



OPEN ACCESS

EDITED BY

Xin Qi,
Eisai, United States

REVIEWED BY

Guotai Wang,
University of Electronic Science and
Technology of China, China
Chunpeng Wu,
Duke University, United States

*CORRESPONDENCE

André Pedersen
andre.pedersen@ntnu.no

SPECIALTY SECTION

This article was submitted to
Pathology,
a section of the journal
Frontiers in Medicine

RECEIVED 17 June 2022

ACCEPTED 24 August 2022

PUBLISHED 14 September 2022

CITATION

Pedersen A, Smistad E, Rise TV,
Dale VG, Pettersen HS, Nordmo T-AS,
Bouget D, Reinertsen I and Valla M
(2022) H2G-Net: A multi-resolution
refinement approach for segmentation
of breast cancer region in gigapixel
histopathological images.
Front. Med. 9:971873.
doi: 10.3389/fmed.2022.971873

COPYRIGHT

© 2022 Pedersen, Smistad, Rise, Dale,
Pettersen, Nordmo, Bouget,
Reinertsen and Valla. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

H2G-Net: A multi-resolution refinement approach for segmentation of breast cancer region in gigapixel histopathological images

André Pedersen^{1,2*}, Erik Smistad^{3,4}, Tor V. Rise^{1,5},
Vibeke G. Dale^{1,5}, Henrik S. Pettersen^{1,5}, Tor-Arne S. Nordmo⁶,
David Bouget³, Ingerid Reinertsen^{3,4} and Marit Valla^{1,2,5,7}

¹Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway, ²Clinic of Surgery, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway, ³Department of Health Research, SINTEF Digital, Trondheim, Norway, ⁴Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway, ⁵Department of Pathology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway, ⁶Department of Computer Science, UiT The Arctic University of Norway, Tromsø, Norway, ⁷Clinic of Laboratory Medicine, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway

Over the past decades, histopathological cancer diagnostics has become more complex, and the increasing number of biopsies is a challenge for most pathology laboratories. Thus, development of automatic methods for evaluation of histopathological cancer sections would be of value. In this study, we used 624 whole slide images (WSIs) of breast cancer from a Norwegian cohort. We propose a cascaded convolutional neural network design, called H2G-Net, for segmentation of breast cancer region from gigapixel histopathological images. The design involves a detection stage using a patch-wise method, and a refinement stage using a convolutional autoencoder. To validate the design, we conducted an ablation study to assess the impact of selected components in the pipeline on tumor segmentation. Guiding segmentation, using hierarchical sampling and deep heatmap refinement, proved to be beneficial when segmenting the histopathological images. We found a significant improvement when using a refinement network for post-processing the generated tumor segmentation heatmaps. The overall best design achieved a Dice similarity coefficient of 0.933 ± 0.069 on an independent test set of 90 WSIs. The design outperformed single-resolution approaches, such as cluster-guided, patch-wise high-resolution classification using MobileNetV2 (0.872 ± 0.092) and a low-resolution U-Net (0.874 ± 0.128). In addition, the design performed consistently on WSIs across all histological grades and segmentation on a representative $\times 400$ WSI took ~ 58 s, using only the central processing unit. The findings demonstrate the potential of utilizing a refinement network to improve patch-wise predictions. The solution is efficient and does not require overlapping patch inference or ensembling. Furthermore, we showed that deep neural networks can be

trained using a random sampling scheme that balances on multiple different labels simultaneously, without the need of storing patches on disk. Future work should involve more efficient patch generation and sampling, as well as improved clustering.

KEYWORDS

hybrid guiding, refinement network, deep learning, convolutional neural networks, digital pathology, hierarchical sampling, clustering, breast cancer

1. Introduction

Cancer is an important cause of death, and of all cancers, breast cancer has the highest incidence worldwide (1). Cancer diagnostics is based on clinical examination, medical imaging and histopathological assessment of the tumor. The latter includes analysis of specific biomarkers that often guides treatment of the patients. Most pathology laboratories are burdened by an increasing number of biopsies and more complex diagnostics (2). To reduce workload for pathologists, automatic assessment of tumors and biomarkers would be of value.

A natural first step in automatic tumor and biomarker analysis would be to correctly identify the lesion, thus separating the tumor from surrounding tissue. For automatic biomarker assessment, it is important to ensure that biomarker status is obtained exclusively in the invasive epithelial cancer cells. With the promise of deep learning-based methods in computational pathology (3), accurate segmentation of the cancer region would be beneficial for building new classifiers and facilitate other cancer analysis methods.

Processing histopathological WSIs is challenging due to their large size. WSIs captured at $\times 400$ magnification may be as large as $200k \times 100k$ pixels, and as such, cannot be used directly as input to convolutional neural networks (CNNs). A solution is to downsample the image to a size that is manageable for the CNN. However, this results in loss of information and is therefore often not useful for tumor segmentation. Another widely used approach is to divide the image into smaller patches (4), before each patch is sent to an algorithm to produce an output. The results are then stitched to form a complete segmentation or heatmap of the entire WSI. However, the use of such a patch-wise design based on high-resolution information only, often results in edge artifacts and poor global segmentation of larger structures (5).

Traditional segmentation methods such as geodesic active contour (6) and superpixel (7) have been used for segmentation of histopathological images. In recent years, deep learning-based methods have surpassed traditional methods. From the challenge paper of CAMELYON16 (8), all 32 participating teams used machine learning-based methods, of which 25

teams used deep learning-based methods. The seven non-deep learning-based methods performed poorest on both the semantic segmentation and classification tasks. We therefore only consider deep learning-based approaches in this study.

Schmitz et al. (9) compared multi-scale convolutional autoencoder (CAE) designs, applied in a patch-wise fashion across liver tumors in WSIs. They found that the network benefited significantly from the added multi-scale information, compared to the baseline U-Net (10). They also proposed non-overlapping inference to reduce runtime at the cost of reduced accuracy along patch edges. For handling these edge artifacts, Priego Torres et al. (5) proposed a conditional random field-based, patch-wise, merging scheme.

To improve the patch-wise design, Guo et al. (11) developed a multi-task network for classification and semantic segmentation of breast cancer. They used a pretrained InceptionV3 (12) architecture and fine-tuned it on the CAMELYON16 data set (8). Such transfer learning has the benefit of making training more efficient, as the network is not trained from scratch. Using a more complex backbone, such as InceptionV3, has the potential benefit of improved performance. However, the architecture is computationally expensive, and might therefore not be suitable for real-time applications, such as histopathological diagnostics.

Breast cancers are known for their intra- and intertumor heterogeneity, and thus their morphological appearance varies both within and between tumors. Due to intratumor heterogeneity, the patches generated from a single WSI often contain different tissue types and a varying morphological appearance. Qaiser et al. (13) studied the effect of smart patch selection and balancing in preprocessing, to produce models that performed well on varying types of tissue. They demonstrated that a deep clustering approach of patches outperformed the conventional k -means (14) clustering method.

A similar cluster-guiding strategy was performed by (15) using multiple instance learning. They used a pretrained VGG19-encoder (16) for feature extraction. The dimensionality of the features was reduced using principal component analysis (PCA) (17), before performing k -means clustering. Samples were drawn from these clusters and balanced during training.

The number of clusters was set to four, as they assumed that there were four main natural tissue types in the data.

For image classification, MobileNetV2 (18) and InceptionV3 (12) are popular baseline architectures, commonly used in digital pathology (4, 19, 20). For image segmentation, the most commonly used CAE is U-Net (10). To make the U-Net design more efficient, various concepts have been proposed, such as multi-scale input (21), deep supervision (22), and attention (23). These three concepts are used in the two segmentation architectures AGU-Net and DAGU-Net (24). Performing both detection and semantic segmentation in a single step is a challenge. The segmentation result is often suboptimal, and therefore a post-processing method is required. Refinement networks for the CAE itself have therefore been proposed, either in multiple steps (25) or end-to-end, e.g., DoubleU-Net (26).

