Heliyon 10 (2024) e38410

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

HI-Net: A novel histopathologic image segmentation model for metastatic breast cancer via lightweight dataset construction

Fengze Li [a,b], Jieming Ma [b,*], Tianxi Wen [b], Zhongbei Tian [a], Hai-Ning Liang [b]

[a] University of Liverpool, Liverpool, UK
[b] Xi'an Jiaotong-Liverpool University, Suzhou, China

ARTICLE INFO

Dataset link: https://data.mendeley.com/datasets/hd3kw5tt8f/1

Keywords:
Breast cancer
Histopathology slide
Deep learning
Medical imaging
Intelligent diagnosis

ABSTRACT

Since 2020, breast cancer has remained the most prevalent cancer worldwide and the World Health Organisation projects significant increases by 2040, with new cases expected to exceed 3 million annually (a 40% increase) and deaths to surpass 1 million (a 50% increase), highlighting the urgent need for advancements in detection and treatment. Current detection of metastasis is highly dependent on labour-intensive and error-prone pathological examination of large-scale biotissue. Given the high-resolution (100,000 × 100,000 gigapixels) but limited quantity of open-source pathological slide datasets, existing deep learning models face preprocessing challenges. This paper introduces HI-Net, a high-speed panoramic feature-extraction pyramid network for rapid and accurate detection of metastatic breast cancer, balancing panoramic segmentation and local attention. Additionally, a lightweight pathological slide dataset optimised for 512 x 512-pixel resolution, derived from downsampled and reassembled competitive datasets, accelerates training and reduces computational costs. HI-Net demonstrates superior performance on existing medical imaging competition datasets and our lightweight dataset, evidencing its effectiveness across datasets and potential for contributing to the generalisation of intelligent diagnostics.

1. Introduction

1.1. Current developments in breast cancer and pathology

The International Agency for Research on Cancer (IARC), through the World Health Organisation (WHO), has stated that since 2020, breast cancer has been the most commonly diagnosed type of cancer in the world [1,2], with more than 2.26 million cases. By 2022, this number had risen slightly to 2.31 million cases, making it the second most common cancer, but it is still the leading cause of cancer death in women [3]. The WHO highlights the consistent global presence of breast cancer and its status as the most common cancer in women in 157 countries in 2022, underlining its significant health impact [4]. Future projections indicate a substantial increase in breast cancer cases and deaths by 2040, with estimates suggesting more than 3 million new cases annually, a 40% rise, and over 1 million deaths, a 50% increase. This anticipated growth emphasizes the urgent need for advancements in detection and treatment, particularly in transitioning countries where the disease's impact is intensifying [5].

One of the critical professional goals of pathology is to provide patients with a definitive diagnosis of disease. Therefore, an accurate, traceable, and experimentally reproducible pathological diagnosis is essential to advance intelligent medical treatment [6].

* Corresponding author.
E-mail address: JiemingMa@xjtlu.edu.cn (J. Ma).

https://doi.org/10.1016/j.heliyon.2024.e38410
Received 2 July 2023; Received in revised form 22 September 2024; Accepted 24 September 2024

Available online 27 September 2024

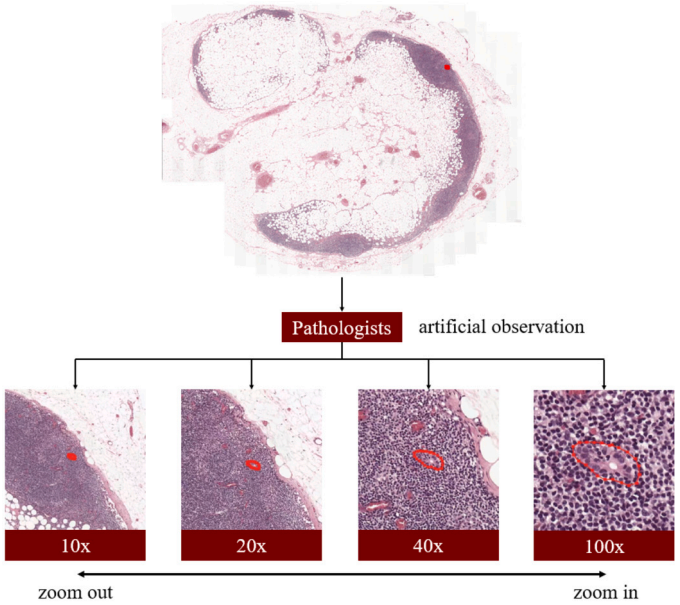

Fig. 1. Slides of tissue patches at different magnifications (the area circled in red highlights the ground truth).

However, in actual clinical practice of breast cancer diagnosis, pathologists are highly dependent on the pathological evaluation of tissue samples under the microscope for a large part of their diagnosis [7]. Although relatively efficient workflows prevail in pathology laboratories as medical equipment is updated, the display of whole slide images (WSIs) [8], on which pathologists rely to observe the morphology of cells, still takes close to 10 minutes. In addition, pathologists are inevitably confronted with the limitations imposed by microscopic images, where the lack of standardisation and empiricism lead to diagnostic errors that can be detrimental to the subsequent treatment of the patient and the physician's maximum cognitive workload [9]. As shown in Fig. 1, this is a pathological tissue patch at various magnifications of nearly 200,000 x 260,000 pixels. Pathologists tend to obtain approximate information at low resolution and then magnify it to more than 40x to fully visualize the relatively accurate microscopic state of the tissue cells and provide a final diagnosis.

The challenges faced by pathologists in the clinical practice of breast cancer diagnosis go far beyond the limitations of evaluating tissue samples under the microscope. First, the inherent heterogeneity of breast cancer tissue can complicate the interpretation of histopathological features, potentially leading to variability in diagnosis among pathologists. Studies have highlighted the subjective nature of histopathological interpretation, which may result in discrepancies in diagnosis, grading, and staging of breast cancer [10]. The impact of such variability can be mitigated through the use of standardized criteria and training, yet it underscores the need for additional diagnostic support tools. Second, the high volume of cases that require evaluation can strain resources and contribute to diagnostic delays. As the prevalence of breast cancer rises, the workload for pathologists intensifies, increasing the risk of fatigue-related errors [11]. Moreover, advancements in breast cancer treatments demand more precise and detailed tissue analyses, requiring pathologists to identify specific biomarkers and molecular subtypes of cancer. This complexity necessitates continuous education and adaptation to new diagnostic criteria and technologies [12]. The accuracy of these assessments is critical for guiding personalized treatment plans and improving patient outcomes.

Using published data from the United States as an example, there are approximately 21,000 active pathologists today, and the trend for future growth in numbers is pessimistic as there have been nearly 20% fewer pathologists in North America in the last 15 years [13]. Deep learning-based methods had already gained initial application in digital pathology prior to the use of WSIs. However, these early studies raised a similar nuisance: the datasets required tedious preprocessing, in these cases being image datasets consisting mainly of previously selected regions of interest (ROIs) [14]. These preprocessing methods mentioned in earlier studies require pathologists to select regions of interest in advance or manually attach weights to the regions, leading to significant labour and time costs. Moreover, such processes require pathologists to have interdisciplinary expertise, not to mention the fact that pathologists can disagree with each other [15], making them technically challenging to implement on a large scale in clinical workflows in the laboratory [16]. The introduction of WSIs into pathological assessment has revolutionized the field, significantly impacting the workflow in pathology laboratories, especially concerning breast cancer diagnosis. WSIs allow for the digitisation of entire tissue slides at high resolution, facilitating a more detailed and comprehensive examination of tissue samples. This technological advancement supports remote diagnostics, collaborative assessments, and enhanced educational opportunities by enabling pathologists to share and discuss complex cases with peers worldwide [17]. The benefits of WSIs in pathological assessment are manifold. They enable a level of detail and precision previously unattainable with traditional microscopy, improving the accuracy of diagnoses. Moreover, the digitisation of slides offers improved storage and retrieval efficiency, reducing physical space requirements and allowing for easier access to historical cases [18]. While it will inevitably be a challenge for artificial intelligence (AI) models to handle WSIs with

huge resolutions of millions of pixels, its capability to implement remote pathology telework and the possibility to build large digital databases of pathology section images make it exceptionally important to develop deep learning models based on WSIs.

