International Journal of Medical Informatics 159 (2022) 104669

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Deep transfer learning based model for colorectal cancer histopathology segmentation: A comparative study of deep pre-trained models

Sara Hosseinzadeh Kassani [a,*], Peyman Hosseinzadeh Kassani [b], Michal J. Wesolowski [c], Kevin A. Schneider [a], Ralph Deters [a]

[a] Department of Computer Science, University of Saskatchewan, Canada
[b] Department of Neurology and Neurological, University of Stanford, United States
[c] Department of Medical Imaging, University of Saskatchewan, Canada

ARTICLE INFO

Keywords:
Colorectal cancer
Deep learning
Transfer learning
Convolutional neural networks

ABSTRACT

Colorectal cancer is one of the leading causes of cancer-related death, worldwide. Early detection of suspicious tissues can significantly improve the survival rate. In this study, the performance of a wide variety of deep learning-based architectures is evaluated for automatic tumor segmentation of colorectal tissue samples. The proposed approach highlights the utility of incorporating convolutional neural network modules and transfer learning in the encoder part of a segmentation architecture for histopathology image analysis. A comparative and extensive experiment was conducted on a challenging histopathological segmentation task to demonstrate the effectiveness of incorporating deep modules in the segmentation encoder-decoder network as well as the contributions of its components. Experimental results demonstrate that shared DenseNet and LinkNet architecture is promising, achieves the state-of-the-art performance, and outperforms other methods with a dice similarity index of 82.74%±1.77, accuracy of 87.07%±1.56, and f1-score value of 82.79%±1.79.

1. Introduction

Colorectal cancer (CRC), also known as colon or bowel cancer, develops from the abnormal or excessive growth of malignant cells in the colon or rectum. CRC is the third most frequent cancer-related death in the United States, in both men and women, after lung cancer and breast cancer [1]. According to the annual report provided by the American Cancer Society [2], approximately 104,610 new cases of colon cancer and 43,340 new cases of rectal cancer will be diagnosed in 2020. Additionally, 51,020 patients are likely to die from colorectal cancer during 2020 in the United States. Most colorectal cancers start as abnormal tissue or benign polyps that grow in the inner linings of the intestine. Although the majority of abnormal tissue or polyps are initially benign, they can become malignant over time if left untreated [3]. Therefore, due to the high incidence and mortality rate of colorectal cancer [4], detecting and removing suspicious tissues in early stages is important to prevent the risk of developing colorectal cancer. Colonoscopy is the reference method for screening and detecting polyps inside the colon. Screening of polyps or other types of abnormal tissue in colonoscopy images or videos depends on the experiences of endoscopists. When the visual inspection of a polyp suggests that it may be malignant, it is recommended to remove the polyp for pathological analysis and determine whether malignancy [5].

Histopathology is the examination of thin sections of suspicious tissue through a microscope. The extracted tissues have been fixed onto glass slides and stained to reveal structures and morphological features. With the availability of whole slide scanning devices, there is a growing trend towards acquiring digitized pathology images of glass histology slides. The advent of digital pathology has enabled the application of existing image analysis techniques in the area of tissue histopathology. This has led to the rise of computational pathology and the possibility of using machine learning to aid the pathologist in diagnostic decision making. In a conventional pathological diagnostic practice, a pathologist has to analyze thousands of tissue sections on glass slide series, under the microscope, and annotate structures and morphological features. The manual segmentation of tumor regions is a laborious, challenging, and time-consuming task. Furthermore, visual examination of tissue slides is tedious when workloads are high. The traditional

* Corresponding author.
E-mail addresses: sara.kassani@usask.ca (S. Hosseinzadeh Kassani), peymanhk@stanford.edu (P. Hosseinzadeh Kassani), mike.wesolowski@usask.ca (M.J. Wesolowski), kevin.schneider@usask.ca (K.A. Schneider), deters@cs.usask.ca (R. Deters).

https://doi.org/10.1016/j.ijmedinf.2021.104669
Received 25 September 2021; Received in revised form 28 November 2021; Accepted 16 December 2021
Available online 31 December 2021
1386-5056/© 2021 Elsevier B.V. All rights reserved.

approach can also be biased due to the subjective nature of the task, which can lead to inter and intra-observer variability [6,7].

