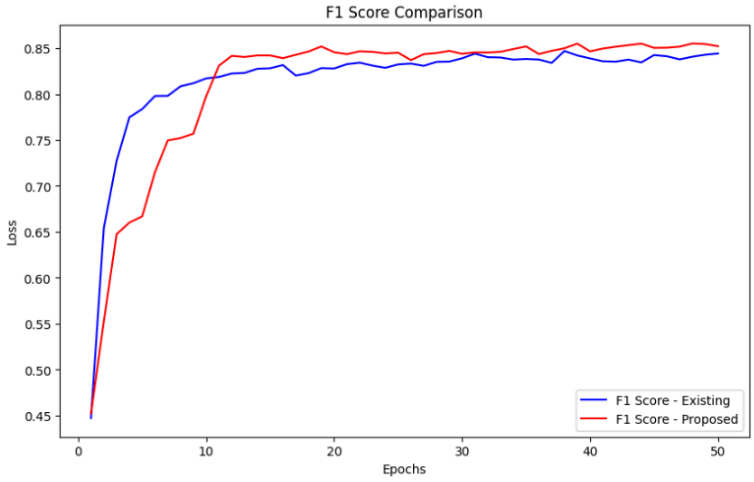


Figure 5.4: F1 Score progression for KMC Liver dataset

Iteration	Existing	Proposed
5	0.783715	0.666940
10	0.817152	0.797742
15	0.828144	0.842108
20	0.827895	0.845555
25	0.832375	0.845104
30	0.838988	0.843871
35	0.838260	0.851912
40	0.838884	0.846507
45	0.842489	0.850380
50	0.844240	0.852116

Table 5.5: F1 Score progression for KMC Liver dataset

Fig. 5.5 illustrates a comparison between the training loss curves of our proposed model and an existing model, utilising the Monuseg dataset. In our experiments with this dataset, we observed that optimal loss values were achieved after approximately 5 epochs, as depicted in Table 5.6. Notably, our model obtained a final loss value of 0.012534, whereas the existing model yielded a slightly higher value of 0.013317.

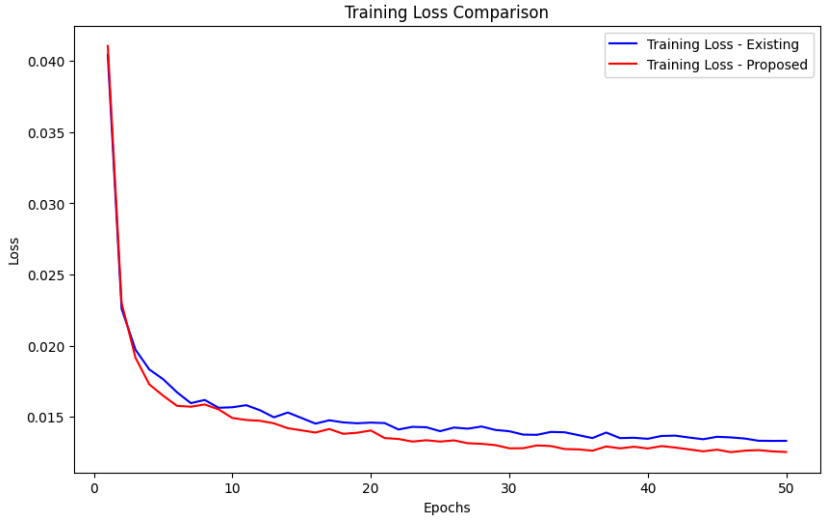


Figure 5.5: Training Loss progression for Monuseg dataset

Iteration	Existing	Proposed
5	0.017645	0.016485
10	0.015675	0.014920
15	0.014921	0.014051
20	0.014598	0.014044
25	0.013995	0.013263
30	0.013990	0.012790
35	0.013715	0.012716
40	0.013459	0.012779
45	0.013600	0.012695
50	0.013317	0.012534

Table 5.6: Training Loss progression for Monuseg dataset

Figure 5.6 displays a contrast of the validation loss between our proposed model and an existing model applied to the Monuseg dataset. Noteworthy is the occurrence of spikes in validation loss for both models, likely stemming from the dataset's relatively modest size for the validation split. Despite grappling with this obstacle, our model notably outperforms the existing model on the validation split, as substantiated by the findings detailed in Table 5.7.