1.2. Deep learning-based approaches for WSIs

The advancement of deep convolutional neural networks is needless to mention, and the introduction of the transformer architecture in 2017 [19] has opened up more possibilities for the efficient processing of medical images and the widespread availability of intelligent medical infrastructures. Both transfer learning-based cancer detection [15,20–22] and transformer-based improved medical image processing networks [23–25], and the continued progress of other branching tasks [26–31], follow a similar pattern. Additionally, the Tubule-U-Net proposed by Tekin et al. in 2023 [32] was tested on five different WSIs, all achieving excellent breast cancer tumour grading results, confirming the dominance of deep learning-based approaches in the detection and localisation of WSIs for cancer metastasis. Models for WSIs are typically trained by extracting small image patches from the whole slide image at the preprocessing step in a split and multi-step manner. These methods are accompanied by the use of a sliding window, starting with sparse sampling at the finest magnification and finishing with performing classification prediction. These methods have demonstrated remarkable success in improving diagnostic accuracy, particularly in the analysis of WSIs, by enabling the detailed examination of tissue samples at various scales and complexities. However, practical implementation faces several challenges, including the need for extensively annotated datasets for training, computational resource requirements, and the integration of these technologies into existing clinical workflows without disrupting standard pathological practices. Additionally, there is a growing need for interpretability and explainability in deep learning models to gain the trust of medical professionals and to comply with regulatory standards.

Based on the characteristics mentioned above brought about by the disease itself and the detection approaches, panoptic segmentation [28] and accurate local information acquisition are the characteristics that potential target models are expected to possess. The ability of the two tasks to be combined and the ability to achieve high accuracy is challenging. Still, if a network can be designed to perform the above task properly, it would be an auspicious approach for megapixel-resolution WSIs. Specifically, a single network can be designed to perform global segmentation and local feature extraction. The backbone of Feature Pyramid Network (FPN)-based networks [28] is a promising choice. By combining lateral connections with top-down connections, feature maps of different resolutions can be obtained, and they all contain the semantic information of the original deepest feature maps. Although the semantic features of the higher layers are passed down through lateral connections, only the semantic information of the feature pyramid is enhanced. It is unavoidable to go through many layers of the network in the middle, so much of the object information at the bottom has been lost. Furthermore, although the lateral connections can address the limitation of significantly increased inference time for each layer of the characterised image pyramid, the generic process of the semantic branch to upsample is relatively redundant.

1.3. Solutions proposed in this paper

In addressing the challenge of pathological image diagnosis, particularly with a focus on achieving precise localisation within histopathologic images, the conventional approach of utilizing multiple lateral connections for sampling may not be essential. This realisation comes from the observation that, from the acquisition of the original image to the derivation of the bottom feature map, a significant portion of the critical object information is lost [33]. This loss is primarily attributed to the information being processed through several layers within neural networks. During these top-down information transfers, the process does not account for the variability in the importance of different features, leading to a non-optimised distribution of weights across the network layers. Therefore, we proposed a novel feature pyramid-based network for histopathologic image segmentation called HI-Net that takes an attention mechanism [34,35] accompanied by weight consideration as a promising solution. Moreover, we proposed a shortcut module within the semantic branch, allowing the input to skip multiple convolutional layers and regularisation operations directly.

Moreover, the task of developing deep learning models for WSIs suffers from two hard-to-avoid issues at this stage: one is cumbersome preprocessing, and the other is the lack of open-source datasets with nonuniform formats. Most existing works tend to simply segment the original image into small patches of the same size [36]. Some works have also suggested that pre-sampling a set of patches in each slide can improve efficiency [7], but this will undoubtedly limit the breadth of patches seen during training. For the limited dataset of open-source WSIs, it has been argued that a data enhancement approach based on geometric reconstruction is an effective solution [37]. However, our proposed lightweight histopathology slide dataset integrates multiple competition datasets. It features a much larger number of images than the competition datasets and has been annotated to better serve the deep learning network for this task. It is also segmented into lightweight resolution slides with uniform colour gamut and annotation form after multiple downsampling while maintaining quality. Thus, this article features the following main contributions:

1. This paper introduces HI-Net, a novel network for histopathology image segmentation with an innovative top-down attention (TDA) module that ensures precise semantic and object information transfer, reducing information loss significantly.
2. Along with HI-Net, a streamlined shortcut module optimises the pyramid architecture's lateral connections, accelerating medical image processing by simplifying convolution and regularisation operations.
3. This paper presents a first-of-its-kind downsampled lightweight slide dataset for histopathologic segmentation, containing 7927 images, surpassing existing datasets in scale and providing a new benchmark for research.
4. HI-Net demonstrates superior performance in both new and existing datasets, outperforming advanced models. The shortcut module markedly improves training speed, showcasing efficiency and accuracy improvements in medical image analysis.

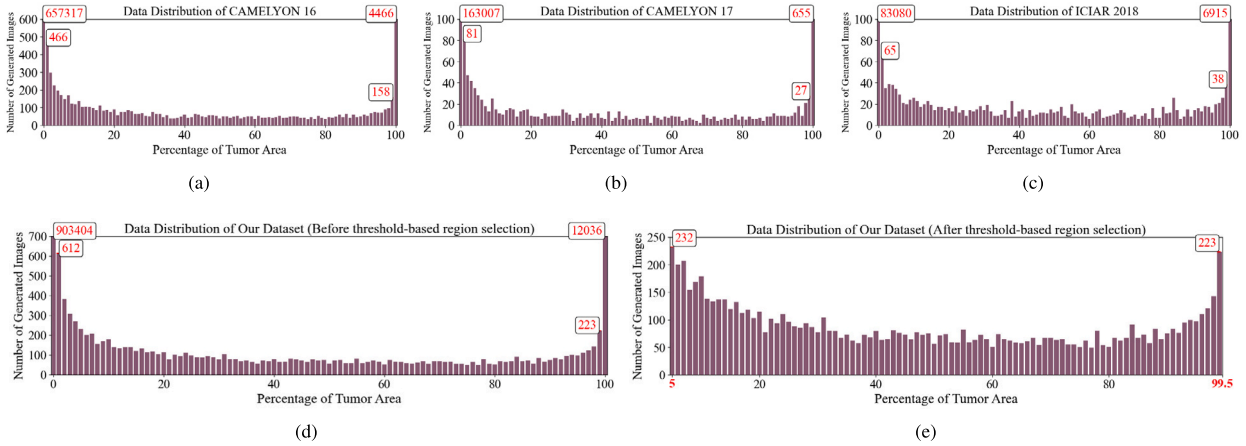

Fig. 2. Overview of the distribution analysis of tumor size percentages across different datasets and the effects of threshold-based region selection. Sub-figures (a), (b), and (c) illustrate the distribution of tumor size percentage images in the comparison datasets CAMELYON 16, CAMELYON 17, and ICIAR 2018, respectively. Sub-figures (d) and (e) present the distributions of the proposed dataset before and after applying threshold-based region selection. The optimized distribution in sub-figure (e) shows a more uniform spread of tumor sizes, achieving a more balanced and effective dataset for subsequent analysis.

2. The proposed lightweight dataset

Despite the limited quantity of gigapixel WSIs available, each contains numerous lesion areas of interest for segmentation. This paradox highlights the rich, yet under-exploited, potential of existing WSI resources for detailed pathological analysis. The challenge of gigapixel WSI's large image resolution, which prevents direct neural network input, is addressed through a standardised sliding window approach. This method simplifies the otherwise cumbersome preprocessing required for each study, significantly reducing the barrier to entry for high-resolution medical image analysis. Variability in annotation methodologies across datasets poses a significant challenge for consistent analysis. Therefore, given a gigapixel pathology image (shown in Fig. 1), the deep learning-based models aim to assist pathologists in classifying whether the image contains a tumor or not and in localizing the tumor. This section of the paper describes the process of constructing our proposed lightweight dataset, where the CAMELYON 16 [38], 17 [39] and ICIAR 2018 [40] datasets provided by the competition organizers are considered as the benchmark, and the Automated Slide Analysis Platform (ASAP) [41] is selected as the tool for ultra-high resolution image display and manual annotation.

The CAMELYON 16 dataset collected from Radboud UMC and UMC Utrecht consists of 400 WSIs, divided into 270 for training and 130 for testing. While CAMELYON 17 follows the data of the previous generation and uses it as lesion-level training data, it expands the dataset to 1000 slides. The ground truth of the CAMELYON series was obtained by professional pathologists for the delineation of metastatic cancer areas on WSIs. The data is provided in two formats: an XML file containing annotated contours of cancer metastasis locations and a WSI bidirectional mask indicating the location of cancer metastasis. The main difference between ICIAR 2018 and the CAMELYON series sits in its relatively small resolution, only approaching 2000 x 1500 pixels. It contains 400 WSIs in which four different cancer states are labelled, with the distinction between benign and cancerous.

Thus, several obvious drawbacks need to be addressed, shown as the bullet points below, and according to the existing obstacles, the lightweight dataset proposed in the paper provides corresponding optimisations.

1. Despite the limited quantity of gigapixel WSIs available, each contains numerous lesion areas of interest for segmentation. This paradox highlights the rich, yet underexploited, potential of existing WSI resources for detailed pathological analysis.
2. The challenge of gigapixel WSI's large image resolution, which prevents direct neural network input, is addressed through a standardised sliding window approach. This method simplifies the otherwise cumbersome preprocessing required for each study, significantly reducing the barrier to entry for high-resolution medical image analysis.
3. Variability in annotation methodologies across datasets poses a significant challenge for consistent analysis. Our dataset introduces optimisations for annotation consistency and colour space uniformity.