We propose a novel cascaded CNN design, H2G-Net, which efficiently utilizes both high-resolution and global information. The design is rapid, does not require ensembling, and can be used on low-end hardware. H2G-Net is a two-stage design. In the first stage, initial segmentation of the tumor region is extracted by applying a patch-wise classifier model. To correct for the fragmentation and pixelation in the generated tumor heatmap, a refinement network is used in the second stage. The network utilizes both the tumor heatmap and a low-resolution version of the WSI to produce the final segmentation. To counter the heterogeneity of breast cancer, a novel sampling scheme is proposed, which simplifies training by handling multiple data imbalance challenges simultaneously and reads patches directly from the raw WSI format during training.

In this paper, we present the following contributions:

- (1) A new, large data set of 624 breast cancer WSIs annotated by pathologists.
- (2) A novel sampling scheme that extracts patches directly from the WSI, handles hierarchical structured data imbalance problems, and enables end-to-end cluster-guiding.
- (3) A novel approach where a cascaded CNN combines high-resolution and global information in histopathological images, producing superior performance over single-resolution approaches.
- (4) The proposed pipeline and trained models are made openly available for use in FastPathology (27).

2. Materials and methods

2.1. Data set and annotation design

In this study, we used 4 μm thick whole sections ($n = 624$) from a cohort of Norwegian breast cancer patients (28), Breast Cancer Subtypes 1 (BCS-1). All tumors were previously classified into histological grade, according to the Nottingham grading system (29). The sections were stained with hematoxylin-eosin (H&E), scanned at $\times 400$ magnification

using an Olympus scanner BX61VS with VSI120-S5, and stored in the cellSens VSI format using JPEG2000 compression.

For each WSI, the tumor area was delineated by pathologists using QuPath (30). The tumor area included not only invasive epithelial cells, but also surrounding stromal tissue including other cells such as fibroblasts and inflammatory cells, and blood vessels. It is well-known that these components are also important for cancer development (31–33), and thus including these in the annotation could be of value in future studies of cancer progression and prognostication. To speed up and assist with the annotation work, automatic and semi-automatic approaches were tested, similarly to the approach used by (34). We used two different approaches for annotation (AN1 and AN2). For both annotation designs, predicted annotations were manually adjusted by the pathologists using the brush tool in QuPath.

The first 150 WSIs were annotated using the AN1 method, which involved using the semi-automatic tissue detection function in QuPath. The following parameters were used for performing segmentation: simple tissue detection threshold 200, requested pixel size 20, and minimum area 100,000. In cases where the algorithm failed, the parameters were adjusted or the tumor was manually annotated from scratch.

The remaining WSIs ($n = 474$) were annotated using the AN2 method (see Figure 1). A patch-wise CNN, similar to the Inc-PW method described in Section 3.2, was trained from a subset of the first 150 annotated WSIs. The model was trained in Python, and the produced model was then applied to the remaining WSIs. The resulting heatmaps were imported in QuPath and converted to annotations. Simple morphological post-processing was then performed before the segmentations were adjusted by the pathologists. Finally, to ensure consistency, all annotations were reviewed by a single pathologist experienced in breast cancer pathology and minor adjustments were made.

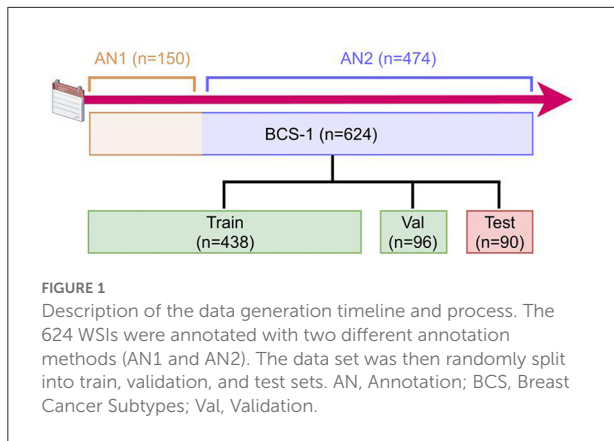
The pathologists' annotations were exported from QuPath as individual PNGs, one for each WSI, with a downsampling factor of four. The PNGs were then converted to tiled, pyramidal TIFFs, using the command line tool `vips`¹, with tiles sized 1024×1024 and a LZW lossless compression. All WSIs were converted to the single-file, pyramidal tiled, generic TIFF format using the command line tool `vsi2tif`². Lastly, the annotated WSIs were randomly distributed into the three sets: training ($\sim 70\%$; $n = 438$), validation ($\sim 15\%$; $n = 96$), and test ($\sim 15\%$; $n = 90$) set.

2.2. Preprocessing

The tissue regions of each WSI were automatically segmented by the following steps: (1) Extract the $\times 1.25$ image

1 <https://github.com/libvips/libvips>

2 <https://github.com/andreped/vsi2tif>



plane of the WSI, (2) convert the image to the HSV (Hue, Saturation, Value) color domain, (3) extract and threshold the saturation channel image using a fixed threshold of 20, (4) perform two consecutive applications of morphological closing using kernel sizes 5×5 and 3×3 , respectively.

Following annotation, we extracted patches sized 256×256 at $\times 100$ magnification level from the WSIs. Only patches containing more than 25% tissue were included. Patches with more than 25% tumor were considered tumor patches, and only patches with no tumor were considered non-tumor. The remaining patches in the range (0,25]% tumor tissue were discarded. For each WSI, the coordinates of accepted patches were stored along with the assigned label i.e., non-tumor/tumor.

2.3. Hierarchical sampling scheme

A batch generator was created to sample patches directly from the raw WSI format. Patches were read using OpenSlide (35), which enabled multi-threading processing. The generator was based on the condition that it is important to balance patches according to the following features: class label, tissue type, tissue and tumor area, and histological grade.

Patches were sampled in a hierarchical sampling scheme (see Figure 2), conducted as a tree structure uniformly distributed at each respective stage. The goal was to make all relevant outcomes equally probable. The sampling scheme was defined in the ordered stages: (1) Randomly select a histological grade, (2) from the grade select a WSI, (3) from the WSI select a class label, (4) from the class label select a patch.

To include patch-level tissue type label in the balancing scheme, we used our sampling generator to train a k -means clustering model, similar to (15). From a set of 100 batches of size 32, features were extracted using a VGG-16 (16) backbone pretrained on the ImageNet data set (36). The extracted features were then standardized using Z-score normalization, before PCA was performed. The number of principal components

was chosen such that 95% of the variance of the data was explained. The k -means model was then trained using $k = 10$ number of clusters, as recommended in a related study (37). The clustering model was implemented using the Python library scikit-learn (38).

To utilize the trained clustering model in the patch-wise CNN, TensorFlow (39) equivalents of the standardization, PCA and k -means transform methods were implemented, which was defined as a TensorFlow graph. The scikit-learn trained weights were then loaded for each corresponding component. Finally, for training the CNN classifier, each patch was passed through two different graphs; (I) a frozen pipeline that performed clustering and (II) a learnable deep neural network that performed classification. The outputs from both models were then passed to the loss function.