In contrast to the manual approaches, pathologists could benefit from the automation of repetitive tasks such as segmentation, classification and object detection. Accurate segmentation of structures such as cancerous regions, nuclei and glands are of crucial importance for a pathologist in assessing the degree of cancer malignancy and localization of tumor regions in Hematoxylin and Eosin (H&E)-stained slides. Computer-assisted diagnosis systems and quantitative image analysis could help to improve tissue analysis for image-based bio-marker discovery and tissue diagnosis with reproducible results. Given the increasing number of CRC cases and shortcomings of the conventional diagnosis system, the demand for developing precise and reliable segmentation algorithms is increasing [8,9]. Accurate segmentation, as well as, grading and scoring the aggressiveness of cancerous regions using automatic computer-aided diagnosis systems could greatly assist pathologists in accelerating the diagnostic process (see Fig. 1).

In this study, a number of convolutional neural network (CNN) architectures are investigated to determine the quality of their segmentation performance. A new transfer learning-based fusion approach, in the encoder part of segmentation backbones, is proposed to train CRC tissue segmentation. To do so, different architectures were designed so that each of them assembles multiple pre-trained CNN feature extractors on three segmentation architectures, namely U-Net, LinkNet and FPN. The robustness of 17 deep feature extractors belonging to 7 architecture families were investigated [10]. Finally, a comprehensive analysis comparing the proposed transfer learning approach on different CNN models is provided.

2. Related work

Chen et al. [11] presented a deep learning model and image processing methods for nuclei detection and segmentation from microscopy images. A multi-layer convolutional neural network was employed for feature extraction from both spatial and color information. The proposed U-Net model has been implemented to be trained on H&E-stained microscopy images containing seven different tissue samples for segmenting the boundary and detecting the geometric center of the nuclei. Kainz et al. [12] proposed a CNN-based model for gland segmentation in H&E stained histopathological images of colorectal cancer. In the proposed architecture, the learned gland-separating structures are refined to regulate the trade-off between precision and recall while improving the F1-score, Dice score and Hausdorff. BenTaieb et al. [13] designed a multi-loss convolution network that performs both classification and segmentation for colon adenocarcinomas diagnosis. In BenTaieb's method two distinct loss functions employed to optimize a single unified deep CNN. The proposed CNN architecture composed of two components organized symmetrically. For the first stage, the classification component determines the type of tumor using stacked layers of convolution and subsampling operations. Then, at the second stage, the segmentation component performs deconvolution and up-sampling operations to segment out the identified glands from the first stage. To integrate the feature maps from the first part (classification stage) to the segmentation stage, cross-network spatial activation maps was introduced. The performance of presented model also compared to the U-Net and AlexNet models. BenTaieb's proposed method by including class-specific spatial priors was able to achieve a more accurate segmentation network. sun2019comparativelung used a deep CNN and an FCN for the segmentation to design an ensemble deep learning model. The proposed model was trained on patches of the size of 224 × 224 for CNN and 256 × 256 pixels for FCN parts. The FCN part of the proposed architecture is based on DenseNet model. The predicted mask is produced using probability maps generated by each architecture with a dice similarity index of 77%. The obtained results in this study showed that CNN model was more precise to segment out the cancerous regions than FCN model. The performance of VGGNet and ResNet have been compared in a study conducted by Šarić et al. [14]. To generate the dataset, patches of the size of 256 × 256 were extracted from 25 whole slide images. Each patch was separated from another using a stride of 196 pixels to provide a sufficient degree of patch overlapping. Obtained results demonstrated that the VGG16 achieved better performance with higher true positive rates with 75.41% accuracy than ResNet50 with 72.05%. Sun et al. [15] employed a cascade model using convolutional neural networks for fast segmentation of whole slide images. The proposed cascade network has consisted of two U-Net architectures. The first U-Net architecture focuses on cancerous regions by filtering out non-informative areas such as blank or healthy regions and reduce computational complexity. Then, the second U-Net architecture is used to refine the predicted segmented regions from the first stage by reducing the over-segmented (false positive) areas. The obtained result of the proposed model achieved a 69.4% dice score and a precision of 94.7%. One major limitation of the proposed method is the computational costs, especially for real-time clinical applications. Figueira et al. [16] proposed a method for lung, breast and colon tumor detection based on adversarial domain adaptation networks (ABDA-Net) with ResNet50 as feature encoder. Non-overlapping patches of the size of 256 × 256 were extracted from the input whole slide images. The proposed method achieved 86.61% accuracy on the lung cancer dataset. The poor performance of the proposed method compared to the number of annotated images is the main limitation of this study. Koyun and Yildirim [17] designed an unsupervised nuclei segmentation model using a cycle-consistent generative adversarial network. The proposed model is a U-Net encoder-decoder architecture with nine residual connections between the encoder and decoder parts. The generator part of the proposed CycleGAN is able to generate different regular and irregular nuclei. To improve the predicted mask, different post-processing