2.1. Downsampling and threshold-based region selection

Our proposed lightweight histopathology breast cancer slide image dataset is based on the three competition datasets mentioned above to downsample the CAMELYON series dataset 4 times according to the original WSI map resolution size, but not for ICIAR 2018. The downsampled images of the new dataset were then slide-intercepted as 512 $\times$ 512 pixel patch slides. To save computational costs and to integrate the cumbersome data preprocessing of WSIs in other studies, the article opted to focus the analysis on the regions on the slides that were most likely to contain cancer metastases. Identifying tissues within WSIs to exclude background white space and removing WSIs where the area of the cancer region is disproportionately large for the entire image were the goals of the dataset creation, shown as Fig. 2. Therefore, we introduced a threshold-based segmentation method to automatically detect eligible segmented slides, where the downsampled images were shifted from RGB colour space to V colour space, and then the installation of

Table 1
Number of slides in the benchmark datasets and ours.

Dataset	Training set	Test set	Validation set	Total
CAMELYON 16 [38]	270	130	-	400
CAMELYON 17 [39]	500	500	-	1000
ICIAR 2018 [40]	300	100	-	400
Ours	**6345**	**800**	**782**	**7927**

the Otsu algorithm [42] calculated the optimal threshold for each channel, allowing the mask generation for the **H** and **S** channels to be combined into the final masked image. This optimisation process was guided by the goal of achieving the highest possible diagnostic precision within our dataset.

Addressing the exclusion criteria for slides with tumor presence below 5% or above 95%, our decision was driven by a rigorous evaluation of data balance and training efficiency. Initially, including slides with very high tumor coverage (>95%) appears to introduce valuable data for feature extraction. However, our analyses indicated that retaining these slides could lead to a dataset imbalance, with an overrepresentation of tumor-rich images. Such imbalance would skew the model's training towards predominantly identifying tumor regions, potentially compromising its ability to accurately distinguish between tumor and non-tumor tissues in a balanced manner. To mitigate this risk and enhance the model's robustness, we chose to exclude slides with tumor coverage beyond 99.5%. Moreover, our investigation revealed that slides with minimal tumor presence (<5%) are more abundant than those with higher tumor coverage. Considering the relatively lower informational value of such minimally affected slides for training purposes, we aimed to further balance the dataset and improve training efficiency by excluding these extremes. To validate this approach, we conducted extensive quantitative experiments across various datasets, including our newly established dataset, to meticulously assess the distribution of tumor area ratios. Shown as Fig. 2(e), these experiments confirmed that our threshold settings (excluding slides with tumor presence below 5% and above 99.5%) are well-justified, striking a balance between dataset representativeness and training efficacy.

Typically, a single original image divided by a fixed scale with conditional selection can generate more than 4 slides with partial cancer areas that meet the requirements, which results in the number of our proposed dataset that contains part of the cancer area and is targeted for the medical image deep learning task is expanded extremely. We divided the resulting 7927 new slides into a training set, a test set and a validation set in a ratio of 8:1:1, shown as Table 1. Each image has a corresponding label associated with it.

2.2. Chromosomal regularisation

Manual relabeling is a prerequisite for a complete annotation of the tumor area to be performed with the assistance of ASAP. Moreover, to address the problem of varying colour fields of annotation in different datasets, the paper conducted chromosomal regularisation of the newly generated slides, where the selected hematoxylin staining resulted in a uniform violet bias in the regularised images. Notably, WSI binary masks were also recreated and added to the dataset for validation. Thus, during model training, the input fixed-size slides and ground truth image annotations would indicate the location of regions containing metastatic cancer in each WSI.

An overall sample dataset presentation is shown in Fig. 3. After optimising the highly repetitive and complex data preprocessing conducted in most other studies into a uniform downsampling operation, and consolidating it at a fixed size of 515 x 512 pixels patch into a remarkably lightweight public dataset relative to the 100,000 x 100,000 gigapixels competition dataset, deep learning models for histopathology images, including the proposed HI-Net described later in this paper, can better assist pathologists in segmenting and localising cancer regions.

3. HI-Net

As shown in Fig. 4, multi-scale semantic solid features of the histopathologic image segmentation model called HI-Net are constructed in this paper. The WSIs incorporated into the network exhibit robust semantic features at all scales, which is attributed to the structure of the featured image pyramid. This structure enhances feature transmission by combining low-resolution features to create a robust semantic representation, which achieves consistent semantic understanding across all scales. The method works by seamlessly integrating two types of features: low-resolution features that carry high semantic intensity, and high-resolution features that have lower semantic intensity. This fusion ensures that both detailed and broader semantic information are effectively captured and represented. The synergy of top-down pathways and lateral connections facilitates a holistic capture of semantic nuances at various resolutions.

Employing ResNet101 [43] as its backbone, HI-Net revolutionises the semantic feature integration mechanism commonly observed in FPNs through the novel top-down attention (TDA) module, which not only ensures the precision of semantic feature delivery but also significantly reduces computational overhead by refining the FPN's traditional multi-convolution approach to instance segmentation into a streamlined process that boasts high segmentation prediction accuracy.

The work of HI-Net for histopathologic image-based tasks is also reflected in the corresponding optimisation of the feature connection mechanisms in different network layers. Specifically, the backbone network of HI-Net was deliberately selected through comprehensive ablation studies, contrasting various Mask-RCNN structures [44] to identify the most compelling feature extraction framework. Furthermore, the TDA module was proposed to address the problem of massive loss of the underlying object information after many layers of the network by using channel and spatial attention to introduce weighted feature maps.

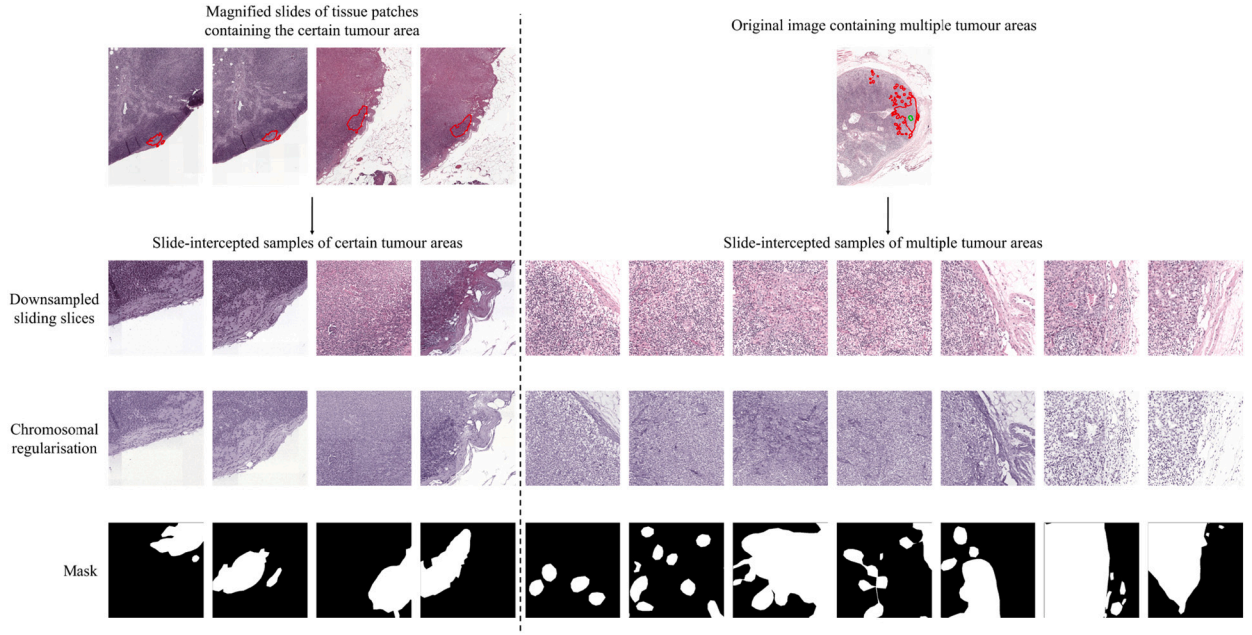

Fig. 3. Schematic process of the construction of our proposed lightweight dataset. The first line to the left of the dotted line represents magnified slides of tissue patches containing the certain tumour area, and to the right is the original image containing multiple tumour regions. The second to fourth rows present the downsampled and scaled slides, the hematoxylin-regularized bias purple image, and the mask.