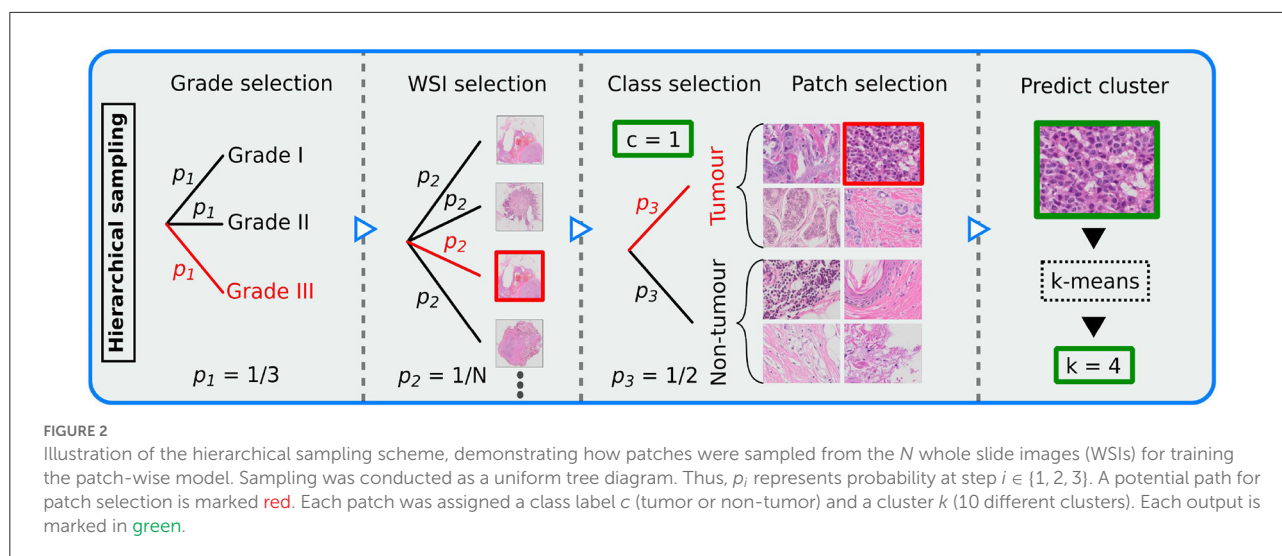
The MobileNetV2 (18) architecture was used for the patch-wise CNN classification of breast cancer tumor tissue, as it is lightweight, efficient, and optimized for low-end processors and thus suitable for real-time deployment. To further reduce the number of parameters, we simplified the classifier head. The updated classifier contained a global average pooling layer, followed by a dense layer of 100 hidden neurons, dropout (40) with a 50% drop rate, ReLU activation function, batch normalization (41), and finally a dense layer with softmax activation function.

2.4. Cluster-guided loss function

To balance on tissue type and thus ensure similar model performance on all predicted clusters, we included the cluster-information in the loss computation. By doing so, we enabled end-to-end cluster-guiding and made the clustering design more scalable, as the cluster heatmap does not need to be generated in a separate step or stored on disk. For a given batch, we calculated the cross-entropy loss for each cluster independently, and then calculated the macro average across each cluster. We named this loss function cluster-weighted categorical cross-entropy (CWCE) loss. The loss can be mathematically described as:

$$\mathcal{L}_{\text{CWCE}} = -\frac{1}{K_b} \sum_{k=1}^{K_b} \sum_{c=1}^C \sum_{i=1}^B \mathbb{1}(q_{i,k} = k) y_{i,c} \log(p_{i,c}) \quad (1)$$

where $i \in \{1, \dots, B\}$ represents sample i in a batch of size B , $k \in \{1, \dots, K_b\}$ cluster in a mini-batch b of size B of K_b represented clusters, $c \in \{1, \dots, C\}$ class, p class prediction, q cluster prediction, and ground truth tumor class. Note that the number of clusters K_b may vary between mini-batches.



2.5. Heatmap generation and refinement stage

The tumor heatmaps were generated by: (1) extracting the tissue region as described in Section 2.2 and generating image patches from the tissue region (see Figure 3A), (2) applying the trained patch-wise model in a sliding window fashion across the WSI, and (3) stitching the tumor class softmax prediction of each patch to form a tumor heatmap (see Figure 3B). Heatmaps were generated for all WSIs in the training and validation sets, as these were relevant for the second stage. Alternatively, if only a patch-wise model was used, these heatmaps were thresholded by 0.5 to produce segmentations of the tumor region. The latter represents the segmentation designs (III)-(V) mentioned in Section 3.2).

To improve the result from the patch-wise detection, we combined the heatmap with the low-resolution WSI (see Figure 3C). This was done in an additional stage using a Refinement CNN. A suitable magnification level was chosen ($\geq 1024 \times 1024$ pixels), and a low-resolution version of the original WSI was extracted from the image pyramid. The image was then normalized to $[0, 1]$, before both the resulting image and the heatmap were resized to 1024×1024 using bilinear interpolation. A separate fully-convolutional neural network was then used to refine the resulting heatmaps from the detection stage. We used the U-Net (10) architecture, which took the concatenated low-resolution three-channel WSI and predicted heatmap as input.

2.6. From development to deployment

After training the patch-wise and refinement models, the models are ready to be used for inference. The inference pipeline

is illustrated in Figure 3. The trained TensorFlow models were converted to the ONNX (42) standard format, to enable efficient inference on both graphics processing unit (GPU) and central processing unit (CPU) with different frameworks. The models were then integrated into the FastPathology (27) platform by writing a FAST (43) text pipeline, containing information about the inference pipeline and how the models should be handled (e.g., input shape, node names, and inference type). Thus, the proposed pipeline can be used through a graphical user interface (GUI) without programming. Binary release of FastPathology, trained models, test data, and source code can be accessed on GitHub³.

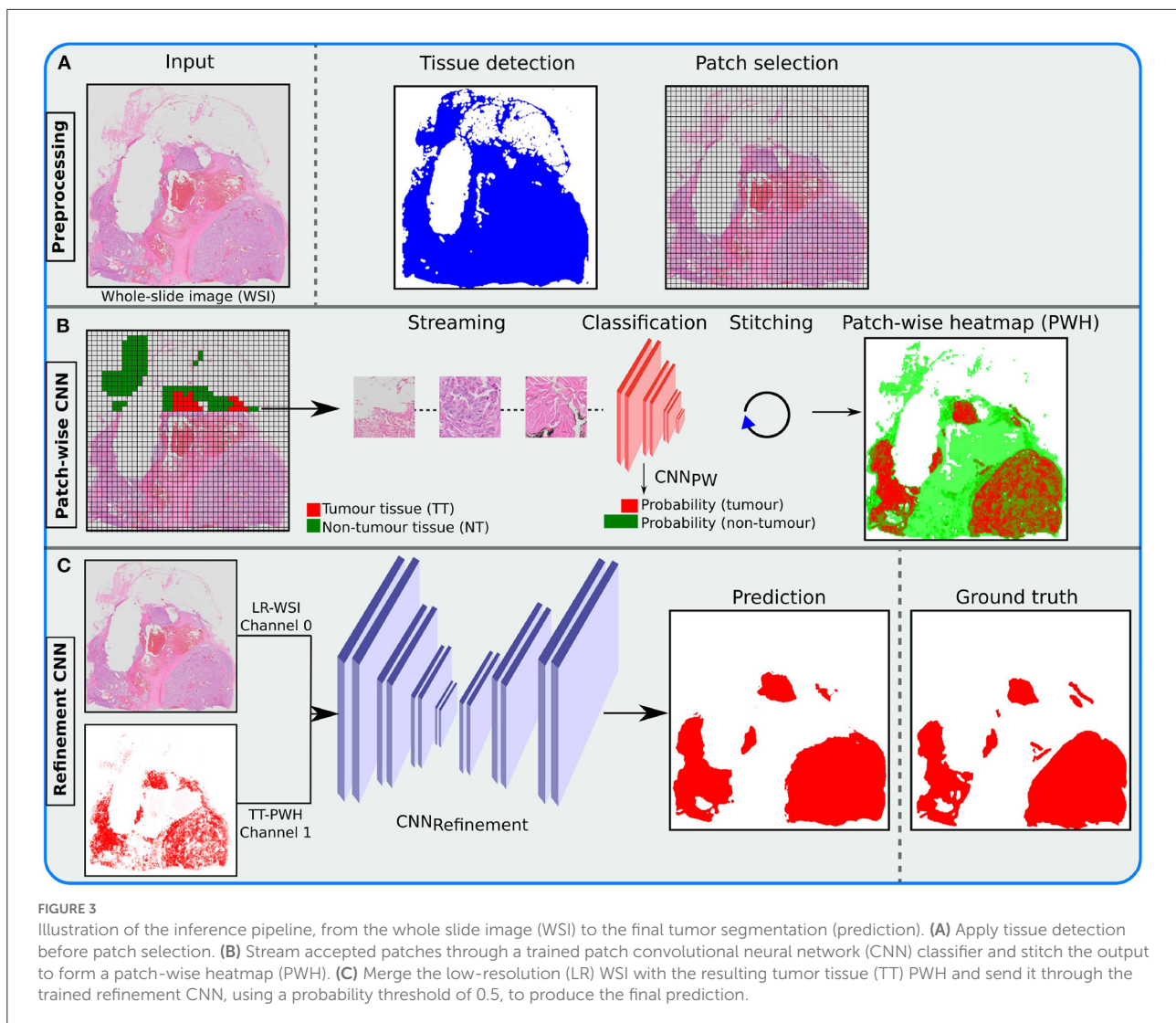
3. Validation study

3.1. Experiments

We conducted an ablation study to evaluate our design. The experiments conducted were:

- (1) To assess the importance of architecture complexity in breast cancer tumor detection in WSIs, we compared CNN classifiers using the two backbone architectures InceptionV3 and MobileNetV2.
- (2) To evaluate the cluster-guiding approach, we conducted experiments with and without k -means using the MobileNetV2 backbone.
- (3) To assess the effect of post-processing on the predicted heatmap, we compared state-of-the-art CAEs against simple baseline methods.
- (4) To evaluate the importance of having a GPU for inference, runtime measurements of the best performing method

³ <https://github.com/AICAN-Research/FAST-Pathology>



were performed with and without using the GPU, using TensorRT (44) and OpenVINO (45) for GPU and CPU inference, respectively.

3.2. Baseline segmentation methods

Using our pipeline, any architecture or component can be removed, added, or substituted. It is therefore valuable to assess the importance of each component in the pipeline. To evaluate the pipeline, we used existing, well-documented, state-of-the-art architectures. For patch-wise classification we used the MobileNetV2 (~ 2.39M params.) and InceptionV3 (~ 22.00M params.) backbones pretrained on the ImageNet data set, and used the same simplified classifier head for both architectures, as described in Section 2.3.

For image segmentation refinement, we compared the CAE architectures U-Net (~ 11.58M params.), AGU-Net (~ 7.68M params.), DAGU-Net (~ 9.99M params.), and DoubleU-Net (16.04M params.). In addition, we included a traditional, widely used tissue segmentation method (46), to serve as a minimal baseline measure. This method simply segments all tissue, and thus all tuned methods should outperform it. For this method, the image was resized to 1024×1024 , before being converted to the HSV color domain. Then, the saturation image was thresholded using Otsu's method (47).

The autoencoders were slightly modified to work better for our use case and data set. To make comparison fair, all autoencoders had similar depth and filter arrangement. Our U-Net architecture consisted of nine encoder and decoder blocks, with the following filters (from top to bottom): {8, 16, 32, 64, 128, 128, 256, 256, 512}. The encoder block used the following

operations (Convolution-BatchNormalization-ReLU) \times 2-MaxPooling. The AGU-Net, DAGU-Net, and DoubleU-Net architectures used the same filter arrangement as U-Net, but based on eight levels without the last 512 filter due to memory constraints. Furthermore, DoubleU-Net used 42 atrous spatial pyramid pooling filters. All autoencoders were symmetric, trained from scratch, modified to support two inputs (image and heatmap), and used softmax activation function in the last layer. Implementations of all architectures used in this study are openly available in our GitHub repository⁴.

In summary, the following segmentation designs were compared:

- (I) **Otsu**: Intensity-based thresholding for tissue segmentation.
- (II) **UNet-LR**: Segmentation of low-resolution WSI using a U-Net architecture.
- (III) **Inc-PW**: Patch-wise classification using an InceptionV3 architecture.
- (IV) **Mob-PW**: Patch-wise classification using a MobileNetV2 architecture.
- (V) **Mob-KM-PW**: Same as (IV), with k -means guiding.
- (VI) **Mob-PW-UNet**: Same as (IV), with a U-Net refinement network, without k -means guiding.
- (VII) **Mob-PW-AGUNet**: Same as (IV), with an AGU-Net refinement network.
- (VIII) **Mob-PW-DAGUNet**: Same as (IV), with a DAGU-Net refinement network.
- (IX) **Mob-PW-DoubleUNet**: Same as (IV), with a DoubleU-Net refinement network.

3.3. Statistical evaluation

All patch-wise and refinement models were trained using the same training set, and the best models were selected based on the performance on the validation set. The test set was used as a hold-out sample for an unbiased, final evaluation.

As our use case is on breast cancer segmentation, we only consider binary segmentation in this study. However, H2G-Net supports multiclass semantic segmentation. A threshold of 0.5 was used to distinguish between the tumor and non-tumor classes. Metrics were reported WSI-wise, and only on the test set. For each respective metric, macro average and standard deviation were reported. The specific metrics used to assess performance were pixel-wise recall, precision, and the Dice similarity coefficient (DSC). To further assess the robustness of the design, we also reported DSC for each histological grade. We performed multiple pairwise Tukey's range tests, comparing the DSC measures for all deep learning-based designs. The p -values were estimated for the test set (see [Supplementary Table 1](#)).

⁴ <https://github.com/andreped/H2G-Net>

3.4. Training parameters

For training the classification models, we fine-tuned the respective pretrained backbones using the Adam optimizer (48) with an initial learning rate of $1e-4$. For batch generation, 500 and 200 batches of size 64 for training and validation, respectively, were sampled randomly for each epoch. The models were trained for 100 epochs. Batches were generated in parallel using eight workers with a maximum queue size of 20. Models were trained using the following online data augmentation scheme of which all had a 50% chance of being used: random horizontal/vertical flip, 90° lossless rotations, HSV color augmentation with a random shift of range $[-20, 20]$, and multiplicative brightness augmentation of range $[0.8, 1.2]$.

All segmentation models were trained from scratch using the Adam optimizer with an initial learning rate of $1e-3$. Accumulated gradients using a batch size of four with six accumulation steps were performed. For online data augmentation, simple horizontal/vertical flip, 90° rotations, random zoom of range $[0.8, 1.2]$, and Macenko (49) stain augmentation⁵ using $\sigma_1 = \sigma_2 = 0.1$, with a chance of 50% of being used, were conducted. The models were trained for 1,000 epochs, or until the early stopping criterion with a patience of 100 epochs was achieved.

Implementation was done in Python 3.6, and CNN architectures were implemented in TensorFlow (v1.13.1). Experiments were performed using an Intel Xeon Silver 3110 CPU, with 32 cores and 2.10 GHz, and an NVIDIA Quadro P5000 dedicated GPU.

4. Results

For the test set, all deep learning-based methods outperformed the tissue segmentation method, Otsu, in terms of DSC (see [Table 1](#)). Comparing the patch-wise classifiers, Inc-PW and Mob-PW, no significant difference in DSC was found between the architectures ($p \approx 0.9$, see [Supplementary Table 1](#)). Adding cluster-guiding to Mob-PW, Mob-PW-KM, slightly reduced performance, however, not significantly ($p \approx 0.9$). Among the best single-resolution designs (i.e., UNet-LR, Inc-PW, and Mob-PW), the patch-wise approaches performed slightly better in terms of DSC, but not significantly ($p \approx 0.9$). The low-resolution approach (UNet-LR) achieved better recall, but with the cost of poorer precision.

Introducing a U-Net-inspired refinement network (using both the low-resolution WSI and the resulting heatmap from Mob-PW) resulted in significant improvement compared to the best single-resolution approach (Mob-PW-UNet vs. Inc-PW, $p \approx 0.012$). All methods using a refinement network significantly outperformed the single-resolution approaches.

⁵ <https://github.com/Peter554/StainTools>

TABLE 1 Test set segmentation performance for the different designs.