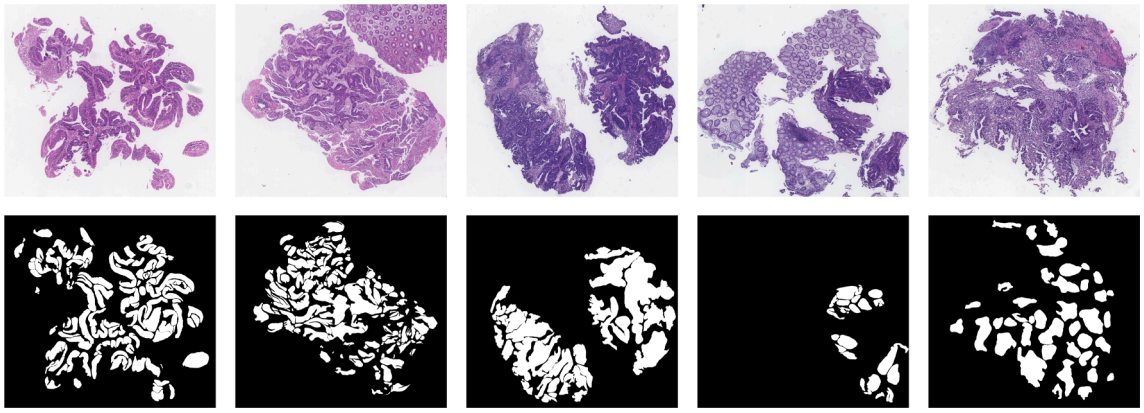

Fig. 1. Some examples of WSIs (first row) and their corresponding manual annotated masks provided by expert pathologists (second row).

methods were applied to the generated output. The authors also discussed that the proposed model might be sensitive to illumination and contrast variations of the input data. Khanagha and Kardehdeh [18] developed a deep learning model using a Fully Convolutional Neural Networks (FCNN) for whole slide images annotation. The model was trained on the patches of the size of 1024×1024. Three custom FCNN architectures (small-FCN-16, small-FCN-32 and small-FCN-512) were designed to measure the performance of deep FCNN on the tissue segmentation. In the proposed approach, the issue of the limited dataset size was addressed by avoiding the updating weights in the up-sampling layers. The proposed method achieved a 71.24% average dice similarity index. In a similar study conducted by Mahmood et al. [19], a dual generative adversarial network were adopted for a multi-organ nuclei segmentation task from histopathological images. The synthetically pathology images generated from the proposed model were combined with real input images to train CNNs and perform final nuclei segmentation. The backbone architecture was designed using an encoder-decoder architecture similar to U-Net with skip connections. The proposed model by using an adversarial pipeline with larger receptive field was able to capture more global information in comparison with standard CNNs. The proposed model was trained on patches of 256×256 of H&E-stained slides of histopathological images from different organs such as breast, ovary, esophagus, liver, kidney, prostate, bladder, colorectal and stomach. Zeng et al. [20] proposed a modified U-Net for nuclei segmentation in histopathology images. For stain normalization of H&E-stained images, Macenko et al. [21] method was employed by authors. The proposed Residual-Inception-Channel attention-UNet (RICUNet) architecture was able to identify different types of cell shapes and scales. The presented model inspired by studies of Chen et al. [22] and Yu et al. [23], which was applied on nuclei segmentation of different organs such as kidney, prostate, breast, liver, stomach colon and bladder, consisting of both benign and diseased tissue samples. In Chen et al. [22], a multi-task learning framework was employed to segment the nucleus and cell contour simultaneously. Yu et al. [23] employed CAB (Channel-Attention-Block) module in up-sampling part. The features from cell contour acts as auxiliary features to differentiate dense and overlapped cells and assist in reducing errors at the object level. In the proposed U-Net model, convolutional layers were replaced with Residual-Inception-blocks.