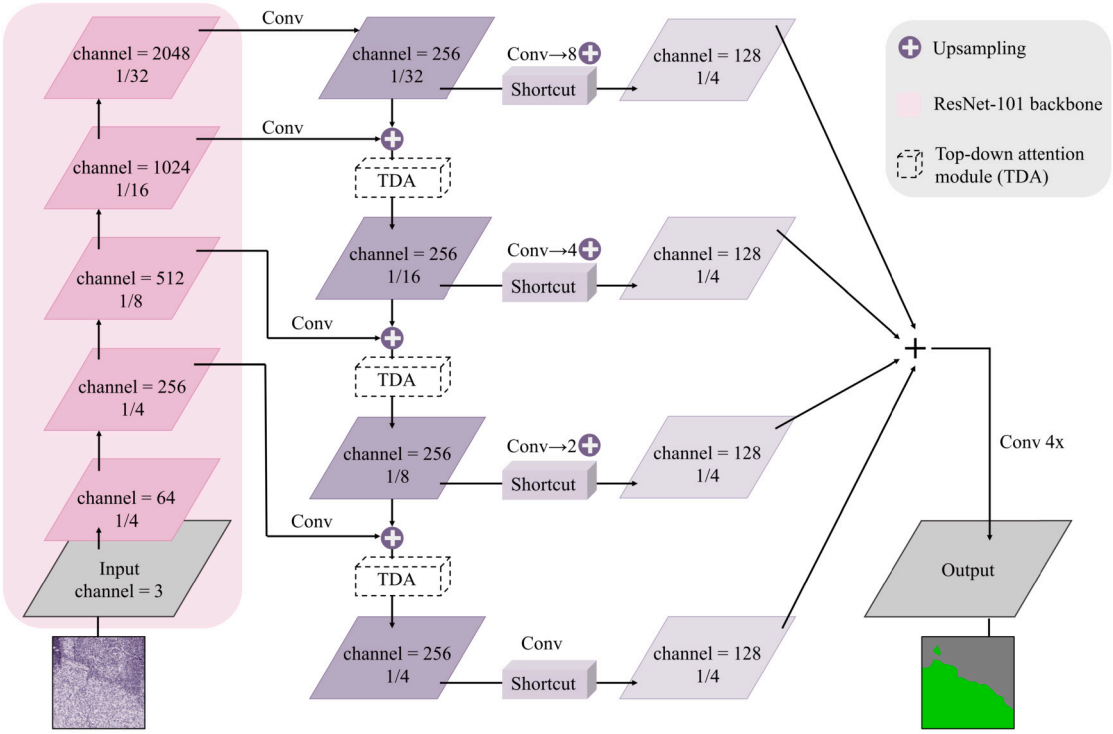

Fig. 4. Overview of the proposed HI-Net. The backbone network begins by extracting multi-scale image features. The TDA module then enhances semantic information transfer between layers, reducing information loss. Subsequently, the shortcut module optimises the process by redirecting redundant operations towards the upsampling process in the pyramid architecture, thereby improving the overall network performance.

Moreover, for histopathologic image segmentation or localisation, since the newly added attention mechanism has enhanced the prediction accuracy, the lateral connection using the coarse map with the same size and number of channels to increase the accuracy becomes redundant. Thus, considering the new architecture can render traditional laterally connected coarse mapping methods aimed at improving accuracy through size and channel matching both obsolete and overly complex. The emergence of such complex attention mechanisms requires us to move away from traditional methods and adopt a more streamlined and efficient model architecture. Therefore, a novel skip connection was urgently required, which is why this paper proposes a shortcut module for multilayer features.

Our HI-Net for histopathologic image segmentation can be conceptually divided into three distinct components. The initial component involves the extraction of the feature map from the input image. In this study, a bottom-up pathway, specifically ResNet101, was selected for feature extraction following comparative experiments with other backbones, including ResNet50 and ResNext101. The second component involves the transformation of the shared feature map into a featured image pyramid structure. This process generates bottom-object features enriched with semantic information via the TDA module. Subsequently, the shortcut module is utilised to obtain a refined prediction frame. The final component is dedicated to the classification of the prediction frame, regression localisation, and the prediction of the mask for each ROI within the image. This process aims to complete a pixel-level segmentation task. The structure and function of HI-Net, as described, ensure a comprehensive and efficient approach to histopathologic image segmentation.

HI-Net employs a multi-task loss function, which is an amalgamation of classification loss (L_{cls}) and mask loss (L_{mask}). The formulation of this function is depicted in Equation (1). This multi-task loss function serves as a comprehensive measure to optimize the performance of HI-Net in histopathologic image segmentation tasks.

$$\mathrm{L} = \frac{1}{N_{cls}} \sum_i L_{cls}\left(p_i, p_{\mathrm{i}}'\right) + \frac{1}{N_{\mathrm{mask}}} \sum_i L_{\mathrm{mask}}\left(m_i, m_{\mathrm{i}}'\right), \tag{1}$$

where N_{cls} denotes the total number of classification categories, and the parameter λ serves as a balancing factor between the magnitudes of N_{cls} and N_{reg}. The classification loss function is represented by L_{cls}. Furthermore, p_i symbolizes the confidence level associated with the object being identified as the target. Conversely, p_{i}' is a binary function, assuming values of either 0 or 1. In detail, L_{cls}, representing the classification loss function, can be expressed as Equation (2):

$$\mathrm{L}_{cls}\left(p_i, p_i'\right) = \log\left[p_i p_i' + \left(1 - p_i'\right)\left(1 - p_i\right)\right], \tag{2}$$

where p_i denotes the confidence level associated with the object being identified as the target. Conversely, p_i' is a binary function, assuming values of either 0 or 1. The function assumes a value of 1 when the object is a positive example and 0 when it is a negative example. Notably, the mask branch uses an average binary cross-entropy loss function, as shown in Equation (3).

$$\mathrm{L}_{\mathrm{mask}} = -\left[m_i \cdot \log(m_i') + (1 - m_i) \cdot \log(1 - m_i')\right], \tag{3}$$

where $\mathrm{L}_{\mathrm{mask}}$ represents the mask loss, which is the quantity the model aims to minimize during training. The binary value m_i denotes the ground truth mask for the i^{th} pixel in the image and m_i' is the predicted mask for the i^{th} pixel in the image, as output by the model, showing the model's confidence that the pixel belongs to the object of interest.

Regarding feature extraction, ResNet101 is employed as the backbone architecture for the extraction of features from breast cancer WSIs. ResNet101 utilises a cross-layer connection strategy, wherein each residual block is implemented via a feed-forward neural network and a shortcut connection. The structure of these residual blocks is defined as per Equation (4). This approach ensures a robust and efficient extraction of image features, thereby enhancing the performance of the overall network.

$$\mathrm{y}_l = x_l + F\left(x_l, \left\{W_l\right\}\right), \tag{4}$$

where x_l denotes the input to the l^{th} residual block, while y_l signifies the output from the same. The weight coefficient within the l^{th} residual block is represented by W_l. The residual function is denoted by $F\left(x_l, \left\{W_l\right\}\right)$, which encapsulates the transformation applied to the input x_l by the l^{th} layer of the network, parameterised by the weights $\left\{W_l\right\}$.

The chosen backbone comprises numerous convolutional layers that maintain the same scale. In the context of the proposed HI-Net, the last layer's residual block from each group of convolutional layers with identical scale is extracted. This process facilitates the construction of the feature pyramid (from 'channel = 64' to 'channel = 2048', shown in the backone part of Fig. 4), thereby enhancing the network's ability to capture and represent complex features at various scales. This comprehensive representation of variables and parameters contributes to a nuanced understanding of the underlying mathematical model and its practical implementation.

3.1. Top-down attention module

The task of object detection across varying scales presents challenges, particularly in the context of minuscule regions within expansive images. This complexity is exemplified in the endeavour to identify lesioned regions from WSIs, a task that is compounded by the highly localised nature of traditional convolution-based models. These models' receptive field overlay, constructed through the stacking of multiple layers, often results in a receptive field that is insufficiently extensive. This limitation persists even when the theoretical receptive field does not attain the size of the target.