Designs	Recall	Precision	Dice similarity coefficient
(I) Otsu	0.990 ± 0.027	0.534 ± 0.200	0.669 ± 0.179
(II) UNet-LR	0.931 ± 0.113	0.851 ± 0.165	0.874 ± 0.128
(III) Inc-PW	0.881 ± 0.118	0.909 ± 0.099	0.887 ± 0.089
(IV) Mob-PW	0.879 ± 0.123	0.907 ± 0.100	0.885 ± 0.094
(V) Mob-KM-PW	0.853 ± 0.124	0.909 ± 0.097	0.872 ± 0.092
(VI) Mob-PW-UNet	0.944 ± 0.074	0.929 ± 0.088	0.933 ± 0.069
(VII) Mob-PW-AGUNet	0.954 ± 0.066	0.909 ± 0.097	0.927 ± 0.072
(VIII) Mob-PW-DAGUNet	0.942 ± 0.075	0.922 ± 0.091	0.928 ± 0.072
(IX) Mob-PW-DoubleUNet	0.949 ± 0.073	0.919 ± 0.093	0.929 ± 0.074

DSC, Dice Similarity Coefficient; Inc, InceptionV3; Mob, MobileNetV2; PW, Patch-wise; KM, *k*-means; LR, Low-resolution. Results are reported as mean ± standard deviation.

Comparing the refinement architectures, the best performance in terms of precision and DSC was found from the U-Net design, Mob-PW-UNet, but the difference was not statistically significant ($p \approx 0.9$ for all comparisons). No benefit of using more advanced CAE architectures was found.

When each histological grade was analyzed separately, Mob-PW-DoubleUNet was the best performing method for grade I and Mob-PW-UNet performed best on grade II and III (see Table 2). All designs guided by the hierarchical sampling scheme (designs (II)-(IX)) performed similarly across all histological grades, indicating that performance was somewhat invariant to histological grade.

The best performing method, Mob-PW-UNet, took approximately ~ 58 s to run on a representative $\times 400$ WSI using the CPU (see Table 3). Using the GPU, runtime was reduced to 40.26 s. The patch-wise method dominated the overall runtime, with $\sim 1\%$ of the total runtime being used on the refinement stage.

5. Discussion

In this paper, a novel approach for breast cancer segmentation in WSIs was proposed. The method, called H2G-Net, includes a cascaded CNN architecture combining high-resolution and global information with negligible increase in runtime. A random sampling scheme was also proposed, which extracts patches directly from the raw WSI format, handling multiple data imbalance challenges simultaneously. We made the pipeline and trained models openly available in FastPathology (27), enabling the use of the method without programming. To conduct our experiments, we developed a novel data set comprising 624 breast cancer WSIs annotated by pathologists. We have presented each component in the pipeline and assessed the impact of each component in an ablation study. Using multiple guiding components, we significantly

improved segmentation performance, while reducing disk storage requirements compared to conventional training pipelines.

The best performing architectures utilized both low and high-resolution information from the WSI. A similar approach is used by pathologists when separating the tumor from surrounding tissue. The low-resolution image provides a coarse outline of the tumor, whereas higher resolution is often necessary for accurate delineation.

5.1. Cascaded design and related work

For segmentation, using low-resolution as the first step could reduce total runtime by filtering patches during preprocessing. However, UNet-LR, a U-Net using only low-resolution information, results in low sensitivity and should therefore not be the first step for breast cancer segmentation. In this work, we used a patch-wise, high-resolution method as a first step to optimize detection. A U-Net could then be trained at a later stage to refine the produced heatmap, using both the heatmap and the low-resolution WSI as input. Using the heatmap generated from the patch-wise method alone, some areas of the tumor, such as areas with abundant stromal or fatty tissue, may not be recognized, thus resulting in a fragmented heatmap (see Figure 4). We show that the network benefits from having the low-resolution image, together with the heatmap. Using this approach, the segmentation become more similar to the ground truth (the pathologists' annotations).

Tang et al. (50) used a two-step procedure to perform instance segmentation of objects in the Cityscapes data set (51). These images have an initial resolution of 1024×2048 pixels, but the classes of interest can easily be distinguished at lower resolution. In contrast to our design, they first performed semantic segmentation on the low-resolution, before refining the initial segmentation using a patch-wise design. They also used overlapping predictions along the border of the initial segmentation. Their approach might be an interesting refinement method to further improve the segmentation performance of our resulting low-resolution segmentation.

A similar refinement approach to (50) was used by (52) for cervical intraepithelial neoplasia segmentation of H&E stained WSIs. However, introducing a new network for border refinement will make the overall runtime longer and introduce more complexity to the final pipeline. Isensee et al. (53) conducted a similar two-step procedure for medical volumetric data. They first used a CAE applied on the downsampled version of the full 3D volume (CT/MRI). They then applied a 3D patch-wise refinement model, using both the local volumetric data (CT/MRI) and predicted heatmap as input.

Another similar architecture design to ours, was proposed by (54) for WSI classification of breast cancer. They also used a patch-wise model in the first step, before feeding the

TABLE 2 Test set segmentation performance for the different designs in histological grades I-III.

Designs		Dice similarity coefficient ($n = 90$)		
		Grade I (11)	Grade II (48)	Grade III (31)
(I)	Otsu	0.732 ± 0.151	0.659 ± 0.186	0.664 ± 0.174
(II)	UNet-LR	0.880 ± 0.127	0.862 ± 0.142	0.890 ± 0.099
(III)	Inc-PW	0.901 ± 0.072	0.882 ± 0.088	0.890 ± 0.095
(IV)	Mob-PW	0.887 ± 0.089	0.882 ± 0.092	0.890 ± 0.100
(V)	Mob-KM-PW	0.851 ± 0.111	0.872 ± 0.089	0.880 ± 0.088
(VI)	Mob-PW-UNet	0.936 ± 0.073	0.931 ± 0.058	0.935 ± 0.083
(VII)	Mob-PW-AGUNet	0.933 ± 0.082	0.926 ± 0.060	0.927 ± 0.083
(VIII)	Mob-PW-DAGUNet	0.935 ± 0.075	0.926 ± 0.058	0.929 ± 0.088
(IX)	Mob-PW-DoubleUNet	0.942 ± 0.070	0.924 ± 0.066	0.934 ± 0.085

DSC, Dice Similarity Coefficient; Inc, InceptionV3; Mob, MobileNetV2; PW, Patch-wise; KM, *k*-means; LR, Low-resolution. Results are reported as mean ± standard deviation.

TABLE 3 Runtime measurements of the proposed method, Mob-PW-UNet.

	Patch-wise	Refinement	Total
OpenVINO	57.32 ± 0.20	0.75 ± 0.01	58.07 ± 0.20
TensorRT	39.88 ± 0.62	0.38 ± 0.00	40.26 ± 0.62

Experiments were repeated ten times and respective average and standard deviation are reported in seconds. OpenVINO and TensorRT were used as inference engines for CPU and GPU inference, respectively.

resulting heatmap to a second CNN that performed WSI-level classification. In addition, they used skip connections to propagate learned features from the patch-wise CNN to the latter CNN. Thus, our method could be seen as an adaption to their design applied to image segmentation. This style of skip connection is similar to the DoubleU-Net approach by (26). In this study, we did not explore skipping features from the classifier to the refinement network. This could be explored in future work. Our design is also similar to the work of (55), where a similar two-stage, cascaded CNN design was deployed, but for image registration of WSIs.