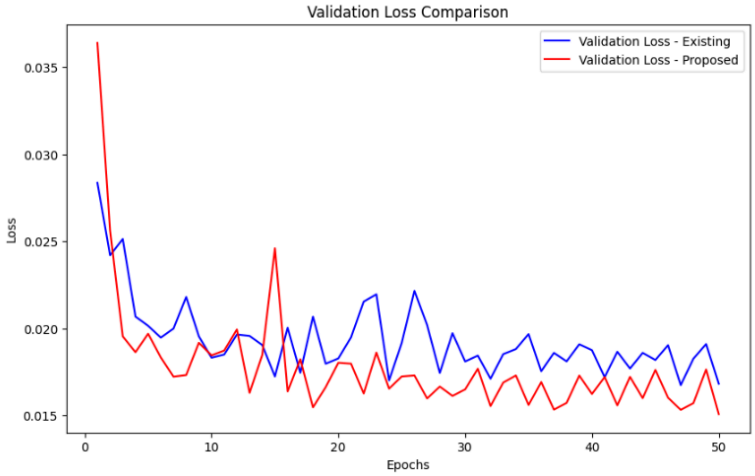


Figure 5.6: Validation Loss progression for Monuseg dataset

Iteration	Existing	Proposed
5	0.020152	0.019697
10	0.018317	0.018454
15	0.017239	0.024607
20	0.018278	0.018028
25	0.019173	0.017243
30	0.018103	0.016499
35	0.019677	0.015603
40	0.018749	0.016238
45	0.018186	0.017617
50	0.016824	0.015076

Table 5.7 Validation Loss progression for Monuseg dataset

In Figure 5.7, we compare F1 Scores of our proposed model with an existing one, using the Monuseg dataset. Despite initially lower scores due to higher training loss, our model reaches its optimal value earlier than the existing one, as in Table 5.8. Our model achieves an F1 score of 0.825329, surpassing the existing model's 0.821524, showcasing superior training performance.

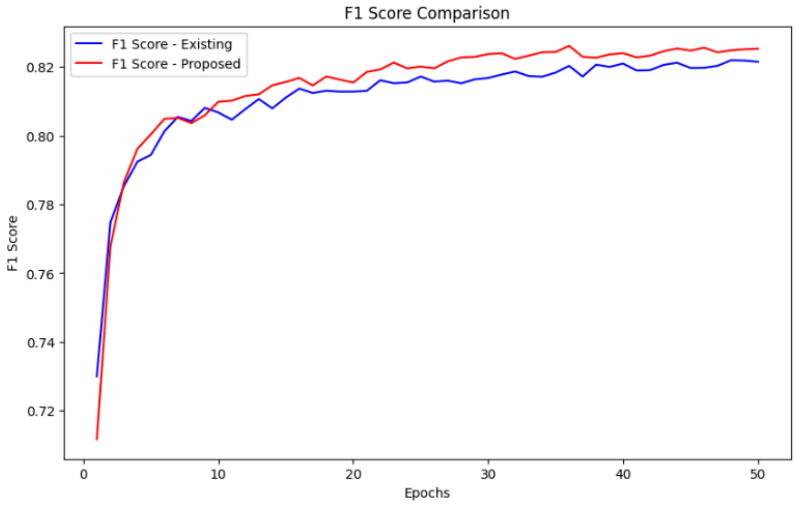


Figure 5.7 F1 Score progression for Monuseg dataset

Iteration	Existing	Proposed
5	0.794418	0.800398
10	0.806771	0.809877
15	0.811101	0.815626
20	0.812827	0.815466
25	0.817243	0.820130
30	0.816816	0.823759
35	0.818373	0.824369
40	0.821007	0.824032
45	0.819720	0.824786
50	0.821524	0.825329

Table 5.8 F1 Score progression for Monuseg dataset

Chapter 6

Conclusion and Future Work

In conclusion, our research introduces a novel approach for nucleus segmentation in histopathology images by incorporating edge-guided attention and global-local encoders into the UNET architecture. By leveraging algorithmic enhancements to improve segmentation accuracy, the proposed methodology demonstrates promising results in accurately delineating nucleus boundaries in challenging histopathology images.

Through extensive experimentation on the KMC Liver dataset and the Monuseg dataset, the research evaluates the performance of the proposed model against existing UNET variants. The comparison of metrics such as Dice score, IoU, accuracy, and precision showcases the effectiveness of the proposed methodology in enhancing nucleus segmentation accuracy and reducing errors along edges.

The research contributes to the advancement of automated nucleus segmentation in histopathology images by addressing limitations in traditional UNET models and introducing innovative techniques to extract edge features and capture global dependencies. By refining the segmentation mask through edge-guided attention and global-local encoders, the proposed model offers a more robust and accurate solution for automated nucleus segmentation tasks.

Future research directions stemming from this study include:

1. **Advanced Edge Extraction Techniques**: In addition to Canny, Prewitt, and Roberts filters, we plan to explore cutting-edge edge detection algorithms and techniques to further enhance the extraction of edge features from input images. By leveraging state-of-the-art edge detection methods, we aim to improve the model's ability to capture intricate details and boundaries, ultimately enhancing nucleus segmentation accuracy.
2. **Exploration of Novel Transformer Architectures**: Beyond the Swin transformer, we will investigate other transformer encoders with unique capabilities for capturing global dependencies and long-range interactions in histopathology images. By exploring a diverse range of transformer architectures, we aim to identify the most effective model

for enhancing nucleus segmentation accuracy and performance.