The pyramid structure can serve as a mechanism to augment semantic information, as it generates feature maps wherein each layer is enriched with semantic information, including the high-resolution lower layers. However, feature maps are closer to image

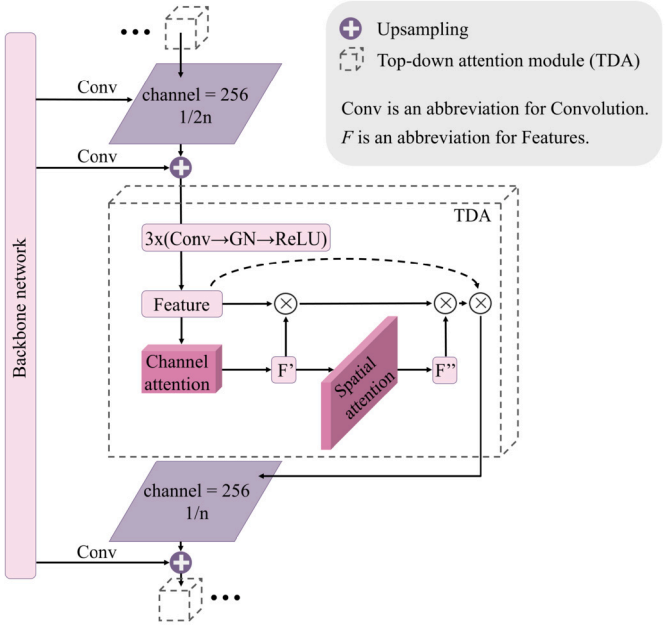

Fig. 5. The schematic illustration of the TDA module.

layers, which are obtained from low-level structures and are ineffective for accurate object detection. The multiscale feature map has better information quality. It facilitates object detection, but as shown above. However, the semantic features at higher levels are passed down through lateral connections; only the semantic information of the feature pyramid is enhanced. If the features are directly passed down from the original graph, many layers of the network have to be passed in between, and the object information at the bottom layer is already lost.

To address the problem of CNNs being overly local and globally insufficient so as to obtain global contextual information, the article argues that a redesigned module based on the attention mechanism would be a positive solution. The Convolutional Block Attention Module (CBAM) [34] is a groundbreaking approach that integrates spatial and channel attention mechanisms. This integration enables access to aggregated feature mapping of spatial information, weighted by importance, marking a significant advancement over the Squeeze-and-excitation network (SeNet) [45], which primarily concentrates on the channel. As a result, this paper advocates for the incorporation of a CBAM-inspired attention mechanism that amalgamates spatial and channel attention and then names it as the TDA module, serving the top-down path and effectively reducing unnecessary information loss inherent in multi-layer networks. Precisely, the top-down path engenders features of higher resolution by upsampling spatially coarser, yet semantically stronger, feature maps from a higher level of the pyramid. These features are subsequently enhanced via lateral connections that utilise features from the bottom-up path. Each lateral connection amalgamates feature maps of the same spatial size from both the bottom-up and top-down paths. Despite possessing a lower semantic level, the bottom-up feature map demonstrates more accurately localised activation due to a reduced number of subsamples.

Fig. 5 illustrates the architecture of the proposed TDA module. The design of TDA facilitates the passage of the feature from a deeper layer through a transition module, which is comprised of a convolutional layer, supplemented by a Group Normalisation (GN) and a Rectified Linear Unit (ReLU) activation function. This transition module then feeds into a channel and spatial-based attention mechanism, followed by a combination of sequential features in tandem, which ensures a robust and efficient processing of features.

In detail, the input feature map (H, W, C) to the TDA module initially traverses the 'Conv-GN-ReLU' transition module, as shown in Fig. 5. Although the benefits of Batch Normalisation (BN) for image optimisation and network convergence are acknowledged, it is noted that for datasets with WSIs, the mean and variance exhibit significant fluctuations. This variance can lead to inconsistencies between the training and testing phases, potentially causing complications. Therefore, the adoption of Group Normalisation allows both the input into the overarching module and the TDA feature maps to achieve feature grouping and intra-group regularisation. The expression of set S_i of GN is shown as Equation (5) below:

$$S_i = \left\{ k \mid k_N = i_N \text{ and } \left\lfloor \frac{k_C}{C/G} \right\rfloor = \left\lfloor \frac{i_C}{C/G} \right\rfloor \right\}, \tag{5}$$

where S_i denotes the set of indices corresponding to the elements in the i^{th} group. k represents the index of an element within the feature map. k_N refers to the batch dimension; $k_N = i_N$ indicates that we're considering the same sample in the batch for comparison. And k_C denotes the channel dimension; k_C and i_C denote specific channels within the feature map. The subscript N typically refers to the batch dimension, implying that k_N and i_N represent the k^{th} and i^{th} samples in the batch, respectively. Similarly, the subscript C typically refers to the channel dimension, indicating that k_C and i_C correspond to the k^{th} and i^{th} channels in the feature map,

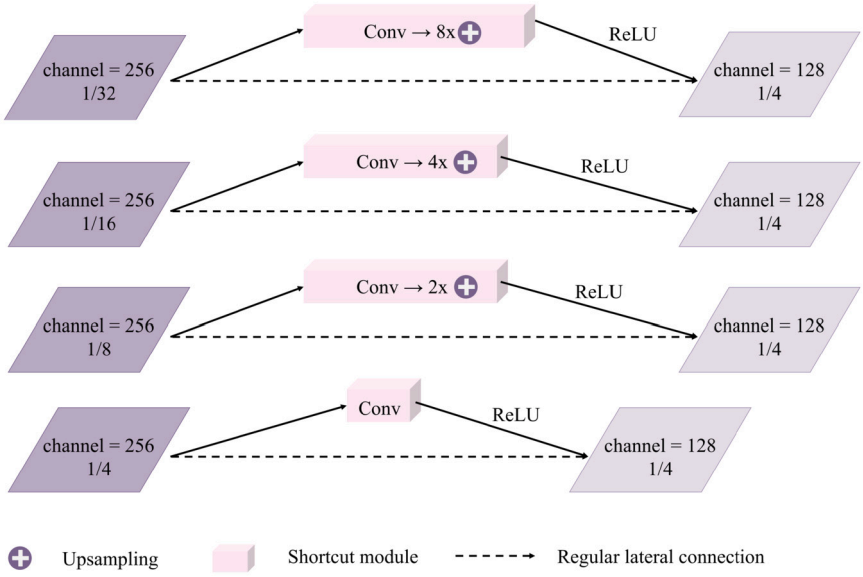

Fig. 6. The overview of the proposed shortcut module and the comparison with lateral connection that are regularly used.

respectively. The variable G signifies the number of groups, a hyperparameter set to 6 in this study, while C/G represents the number of channels per group. The floor function $\lfloor\rfloor$ denotes a rounding down operation, hence $\left\lfloor \frac{k_C}{C/G} \right\rfloor = \left\lfloor \frac{i_C}{C/G} \right\rfloor$ implies that k and i are channels within the same group, representing each group of channels arranged in order along the C axis. This approach maintains a low error rate across different batch sizes, rendering it particularly suitable for training the WSIs dataset.

The normalised feature map (H, W, C) undergoes spatial compression via the top-down pathway through the channel attention module, resulting in a one-dimensional vector. Subsequently, the spatial information of the feature map is aggregated through average pooling and max pooling, generating a channel attention map. This map is then processed by a spatial attention module in a sequential manner, and the math principle and process can be expressed as the Equation (6):

$$\begin{aligned} \mathbf{F}' &= \mathbf{M_c}(\mathbf{F}) \otimes \mathbf{F}, \\ \mathbf{F}'' &= \mathbf{M_s}\left(\mathbf{F}'\right) \otimes \mathbf{F}', \end{aligned} \tag{6}$$

where $\mathbf{M_c}$ and $\mathbf{M_s}$ denote global maximum pooling and global average pooling, respectively. The symbols $\mathbf{F}$, $\mathbf{F}'$, and $\mathbf{F}''$ represent the input feature maps, channel-optimised feature maps, and channel- and spatial-optimised feature maps, respectively. The operator $\otimes$ signifies pixel-by-pixel multiplication. This notation provides a clear and concise representation of the operations and transformations applied to the feature maps in the process.

Specifically for input histopathological sections, the TDA module enables the proposed HI-Net to concentrate more effectively on significant regions of the graph, the cancerous areas. This sequence of operations ensures a robust and efficient extraction and processing of image features, thereby enhancing the overall performance of the network.

3.2. Shortcut module

The Shortcut module is distinct from the residual connections found in architectures like ResNet. While ResNet's residual connections are designed to combat the vanishing gradient problem by allowing the shortcut for gradient flow during backpropagation, our Shortcut module has a different focus and architecture designed to streamline information processing within the network, particularly in the semantic partitioning branch. Our network introduces the Shortcut module to optimize the transformation of deep model outputs into coarse maps with identical channel counts and sizes. This process is crucial for efficient information aggregation and fusion without significantly increasing the computational load. The module consists of a streamlined sequence: Convolution → Group Normalisation → ReLU → Upsampling. This sequence is meticulously designed to adjust the output from the network's deeper layers, ensuring that they are in a suitable form for effective feature fusion across different semantic levels of the network, shown in Fig. 6. The introduction of the Shortcut module addresses a critical concern in deep learning models: the computational cost associated with processing high-dimensional data and the complex inter-layer feature fusion. By simplifying the lateral connections to a direct two-dimensional convolution operation, the Shortcut module significantly reduces the network's computational complexity. This reduction is achieved without compromising the model's ability to understand and segment histopathological images accurately.