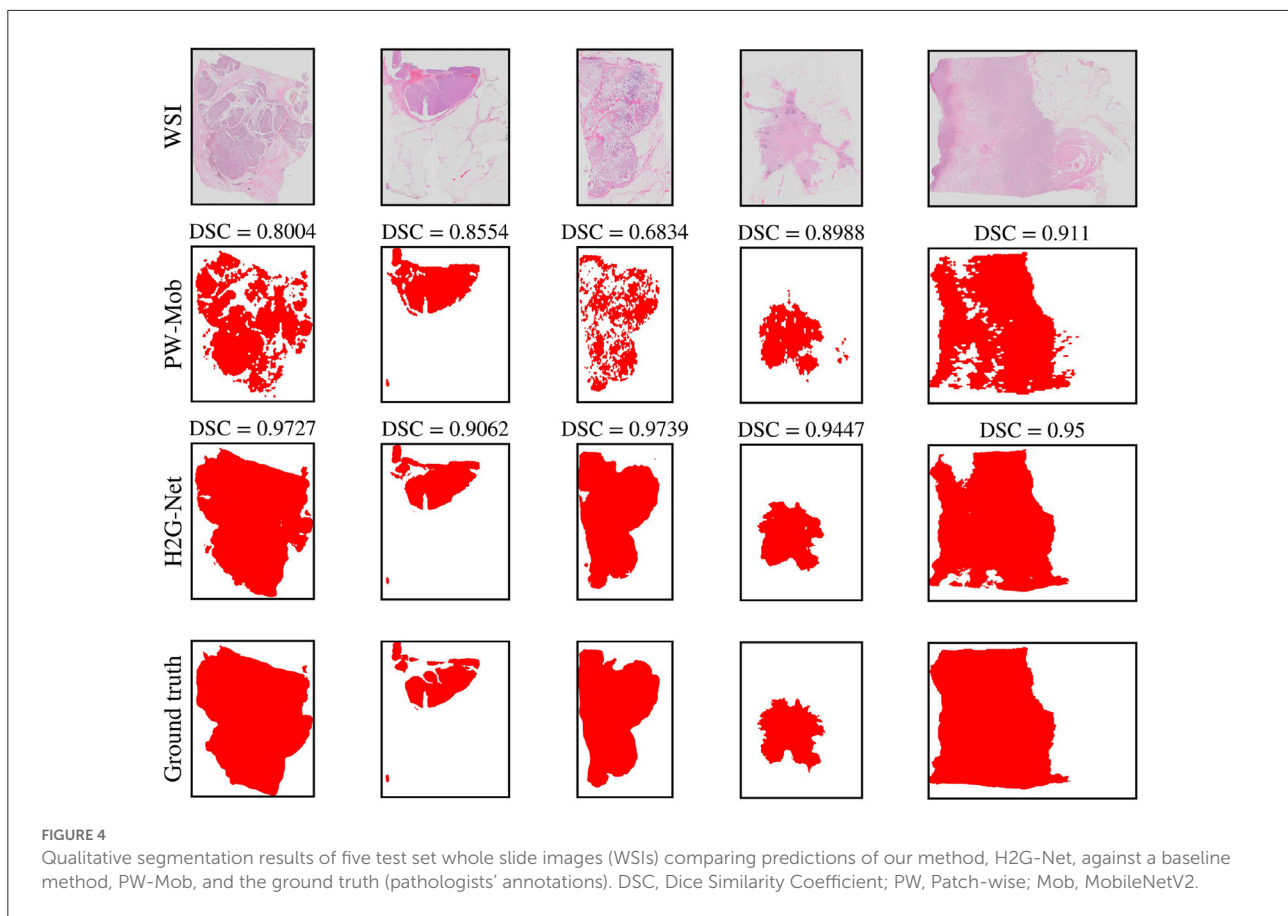
5.2. State-of-the-art comparison

An alternative approach, commonly used in the literature for other segmentation tasks (3), is to perform segmentation on patch-level. We argue that this approach is too slow and memory intensive for routine usage on low-end devices. As tumor segmentation is often the first step in any automatic cancer assessment pipeline, the method must be rapid. Furthermore, the annotations we used for training marked the outline of the tumor, thus including components like stromal tissue, inflammatory cells, and blood vessels, in addition to invasive epithelial cells. Within the tumor area, the different patches

will have a varying morphological appearance. Therefore, using a patch-wise classifier alone, some patches may not contain sufficient information to make a qualified decision. This is even more challenging for semantic segmentation. Therefore, comparing our design to a high-resolution patch-wise segmentation models was not explored. The natural alternative state-of-the-art method could be to apply a patch-wise CNN classifier, which was the best performing method used for tumor detection and classification in the BACH challenge (4). We therefore performed an ablation study to make our design more easily comparable to current state-of-the-art approaches.

Comparison to state-of-the-art methods is challenging, as the data others have used are often not publicly available, may consist of other organs and cancer types, or have been annotated in a different manner. The BACH challenge data set includes 100 image patches from only ten annotated WSIs of invasive carcinomas. Another public data set is the CAMELYON17 (56) challenge data set comprising 1,399 annotated lymph node metastasis sections, where the task and tissue type are not comparable to our data. Lastly, the PAIP 2019 (57) challenge data set includes annotations made for the whole tumor area, using annotations similar to ours. The best performing team in the PAIP 2019 challenge used a patch-wise CAE and reached a Jaccard index of 0.7890, which translates to a DSC of 0.8821. This performance is comparable to the single-resolution approaches we tested (see Table 1). However, the sections they used were from liver cancer tissue. The training data set included only 50 annotated WSIs, in contrast to our data set, including 624 WSIs. Hence, a fair comparison cannot be made.

Cruz-Roa et al. (58) used a local data set of ~ 500 annotated WSIs from breast cancer for training and validating a patch-wise CNN classifier for detecting invasive breast cancer. They also proposed a sampling scheme based on Quasi-Monte Carlo sampling similar to ours, however, a heatmap refinement method was not explored. On their local validation set they achieved a DSC of 0.670, and on an independent test set of



195 WSIs from the The Cancer Genome Atlas (TCGA), they achieved a DSC of 0.759. Le et al. (59) used a local data set of 109 annotated WSIs from breast cancer to train and validate a patch-wise CNN classifier. They also used the same TCGA data set as (58) for evaluation, but used a ResNet-34 CNN backbone and a patch aggregation refinement method based on a sliding window pooling technique. They reported a DSC of 0.820.

As our best performing method is integrated into an open software (FastPathology), it is possible for researchers to make comparisons of their methods to ours, by running our pipeline on their own data. However, it is challenging to compare machine learning methods if they have not been trained on the same data, as machine learning models are a consequence of the training data. A large, public benchmark data set for breast cancer segmentation is therefore needed.

5.3. Architecture depth, patch generation, and clustering

An interesting observation in this study is that using the deepest and most complex network for tumor segmentation is not necessarily better. From Table 1, we found no statistical significant difference between using InceptionV3 and

MobileNetV2 for detection of breast cancer tissue. A similar trend could be seen from the refinement network. Choosing more complex CAEs did not significantly improve segmentation performance. This could be due to data that did not cover all possible variations. The quality of the heatmap provided from the detection stage varied in some cases, making it challenging for the refinement network to improve the initial segmentation.

Reading patches from the raw WSI format is time consuming. It is therefore common to preprocess data before training. In this study, we sampled patches directly from the raw WSI format during training. This idea was recently proposed by (60). We further extended on their idea to make it more generic. The approach by (60) cannot handle larger batch sizes, as the cost of batch generation is not scalable. Thus, we used accumulated gradients to speed up batch generation, while simultaneously reducing GPU memory usage. We further introduced the concept of hierarchical sampling, which added direct support for balancing on multiple categories and labels. This design also added support for performing cluster-balancing end-to-end during training. The clustering method can also easily be substituted.

No benefit from using cluster-guiding to detect breast cancer tissue was observed, comparing Mob-KM-PW and Mob-PW by qualitative visual inspection. This was also observed in a study

by (13) on a similar task. We hypothesize that the core reason why cluster-guiding did not improve performance in our use case may be that the fundamental problem of breast region segmentation does not necessarily lie in handling data imbalance between tissue types. Due to the low field-of-view in patch-wise designs, there will most likely be patches where it is not possible to assign the correct label without also including global context. That is probably why our refinement network was more successful.

5.4. Future perspectives

Kaiser et al. (13) demonstrated improved performance by using a more advanced clustering design. Thus, in future work, substituting the traditional ImageNet features + PCA + k -means clustering approach with a more suited clustering design should be explored. A natural next step could then be to provide the predicted cluster heatmap with the tumor heatmap, as it would provide different, representation-informative, high-resolution information to the refinement network.