3. **Multi-Modal Fusion**: Investigating the integration of multi-modal data sources, such as additional staining techniques or imaging modalities, to enhance the model's ability to segment nuclei in diverse histopathology images.
4. **Scalability and Efficiency**: Optimising the model for scalability and efficiency to handle larger datasets and real-time processing requirements in clinical settings.

By pursuing these avenues of future work, researchers can further advance the field of automated nucleus segmentation in histopathology images, ultimately contributing to improved diagnostic accuracy, disease prognosis, and treatment planning in medical practice.

Bibliography

[1] Shyam Lal, Devikalyan Das, Kumar Alabhya, Anirudh Kanfade, Aman Kumar, Jyoti Kini, NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images, Computers in Biology and Medicine, Volume 128, 2021, 104075, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.104075.

[2] Oktay, Ozan Schlemper, Jo Folgoc, Loic Lee, Matthew Heinrich, Mattias Misawa, Kazunari Mori, Kensaku McDonagh, Steven Hammerla, Nils Kainz, Bernhard Glocker, Ben Rueckert, Daniel. (2018). Attention UNet: Learning Where to Look for the Pancreas.

[3] Z. Zhang, Q. Liu and Y. Wang, "Road Extraction by Deep Residual UNet," in IEEE Geoscience and Remote Sensing Letters, vol. 15, no. 5, pp. 749-753, May 2018, doi: 10.1109/LGRS.2018.2802944. keywords: Roads;Training;Neural networks;Remote sensing;Semantics;Image segmentation;Feature extraction;Convolutional neural network;deep residual U-Net;road extraction,

[4] Ronneberger, Olaf Fischer, Philipp Brox, Thomas. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351. 234-241. 10.1007/978-3-319-24574-4 28.

[5] Vaswani, Ashish Shazeer, Noam Parmar, Niki Uszkoreit, Jakob Jones, Llion Gomez, Aidan Kaiser, Lukasz Polosukhin, Illia. (2017). Attention Is All You Need.

[6] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conf. on computer vision and pattern recognition, 2015, pp. 1–9.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conf. on computer vision and pattern recognition, 2015, pp. 3431–3440.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[9] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1857–1866.

[10] Greff, K., Srivastava, R.K., Schmidhuber, J.: Highway and residual networks learn unrolled iterative estimation. arXiv preprint arXiv:1612.07771 (2016)

[11] Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for

volumetric medical image segmentation. In: 3D Vision. pp. 565–571. IEEE (2016)

[12] Ypsilantis, P.P., Montana, G.: Learning what to look in chest X-rays with a recurrent visual attention model. arXiv preprint arXiv:1701.06452 (2017)

[13] Mukku Nisanth Kartheek, Munaga V.N.K. Prasad, and Raju Bhukya. 2023. Local triangular patterns: novel handcrafted feature descriptors for facial expression recognition. Int. J. Biometrics 15, 2 (2023), 194–211. https://doi.org/10.1504/ijbm.2023.129224

[14] Gugulothu P, Bhukya R. Coot-Lion optimized deep learning algorithm for COVID-19 point mutation rate prediction using genome sequences. Comput Methods Biomech Biomed Engin. 2023 Sep 5:1-20. doi: 10.1080/10255842.2023.2244109.

[15] Dasari CM, Bhukya R. Explainable deep neural networks for novel viral genome prediction. Appl Intell (Dordr). 2022;52(3):3002-3017. doi: 10.1007/s10489-021-02572-3.

[16] Raju Bhukya, Archana Kumari, Chandra Mohan Dasari, and Santhosh Amilpur. 2022. An attention-based hybrid deep neural networks for accurate identification of transcription factor binding sites. Neural Comput. Appl. 34, 21 (Nov 2022), 19051–19060. https://doi.org/10.1007/s00521- 022-07502-z

Acknowledgement

It is our privilege and honour to express our gratitude to all the respected personalities who have guided us, inspired us and supported us in the smooth and successful completion of our project.

We thank our beloved faculty supervisor Dr. Raju Bhukya, Associate Professor, Department of Computer Science and Engineering(CSE), National Institute of Technology, Warangal, for his persistent supervision, guidance, suggestions and encouragement during this project. He has motivated us during the low times and gave us courage to move ahead positively.

We are grateful to Dr.R.Padmavathy, Head of the Department, Computer Science and Engineering, National Institute of Technology, Warangal for her moral support to carry out this project. We are also thankful to the Project Evaluation Committee, for their strenuous efforts to evaluate our projects. We wish to thank all the staff members in the department for their kind cooperation and support given throughout our project work. We are thankful to all of our colleague friends who have given valuable suggestions and help in all stages of the development of the project.

Astitiva Verma (207111)

Adarsh Rao (207102)

Akula Pawan Kalyan (207105)