Furthermore, this module's design circumvents the potential for overfitting associated with the addition of more complex modules like the TDA module, which, despite showing some improvements in accuracy, increased the model's parameter count and,

consequently, its vulnerability to overfitting. Considering that HI-Net already employs a multitasking loss function, taking advantage of newly designed Shortcut or dropout operations may be a viable alternative to routinely applying l_1 or l_2 penalty functions.

In terms of the impact on the HI-Net structure, traditional lateral joins typically involve the summation of cross-layer pixel features with different upsampling sizes, which introduces significant complexity and computational requirements. In contrast, the Shortcut module provides a more elegant solution that maintains the integrity and richness of the feature maps across the network. This is particularly beneficial when dealing with WSI for breast cancer diagnosis, where maintaining high-resolution feature details is crucial for accurate segmentation and diagnosis. Thus, in essence, the Shortcut module represents a significant architectural innovation for our network, striking a balance between computational efficiency and diagnostic accuracy. Its streamlined design not only improves the performance of the network but also makes it a more straightforward solution for a wide range of applications, including those with limited computational resources.

4. Implementation details

4.1. Experiment settings

Our experiments are implemented using the PyTorch framework and run efficiently on two NVIDIA GeForce RTX 3080 at a personal workstation. To demonstrate the performance of the proposed HI-Net on different datasets, including our proposed lightweight dataset, we set up a number of experiments to compare representative medical image segmentation or detection models, and the compared methods were trained and tested under the same settings. Cross-entropy loss was used to train the HI-Net and the comparison models. We trained all models for 70 epochs, over 6 batch sizes, and a learning rate of 0.0001. Moreover, multiple ablation experiments are implemented on the proposed TDA module and shortcut module to validate the effectiveness of the proposed method.

4.2. Dataset setting

Our proposed dataset of lightweight histopathological breast cancer slide images is based on three competition datasets: CAMELYON 16, CAMELYON 17, and ICIAR 2018. For CAMELYON, which provides gigapixel slides, we downsampled four times to reduce resolution. The collected downsampled images are slide-intercepted to generate patch slides with a fixed size of 512 $\times$ 512 pixels. In order to save computational cost and integrate WSI's tedious data preprocessing into a unified pipeline, we chose to focus our analysis on regions on slides most likely to contain cancer metastases. Identifying tissue within WSIs to exclude background voids and remove WSIs where the area of the cancer is disproportionate to the overall image was a prerequisite for dataset creation. Therefore, our methodology utilises the Otsu algorithm for precise segmentation of tumor regions in WSIs, with the optimal threshold defined through preliminary analysis to ensure maximum contrast and diagnostic accuracy. To avoid dataset imbalance and improve training efficiency, we exclude slides with tumor coverage below 5% and above 99.5%, as confirmed by extensive experiments across multiple datasets. This approach, detailed in Fig. 2 from section 2, ensures a balanced dataset and optimizes the model's ability to distinguish tumor from non-tumor tissues effectively.

Furthermore, manual re-annotation and chromosome regularisation were performed at the suggestion of pathologists, resulting in a uniform format and colour distribution for new slides. The 7927 new slides obtained by the above method are divided into training set, test set, and verification set according to the ratio of 8:1:1 so that the construction of the lightweight data set proposed in the article can be built.

For the control group used as a comparative experiment, the CAMELYON 16, CAMELYON 17, and ICIAR 2018 datasets were selected as benchmarks to validate the robustness of the proposed model. It is worth mentioning that the CAMELYON series with a general resolution above 100,000 $\times$ 100,000 pixels still needs to be downsampled before it can be input into the neural network. Furthermore, since CAMELYON 17 inherits a large part of the training set of CAMELYON 16 and has made phased optimisation, this paper chooses to integrate the data sets of CAMELYON 16 and CAMELYON 17 as one of the benchmarks for comparison, named CAMELYON series.

4.3. Evaluation metrics

In this task, the model prediction accuracy (Acc) corresponding to the semantic segmentation pixel accuracy needs little elaboration, as it intuitively represents the proportion of correct predictions to the total predicted values. However, the proposed Mean Pixel Accuracy (MPA) helps to apply the model to the calculation of the proportion of pixels correctly classified for each class separately, with the following Equation (7):

$$MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{p_{i0} + p_{i1} + \ldots + p_{ij}}, \tag{7}$$

where k represents the total number of pre-determined categories, while p_{ij} denotes the aggregate number of pixels in category i that are predicted as category j.

The Mean Intersection over Union (MIoU) serves as a conventional metric in semantic segmentation. Its uniqueness lies in its computation of the ratio between the intersection and union of two sets, specifically the ground truth and predicted segmentation

Table 2
Ablation experiment 1: Different backbones plus pyramid architecture and their Mean Average Precision (mAP).

Backbone	mAP (Training)	mAP (Testing)
ResNet101 + FPN	**0.919**	**0.843**
ResNet50 + FPN	0.841	0.813
ResNeXt101 + FPN	0.886	0.832

Table 3
Ablation experiment 2: The impact of TDA and Shortcut on the overall performance of the network.

MIoU↑ / Acc↑	Resnet101 + FPN	Resnet101 + FPN + TDA	Resnet101+ FPN + shortcut	Resnet101+ FPN + TDA + shortcut (HI-Net)
Our Dataset	0.821 / 0.833	0.862 / 0.871	0.823 / 0.839	**0.885** / **0.904**
CAMELYON series [38,39]	0.786 / 0.821	0.858 / 0.859	0.786 / 0.825	**0.876** / **0.884**
ICIAR 2018 [40]	0.780 / 0.801	0.857 / 0.868	0.778 / 0.803	**0.873** / **0.888**

in the context of image segmentation. This is encapsulated in Equation (8). Conversely, the Frequency Weighted Intersection over Union (FWIoU) enhances the MIoU by assigning weights to each class based on its frequency of occurrence, as depicted in Equation (9):

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}, \tag{8}$$

$$FWIoU = \frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k}p_{ij}}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}}, \tag{9}$$

where p_{ij} and p_{ii} maintain the same definitions as previously stated. Additionally, p_{ij} and p_{ji} correspond to false positives and false negatives, respectively. The term $k+1$ signifies the total number of categories, inclusive of empty categories. Typically, MIoU is computed on a per-class basis, whereby the IoU for each class is determined, subsequently summed, and averaged to yield a global evaluation.

5. Experiments and results

Upon training several benchmark models alongside our proposed HI-Net under uniform standards within a particular environment, the study initially demonstrates the rationale for selecting ResNet101 as the backbone network through the first ablation experiment. This experiment underscores the superiority of ResNet101 relative to other benchmark backbone networks.

A series of ablation experiments are executed on the two innovative modules introduced in this study, namely, the TDA and the shortcut module. These experiments aimed to illustrate the degree to which these modules have been optimised for precision and training speed in histopathologic image segmentation tasks, as delineated in the preceding subsections. Moreover, a comprehensive comparative experiment is carried out using several cutting-edge image processing models as benchmarks. These models span diverse domains, including salient object detection, image segmentation, and medical image detection. Through this extensive experiment, utilising various datasets, this study seeks to investigate and elucidate the performance of the proposed lightweight dataset and HI-Net in histopathologic image segmentation tasks, thereby further substantiating its superiority.

5.1. Ablation experiment

The first ablation experiment was designed to investigate the effect of different backbone networks combined with pyramidal structures on the accuracy of the network. In the article, several classical Mask-RCNN structures are selected as the control group for comparison experiments to explore the advancement of ResNet101 as a backbone network.

The experimental results are shown in Table 2. Compared to the other two ResNet as backbone networks, ResNet101 demonstrated its advanced feature extraction capability through superior prediction accuracy. The experiments also justified the reason for choosing ResNet101 for the article.

Subsequent extensive ablation experiments were designed to investigate the impact of the proposed TDA and Shortcut module on the network's overall performance. These experiments were conducted utilizing our proposed lightweight dataset, supplemented by several additional competition datasets. The objective was to mitigate the overfitting effect induced by the exclusive use of a single dataset for a specific network, thereby ensuring the attainment of as objective experimental results as possible.