Multiple instance learning (MIL) is a promising approach that tackles the challenge of weak supervision and noisy ground truth (61). Exchanging the MIL design with the single-instance CNN classifier is possible. In this framework, one could still perform clustering in preprocessing, and sample patches to the bag, as done in a recent study (37). However, an interesting approach proposed by (62) was to perform clustering directly within the MIL design on bag-level. Using attention, one could train a network, not only to solve a task, but to learn subcategory structures in the data, while simultaneously filtering redundant clusters and noisy patches. This was demonstrated by (63). However, they did not assess the impact of the clustering component. Future work should involve replacing the single-instance CNN with the MIL design, incorporating clustering in an end-to-end fashion, and properly assessing its impact.

5.5. Strengths and limitations

The main strengths of the study are that the models were trained on a large set of breast cancer WSIs. Tumor annotations were created in a (semi-)automatic manner, and manually corrected by pathologists. To ensure consistency, all annotations were assessed by a pathologist experienced in breast cancer pathology. Validation studies were conducted using an independent test set. The performance of the different designs was also evaluated for each histological grade separately. An ablation study was performed to assess the impact of each component in the multi-step pipeline. The proposed design was validated against baseline methods, and the best method was integrated in an open platform, FastPathology (27).

The main limitations are that the models were trained using sections that were H&E-stained in the same laboratory and

scanned in a single scanner. We demonstrated that the model generalized well to the test set, however, we have not tested our model on WSIs from other institutions. It is possible to carry out data augmentation to make the models more invariant, but it is challenging to mimic different staining and scanning effects (64). Thus, in the future, data from different laboratories and scanners will be added for training the models. We did not perform stain normalization as it would have added an additional layer of uncertainty and dependency in the pipeline. Furthermore, it would be interesting to assess the extent of generalization capability of our models to cancers of other origins, such as lung or gastrointestinal cancer. Lastly, our hierarchically-balanced sampling scheme was only used on a single task and not included in the ablation study.

6. Conclusion

Through our hybrid guiding scheme, we demonstrated a significant improvement in segmentation of breast cancer tumors from gigapixel histopathology images. The model outperformed single-resolution approaches and introduced a simple, fast, and accurate way to refine segmentation heatmaps, without the need for overlapping inference or ensembling. We also presented a hierarchical sampling scheme, that enabled patches to be streamed from the raw WSI format concurrently during training. Furthermore, we demonstrated that tissue type balancing can be performed end-to-end, using a novel loss function. The hierarchical sampling scheme and the novel loss function were introduced to make training methods more scalable and to reduce storage requirements.

Data availability statement

The data sets presented in this article are not readily available because the ethical approval does not allow for making the data set publicly available. Requests to access the data sets should be directed to Marit Valla, marit.valla@ntnu.no. The source code to reproduce the experiments is made available at <https://github.com/andrepd/H2G-Net>. The best-performing method is integrated into an open software, FastPathology, which can be downloaded at <https://github.com/AICAN-Research/FAST-Pathology>.

Ethics statement

The studies involving human participants were reviewed and approved by the Central Norway Regional Committee for Medical and Health Research Ethics (reference numbers 2018/2141 and 836/2009). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

AP: conceptualization, methodology, investigation, annotation, writing—original draft, review, and editing, software, and validation. ES: methodology, investigation, supervision, writing—original draft, review, and editing, and software. TR, VD, and HP: annotation, writing—review and editing, and pathology expertise. DB: methodology, supervision, and writing—original draft, review, and editing. T-AN: methodology, investigation, and writing—review and editing. IR: conceptualization, methodology, supervision, and writing—original draft, review, and editing. MV: conceptualization, methodology, investigation, annotation, supervision, data curation, and writing—original draft, review, and editing, resources, pathology expertise, and project administration. All authors reviewed and approved the final version of the manuscript.

Funding

This work was supported by The Liaison Committee for Education, Research and Innovation in Central Norway [Grant Number 2018/42794]; The Joint Research Committee between St. Olavs Hospital and the Faculty of Medicine and Health Sciences, NTNU (FFU) [Grant Number 2019/38882]; The Cancer Foundation, St. Olavs Hospital, Trondheim University

References

1. Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Pineros M, et al. *Global Cancer Observatory: Cancer Today*. Lyon: International Agency for Research on Cancer (2020).
2. The Royal College of Pathologists. *Meeting Pathology Demand: Histopathology Workforce Census* (2018).
3. Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: a survey. *Med Image Anal.* (2021) 67:101813. doi: 10.1016/j.media.2020.101813
4. Aresta G, Arajo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, et al. BACH: grand challenge on breast cancer histology images. *Med Image Anal.* (2019) 56:122–39. doi: 10.1016/j.media.2019.05.010
5. Priego Torres B, Morillo D, Granero MA, Garcia-Rojo M. Automatic segmentation of whole-slide H&E stained breast histopathology images using a deep convolutional neural network architecture. *Expert Syst Appl.* (2020) 151:113387. doi: 10.1016/j.eswa.2020.113387
6. Xu J, Janowczyk A, Chandran S, Madabhushi A. A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: applications to histopathology image segmentation. In: *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. (2010). p.7623. doi: 10.1117/12.845602
7. Zhou J, Ruan J, Wu C, Ye G, Zhu Z, Yue J, et al. Superpixel segmentation of breast cancer pathology images based on features extracted from the autoencoder. In: *2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN)*. (2019). p. 366–70. doi: 10.1109/ICCSN.2019.8905358
8. Ehteshami Bejnordi B, Veta M, P J, Beca F, Albarqouni S, Cetin-Atalay R, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* (2017) 318:2199–210. doi: 10.1001/jama.2017.14580
9. Schmitz R, Madesta F, Nielsen M, Krause J, Steurer S, Werner R, et al. Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. *Med Image Anal.* (2021) 70:101996. doi: 10.1016/j.media.2021.101996

Hospital [Grant Number 13/2021]; and The Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital [Grant Number 2020/14728-49].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmed.2022.971873/full#supplementary-material>

10. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical Image Computing and Computer-Assisted Intervention-MICCAI*. Cham: Springer International Publishing (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28
11. Guo Z, Liu H, Ni H, Wang X, Su M, Guo W, et al. A fast and refined cancer regions segmentation framework in whole-slide breast pathological images. *Sci Rep.* (2019) 9:882. doi: 10.1038/s41598-018-37492-9
12. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV (2016). doi: 10.1109/CVPR.2016.308
13. Qaiser T, Tsang YW, Taniyama D, Sakamoto N, Nakane K, Epstein D, et al. Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Med Image Anal.* (2019) 55:1–14. doi: 10.1016/j.media.2019.03.014
14. Steinhaus H. Sur la division des corps materiels en parties. *Bulletin de l'Academie Polonaise des Sciences.* (1956) 12:801–4.
15. Yao J, Zhu X, Huang J. Deep multi-instance learning for survival prediction from whole slide images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer (2019). p. 496–504. doi: 10.1007/978-3-030-32239-7_55
16. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *International Conference on Learning Representations*. (2015).
17. Karl Pearson FRS. LIII. On lines and planes of closest fit to systems of points in space. *Lond Edinburgh Dublin Philos Mag J Sci.* (1901) 2:559–72. doi: 10.1080/14786440109462720
18. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2018). p. 4510–20. doi: 10.1109/CVPR.2018.00474