As shown in Table 3, the incorporation of the TDA module substantially improved the predictive accuracy of the original amalgamation of the backbone network and feature pyramid structure. This enhancement was consistently demonstrated across all datasets. Conversely, the introduction of the shortcut module had a minimal impact on network accuracy. While the majority of experiments

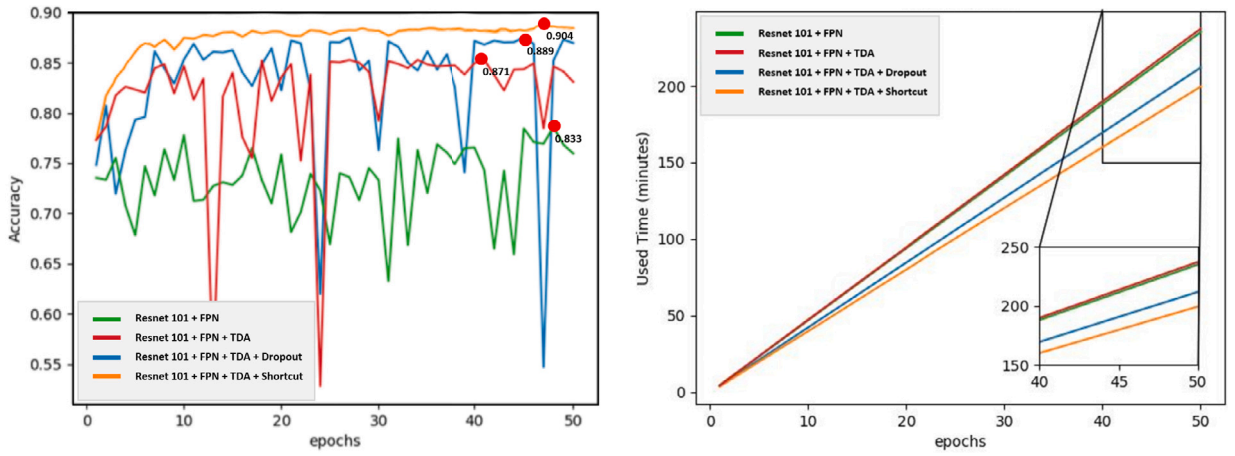

Fig. 7. Ablation experiment 3 and 4: Impact of Dropout operation on accuracy (a) and comparison of Dropout operation and proposed Shortcut module on model training time (b).

Table 4
The comparison between HI-Net and other methods based on our proposed lightweight dataset and other competition datasets.

Algorithms	Our Dataset				CAMELYON series				ICIAR 2018			
	Acc↑	MPA↑	MIoU↑	FWIoU↑	Acc↑	MPA↑	MIoU↑	FWIoU↑	Acc↑	MPA↑	MIoU↑	FWIoU↑
PSPNet [46]	0.835	0.813	0.708	0.706	0.830	0.811	0.685	0.701	0.826	0.803	0.699	0.694
FPN [28]	0.843	0.828	0.697	0.667	0.836	0.820	0.657	0.688	0.839	0.833	0.679	0.682
Mask-RCNN [44]	0.851	0.848	0.701	0.680	0.842	0.832	0.710	0.695	0.839	0.836	0.715	0.706
UNet [26]	0.876	0.870	0.774	0.861	0.797	0.784	0.730	0.744	0.848	0.822	0.770	0.749
AttentionUNet [27]	0.868	0.861	0.762	0.767	0.859	0.831	0.749	0.751	0.850	0.821	0.762	0.755
TransUNet [23]	0.881	0.875	0.782	0.787	0.879	0.871	0.780	0.792	0.872	0.866	0.792	0.778
HI-Net	**0.904**	**0.885**	**0.813**	**0.796**	**0.884**	**0.876**	**0.781**	**0.783**	**0.888**	**0.873**	**0.800**	**0.806**

indicate that it exerts a certain positive influence on overall performance, its primary effect is manifested in the reduction of computational cost. Overall, a structure like HI-Net yields optimal outcomes on both the competition dataset and our proposed lightweight dataset. Crucially, the central columns of the experiment further corroborate the efficacy of the proposed module.

In addition to the two novel modules proposed in this paper, the dropout operation is also mentioned in the subsection 'Shortcut module', which states that the dropout operation may be a way to save computational costs and reduce the complexity of the model. Therefore, this section also presents an ablation and comparison experiment between the dropout operation and our proposed Shortcut module to verify the effectiveness of the dropout operation and to re-emphasise the superiority of the proposed Shortcut module, shown as the Fig. 7.

As shown in Fig. 7(a), the Dropout operation indeed optimises potential overfitting issues that may occur when the model is burdened with an excess of parameters due to the integration of TDA. This results in a degree of improvement in model accuracy, albeit not to the extent achieved by the Shortcut module. In contrast, the merits of the Shortcut module are clearly demonstrated in Fig. 7(b), as it significantly curtails the model's training duration and conserves substantial computational resources. While Dropout also effectively reduces training time, outperforming the original model structure in speed, it does not match the results achieved by the Shortcut module. Consequently, the HI-Net proposed in this study, together with its constituent modules and architecture, has demonstrated its advantages. Such a combination can significantly enhance the accuracy of histopathologic image segmentation while preserving computational efficiency. However, subsequent comparisons with other advanced models remain essential for a more profound validation of the model's superiority.

5.2. Experimental results of different detection models

The results of our study, as shown in Table 4, clearly establish the efficacy of HI-Net in the domain of histopathological image segmentation when contrasted with various well-established algorithms across multiple datasets. HI-Net not only achieves the highest accuracy (90.4% in our dataset, 88.4% in the CAMELYON series, and 88.8% in ICIAR 2018) but also excels in precision and intersection over union metrics, critical indicators of segmentation quality. This performance is indicative of HI-Net's adeptness at nuanced differentiation between tumor and non-tumor tissue, a testament to its architectural strengths. Delving deeper into the comparative analysis, it is evident that HI-Net's design facilitates its outstanding performance.

On our proposed lightweight dataset, HI-Net's superior scores highlight its precision in segmenting tumor regions, a critical aspect considering the inherent complexity of histopathological images. Furthermore, the model's robustness is exemplified by its consistent outperformance on external datasets, such as the CAMELYON series and the ICIAR 2018 benchmark, underlining its adaptability and the generalisability of its learning capabilities.

While other models, such as TransUNet, also exhibited commendable performance, they were slightly outperformed by HI-Net. This could be attributed to architectural differences where TransUNet may be more prone to the loss of low-level detail, such as edge information, which is crucial in the precise identification of tumor boundaries. HI-Net's architecture, by contrast, is tailored to retain such critical information, ensuring that the segmentation captures the full spectrum of diagnostic features necessary for accurate pathology assessment.

With the architecture's efficacy confirmed under uniform hyperparameters, the study's findings affirm our anticipation that HI-Net would be a highly effective framework for digital pathology. The results go beyond validating the superiority of our model in terms of traditional performance metrics; they also suggest that HI-Net can meaningfully contribute to the practical field of pathology by providing a tool that enhances diagnostic precision. Furthermore, the consistency in high performance, particularly in delineating tumor regions within histopathological images, speaks volumes about the potential clinical applications of HI-Net. By setting new benchmarks in automated segmentation, our architecture not only stands as a significant contribution to academic research but also holds promise for improving clinical workflows and patient outcomes in breast cancer diagnosis.

6. Conclusion

The findings presented in this study offer a pioneering contribution to the field of histopathology image analysis. We introduce the first lightweight, large-scale, standardized histopathology slide dataset comprising 7,927 slides, meticulously reconstructed and size-normalized from a variety of competition datasets. The novel threshold-based region selection method we propose ensures a balanced and homogeneous distribution of tumor regions across the map area, addressing a gap not previously considered in existing datasets. Additionally, our optimised downsampling process and uniform chromatic regularization significantly enhance dataset usability, eliminating the need for extensive preprocessing while maintaining consistency in format and colour spectrum.

Building on this dataset, we introduce HI-Net, a cutting-edge network specifically designed for breast cancer histopathology image segmentation. HI-Net features a pyramidal network structure augmented with high-speed panoramic feature acquisition capabilities, a novel TDA module, and a Shortcut module that collectively reduces computational overhead while maintaining high segmentation accuracy. The TDA module addresses the prevalent challenge of information loss at the base of pyramidal structures, thereby enabling more efficient semantic feature extraction. Concurrently, the Shortcut module optimizes feature summation across network layers, enhancing computational efficiency without compromising detail and accuracy.

Extensive ablation studies provide robust evidence for the effectiveness of the TDA and Shortcut modules, which play critical roles in enhancing HI-Net's overall performance. Comparative evaluations on multiple benchmark datasets further demonstrate HI-Net's superiority in information transfer and semantic feature integration, achieving a segmentation accuracy of 90.4% on our proposed lightweight dataset—outperforming existing advanced models for object detection and segmentation tasks. Additionally, HI-Net demonstrates a substantial increase in training speed, underscoring its efficiency and scalability. The integration of HI-Net with this new dataset not only provides a strong foundation for advancing histopathology image segmentation techniques but also significantly alleviates the diagnostic workload of pathologists, facilitating more efficient and accurate AI-driven medical diagnostics.

CRediT authorship contribution statement

Fengze Li: Writing – review & editing, Writing – original draft, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jieming Ma:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Formal analysis. **Tianxi Wen:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Zhongbei Tian:** Supervision. **Hai-Ning Liang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The proposed lightweight dataset is available at: https://data.mendeley.com/datasets/hd3kw5tt8f/1.