19. Kassani SH, Kassani PH, Wesolowski MJ, Schneider KA, Deters R. Classification of histopathological biopsy images using ensemble of deep learning networks. In: *Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*. CASCON '19. IBM Corp. (2019). p. 92–9.
20. Skrede OJ, De Raedt S, Kleppe A, Hveem T, Liestel K, Maddison J, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet*. (2020) 395:350–60. doi: 10.1016/S0140-6736(19)32998-8
21. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal*. (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
22. Nie D, Wang L, Trullo R, Li J, Yuan P, Xia J. Segmentation of craniomaxillofacial bony structures from MRI with a 3D deep-learning based cascade framework. In: *International Workshop on Machine Learning in Medical Imaging*. (2017). p. 266–73. doi: 10.1007/978-3-319-67389-9_31
23. Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. *IEEE Trans Pattern Anal Mach Intell*. (2021) 44:3523–42. doi: 10.1109/TPAMI.2021.3059968
24. Bouget D, Pedersen A, Hosainey SAM, Solheim O, Reinertsen I. Meningioma segmentation in T1-weighted MRI leveraging global context and attention mechanisms. *Front Radiol*. (2021) 1:711514. doi: 10.3389/fradi.2021.711514
25. Painchaud N, Skandarani Y, Judge T, Bernard O, Lalonde A, Jodoin PM. Cardiac MRI segmentation with strong anatomical guarantees. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, et al., editors. *Medical Image Computing and Computer Assisted Intervention-MICCAI*. Cham: Springer International Publishing (2019). p. 632–40. doi: 10.1007/978-3-030-32245-8_70
26. Jha D, Riegler M, Johansen D, Halvorsen P, Johansen H. DoubleU-Net: a deep convolutional neural network for medical image segmentation. *arXiv preprint arXiv:2006.04868* (2020). doi: 10.1109/CBMS49503.2020.00111
27. Pedersen A, Valla M, Bofin AM, De Frutos JP, Reinertsen I, Smistad E. FastPathology: an open-source platform for deep learning-based research and decision support in digital pathology. *IEEE Access*. (2021) 9:58216–29. doi: 10.1109/ACCESS.2021.3072231
28. Engstrom M, Opdahl S, Hagen A, Romundstad P, Akslen LA, Haugen O, et al. Molecular subtypes, histopathological grade and survival in a historic cohort of breast cancer patients. *Breast Cancer Res Treat*. (2013) 140:463–73. doi: 10.1007/s10549-013-2647-2
29. Elston C, Ellis I. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*. (1991) 19:403–10. doi: 10.1111/j.1365-2559.1991.tb00229.x
30. Bankhead P, Loughrey M, Fernandez J, Dombrowski Y, Mcart D, Dunne P, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. (2017) 7:16878. doi: 10.1038/s41598-017-17204-5
31. Hanahan D, Weinberg R. The hallmarks of cancer. *Cell*. (2000) 100:57–70. doi: 10.1016/S0092-8674(00)81683-9
32. Hanahan D. Hallmarks of cancer: new dimensions. *Cancer Discovery*. (2022) 12:31–46. doi: 10.1158/2159-8290.CD-21-1059
33. Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. *Cell*. (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
34. Carse J, McKenna S. Active learning for patch-based digital pathology using convolutional neural networks to reduce annotation costs. In: *European Congress on Digital Pathology*. (2019). p. 20–7. doi: 10.1007/978-3-030-23937-4_3
35. Satyanarayanan M, Goode A, Gilbert B, Harkes J, Jukic D. OpenSlide: a vendor-neutral software foundation for digital pathology. *J Pathol Inform*. (2013) 4:27. doi: 10.4103/2153-3539.119005
36. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE (2009). p. 248–55. doi: 10.1109/CVPR.2009.5206848
37. Yao J, Zhu X, Jonnagaddala J, Hawkins N, Huang J. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med Image Anal*. (2020) 65:101789. doi: 10.1016/j.media.2020.101789
38. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. (2011) 12:2825–830. doi: 10.5555/1953048.2078195
39. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. (2015). Available online at: tensorflow.org
40. Srivastava N, Hinton GE, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. (2014) 15:1929–58. doi: 10.5555/2627435.2670313
41. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37. ICML'15*. (2015). p. 448–56.
42. Bai J, Lu F, Zhang K. ONNX: Open Neural Network Exchange. (2019).
43. Smistad E, Astvik A, Pedersen A. High performance neural network inference, streaming and visualization of medical images using FAST. *IEEE Access*. (2019) 7:136310–21. doi: 10.1109/ACCESS.2019.2942441
44. NVIDIA. *TensorRT* (2019).
45. Intel. *OpenVINO Toolkit* (2019).
46. Bandi P, Balkenhol M, Ginneken B, Laak J, Litjens G. Resolution-agnostic tissue segmentation in whole-slide histopathology images with convolutional neural networks. *PeerJ*. (2019) 7:e8242. doi: 10.7717/peerj.8242
47. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybernet*. (1979) 9:62–66. doi: 10.1109/TSMC.1979.4310076
48. Kingma D, Ba J. “ADAM: a method for stochastic optimization. In: *International Conference on Learning Representations*. (2014).
49. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al. A method for normalizing histology slides for quantitative analysis. In: *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. (2009). p. 1107–10. doi: 10.1109/ISBI.2009.5193250
50. Tang C, Chen H, Li X, Li J, Zhang Z, Hu X. Look closer to segment better: boundary patch refinement for instance segmentation. *arXiv preprint arXiv:2104.05239* (2021). doi: 10.1109/CVPR46437.2021.01371
51. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, et al. The cityscapes dataset for semantic urban scene understanding. In: *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2016). doi: 10.1109/CVPR.2016.350
52. Sornapudi S, Addanki R, Stanley R, Stoecker W, Long L, Zuna R, et al. *Cervical Whole Slide Histology Image Analysis Toolbox*. (2020). doi: 10.1101/2020.07.22.20160366
53. Isensee F, Jaeger P, Kohl S, Petersen J, Maier-Hein K. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:1–9. doi: 10.1038/s41592-020-01008-z
54. Nazeri K, Aminpour A, Ebrahimi M. Two-stage convolutional neural network for breast cancer histology image classification. *arXiv preprint arXiv:1803.04054* (2018). doi: 10.1007/978-3-319-93000-8_81
55. Daly A, Geras K, Bonneau R. A convolutional neural network for common coordinate registration of high-resolution histology images. *bioRxiv*. (2020). doi: 10.1101/2020.09.18.303875
56. Bandi P, Geessink O, Manson Q, Van Dijk M, Balkenhol M, Hermsen M, et al. From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. *IEEE Trans Med Imaging*. (2019) 38:550–60. doi: 10.1109/TMI.2018.2867350
57. Kim YJ, Jang H, Lee K, Park S, Min SG, Hong C, et al. PAIP 2019: liver cancer segmentation challenge. *Med Image Anal*. (2021) 67:101854. doi: 10.1016/j.media.2020.101854
58. Cruz-Roa A, Gilmore H, Basavanthally A, Feldman M, Ganesan S, Shih N, et al. High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: application to invasive breast cancer detection. *PLoS ONE*. (2018) 13:e0196828. doi: 10.1371/journal.pone.0196828
59. Le H, Gupta R, Hou L, Abousamra S, Fassler D, Torre-Healy L, et al. Utilizing automated breast cancer detection to identify spatial distributions of tumor-infiltrating lymphocytes in invasive breast cancer. *Am J Pathol*. (2020) 190:1491–504. doi: 10.1016/j.ajpath.2020.03.012
60. Lutnick B, Murali LK, Ginley B, Rosenberg AZ, Sarder P. Histo-fetch-On-the-fly processing of gigapixel whole slide images simplifies and speeds neural network training. *arXiv preprint arXiv:2102.11433* (2021). doi: 10.48550/arXiv.2102.11433
61. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. In: Dy J, Krause A, editors. *Proceedings of the 35th International Conference on Machine Learning*. (2018). p. 2127–36.
62. Xu Y, Zhu JY, Chang E, Tu Z. Multiple clustered instance learning for histopathology cancer image classification, segmentation and clustering. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. (2012). p. 964–71.
63. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. (2021) 5:555–70. doi: 10.1038/s41551-020-00682-w
64. Tellez D, Litjens G, Bandi P, Bulten W, Bokhorst JM, Ciompi F, et al. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Med Image Anal*. (2019) 58:101544. doi: 10.1016/j.media.2019.10.1544