Acknowledgements

This research is supported by the Natural Science Foundation of China (Grant No. 62472361), the Suzhou Science and Technology Project-Key Industrial Technology Innovation (SYG202122), 2024 Suzhou Innovation Consortium Construction Project, the XJTLU

Postgraduate Research Scholarship (Grand No. PGRS1906004), the XJTLU AI University Research Centre, Zooming New Energy-XJTLU Smart Energy Joint Laboratory, Jiangsu Province Engineering Research Centre of Data Science and Cognitive Computation at XJTLU and SIP AI innovation platform (YZCXPT2022103).

References

[1] J. Ferlay, M. Colombet, I. Soerjomataram, D.M. Parkin, M. Piñeros, A. Znaor, et al., Cancer statistics for the year 2020: an overview, Int. J. Cancer 149 (4) (2021) 778–789.
[2] T.G. Debelee, F. Schwenker, A. Ibenthal, D. Yohannes, Survey of deep learning in breast cancer image analysis, Evolv. Syst. 11 (1) (2020) 143–163.
[3] M. Arnold, E. Morgan, H. Rumgay, A. Mafra, D. Singh, M. Laversanne, et al., Current and future burden of breast cancer: global statistics for 2020 and 2040, Breast 66 (2022) 15–23.
[4] M.M. Hassan, M.M. Hassan, F. Yasmin, M.A.R. Khan, S. Zaman, K.K. Islam, et al., A comparative assessment of machine learning algorithms with the least absolute shrinkage and selection operator for breast cancer detection and prediction, Decis. Anal. J. 7 (2023) 100245.
[5] World Health Organization, Breast cancer, https://www.who.int/news-room/fact-sheets/detail/breast-cancer, 2024. (Accessed 16 March 2024).
[6] M. Cui, D.Y. Zhang, Artificial intelligence and computational pathology, Lab. Invest. 101 (4) (2021) 412–422.
[7] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A.H. Beck, Deep learning for identifying metastatic breast cancer, arXiv preprint, arXiv:1606.05718, 2016.
[8] A. Pedersen, M. Valla, A.M. Bofin, J.P. De Frutos, I. Reinertsen, E. Smistad, Fastpathology: an open-source platform for deep learning-based research and decision support in digital pathology, IEEE Access 9 (2021) 58216–58229.
[9] R.E. Nakhleh, Error reduction in surgical pathology, Arch. Pathol. Lab. Med. 130 (5) (2006) 630–632.
[10] J.G. Elmore, G.M. Longton, P.A. Carney, B.M. Geller, T. Onega, A.N. Tosteson, et al., Diagnostic concordance among pathologists interpreting breast biopsy specimens, JAMA 313 (11) (2015) 1122–1132.
[11] A. Dy, N.N.J. Nguyen, J. Meyer, M. Dawe, W. Shi, D. Androutsos, et al., AI improves accuracy, agreement and efficiency of pathologists for Ki67 assessments in breast cancer, Sci. Rep. 14 (1) (2024) 1283.
[12] E.A. Rakha, F.G. Pareja, New advances in molecular breast cancer pathology, in: Seminars in Cancer Biology, vol. 72, Elsevier, 2021, pp. 102–113.
[13] D.M. Metter, T.J. Colgan, S.T. Leung, C.F. Timmons, J.Y. Park, Trends in the US and Canadian pathologist workforces from 2007 to 2017, JAMA Netw. Open 2 (5) (2019) e194337.
[14] N. Bayramoglu, J. Kannala, J. Heikkilä, Deep learning for magnification independent breast cancer histopathology image classification, in: 2016 23rd International Conference on Pattern Recognition (ICPR), IEEE, 2016, pp. 2440–2445.
[15] M.A. Wakili, H.A. Shehu, M.H. Sharif, M.H.U. Sharif, A. Umar, H. Kusetogullari, et al., Classification of breast cancer histopathological images using densenet and transfer learning, Comput. Intell. Neurosci. 2022 (2022).
[16] J.P.M. Rodriguez, R. Rodriguez, V.W.K. Silva, F.C. Kitamura, G.C.A. Corradi, A.C.B. de Marchi, et al., Artificial intelligence as a tool for diagnosis in digital pathology whole slide images: a systematic review, J. Pathol. Inform. (2022) 100138.
[17] J. Pallua, A. Brunner, B. Zelger, M. Schirmer, J. Haybaeck, The future of pathology is digital, Pathol. Res. Pract. 216 (9) (2020) 153040.
[18] A.B. Farris, C. Cohen, T.E. Rogers, G.H. Smith, Whole slide imaging for analytical anatomic pathology and telepathology: practical applications today, promises, and perils, Arch. Pathol. Lab. Med. 141 (4) (2017) 542–550.
[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, et al., Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[20] S. Khan, N. Islam, Z. Jan, I.U. Din, J.J.C. Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, Pattern Recognit. Lett. 125 (2019) 1–6.
[21] E. Deniz, A. Şengür, Z. Kadiroğlu, Y. Guo, V. Bajaj, Ü. Budak, Transfer learning based histopathologic image classification for breast cancer detection, Health Inf. Sci. Syst. 6 (1) (2018) 1–7.
[22] S. Guan, M. Loew, Breast cancer detection using transfer learning in convolutional neural networks, in: 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), IEEE, 2017, pp. 1–8.
[23] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, et al., Transunet: transformers make strong encoders for medical image segmentation, arXiv preprint, arXiv: 2102.04306, 2021.
[24] L. Wang, R. Li, C. Zhang, S. Fang, C. Duan, X. Meng, et al., Unetformer: a unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery, ISPRS J. Photogramm. Remote Sens. 190 (2022) 196–214.
[25] C. Xin, Z. Liu, K. Zhao, L. Miao, Y. Ma, X. Zhu, et al., An improved transformer network for skin cancer classification, Comput. Biol. Med. 149 (2022) 105939.
[26] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[27] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention u-net: learning where to look for the pancreas. arXiv preprint, arXiv: 1804.03999, 2018.
[28] A. Kirillov, R. Girshick, K. He, P. Dollár, Panoptic feature pyramid networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6399–6408.
[29] B.M. Priego-Torres, D. Sanchez-Morillo, M.A. Fernandez-Granero, M. Garcia-Rojo, Automatic segmentation of whole-slide h&e stained breast histopathology images using a deep convolutional neural network architecture, Expert Syst. Appl. 151 (2020) 113387.
[30] M.A. Khalil, Y.C. Lee, H.C. Lien, Y.M. Jeng, C.W. Wang, Fast segmentation of metastatic foci in h&e whole-slide images for breast cancer diagnosis, Diagnostics 12 (4) (2022) 990.
[31] W.R. Drioua, N. Benamrane, L. Sais, Breast cancer histopathological images segmentation using deep learning, Sensors 23 (17) (2023) 7318.
[32] E. Tekin, Ç Yazıcı, H. Kusetogullari, F. Tokat, A. Yavariabdi, L.O. Iheme, et al., Tubule-u-net: a novel dataset and deep learning-based tubule segmentation framework in whole slide images of breast cancer, Sci. Rep. 13 (1) (2023) 128.
[33] Z. Baojun, Z. Boya, T. Linbo, W. Wenzheng, W. Chen, Multi-scale object detection by top-down and bottom-up feature pyramid network, J. Syst. Eng. Electron. 30 (1) (2019) 1–12.
[34] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 3–19.
[35] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, Neurocomputing 452 (2021) 48–62.
[36] D. Sui, W. Liu, J. Chen, C. Zhao, X. Ma, M. Guo, et al., A pyramid architecture-based deep learning framework for breast cancer detection, BioMed Res. Int. (2021) (2021) 1–10.
[37] Y. Liu, K. Gadepalli, M. Norouzi, G.E. Dahl, T. Kohlberger, A. Boyko, et al., Detecting cancer metastases on gigapixel pathology images, arXiv preprint, arXiv: 1703.02442, 2017.
[38] G. Challenge, Camelyon 2016, [EB/OL], https://camelyon16.grand-challenge.org/, 2016. (Accessed 1 December 2022).
[39] G. Challenge, Camelyon 2017, [EB/OL], https://camelyon17.grand-challenge.org/, 2017. (Accessed 1 December 2022).
[40] G. Challenge, Iciar 2018, [EB/OL], https://iciar2018-challenge.grand-challenge.org/, 2018. (Accessed 1 December 2022).
[41] C.P. Group, Asap (automated slide analysis platform), [EB/OL], https://computationalpathologygroup.github.io/ASAP/, 2018. (Accessed 11 December 2022).

[42] N. Otsu, A threshold selection method from gray-level histograms, IEEE Trans. Syst. Man Cybern. 9 (1) (1979) 62–66.
[43] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[44] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
[45] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[46] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2881–2890.