3. Materials and methods

Deep learning architectures are formed by a sequential convolutional layer, alternated by pooling layers. This architecture is able to learn nonlinear hierarchically discriminative features from input data. Different types of CNN architectures can be formulated by stacking convolutional layers [24]. A convolution window slides over the input image and performs a convolution operation (as shown in Eq. 1) to extract high-level discriminative features maps.

$$Y_i^n = f\left(\sum_i^m \omega_{ij}^{n-1} * x_i^{n-1} + b_i^n\right) \tag{1}$$

where m is the number of feature maps, ω_{ij}^{n-1} represents the convolution filter between the i and j feature maps, x_i^{n-1} is the i^{th} map in the $(n-1)^{th}$ layer, b_i^n denotes the bias, function f denotes a nonlinear activation function, and the asterisk $(*)$ represents the convolution operator.

When feature maps are generated from each convolutional layer, another layer termed a pooling layer, is used. The pooling layer aims to reduce the dimensionality of the feature maps, and hence the computational time of the model decreases. If the pooling operation is removed, the amount of CNN parameters increases exponentially in the subsequent layers. Another advantage of pooling operations is to reduce the sensitivity of the model to small transformations, variations, distortions and translations in input data. The two most common pooling strategies are max-pooling and average-pooling [25].

After extracting features using stacked convolutions and pooling layers, another layer which is called a fully connected (FC) layer is used to convert an extracted 2D summarized feature map into a 1D feature vector. An FC layer is similar to a conventional artificial neural network (ANN) or a multi-layer perceptron. The input–output operation in a neuron of the FC layer is defined in Eq. 2. In each neuron unit, the learned weights are multiplied by the corresponding data from the previous layer and the bias values are added. The calculated value is transmitted to the activation function before being passed to the next layer.

$$fc = f(b + \sum_i^m \omega_i x_i) \tag{2}$$

where f represents an activation function, w is the weight vector, x is the input feature vector of the i^{th} neuron, m is the number of feature maps, and b is the bias vector.

The final component of a deep CNN is the output layer to produce the predictive probabilities corresponding to each class. The softmax activation function is the common activation function for classification tasks. The softmax function is as follows:

$$f(x) = \frac{e^{x_i}}{\sum_{k=1}^{n} e^{n_k}} \tag{3}$$

where e^{x_i} is the i^{th} value in the output vector and n is the number of classes.

Fig. 2 presents a conventional deep CNN model, consisting of the input image, convolution layers followed by pooling layers, and finally, FC layers. In the convolution layer, also known as the feature extraction layer, a convolution kernel (yellow square), with a fixed size, convolves over the input image and generates a feature map. Extracted features are fed into the next layers as input. The pooling layer is used to reduce the size of the feature maps. Finally, the FC layer, at the end of the network, outputs the corresponding probabilities. The main idea of a CNN model is to obtain high-level features such as edge, shape, and texture directly from the input image.

3.1. Segmentation architectures

In this section, a short description of each state-of-the-art CNN architecture utilized in this study is provided:

3.1.1. U-Net

U-Net, proposed by Ronneberger et al. [26] in 2015, achieved effective segmentation results when compared with the ConvNet approach and won the ISBI cell tracking challenge. The contracting path or down-sampling layers of the U-Net architecture learns the feature maps using alternating convolutional filters and max pooling layers. The expanding path, or up-sampling layers, act as inputs for a deconvolution process and provide precise segmentation.

3.1.2. LinkNet

LinkNet [27] is an efficient encoder-decoder architecture, which takes advantage of skip connections and residual blocks, to address the problem of spatial information by directly connecting spatial information from the encoder to the corresponding decoder part. The encoder part of the original LinkNet uses ResNet18 to extract features from input images and the corresponding decoder produces the predicted mask.

3.1.3. Feature Pyramid Network (FPN)

The main idea of the FPN architecture is to combine low-level semantically strong features with high-level semantically weak features from each layer, independently, to produce the final pixel classification. Feature pyramids in the FPN architecture are the basic

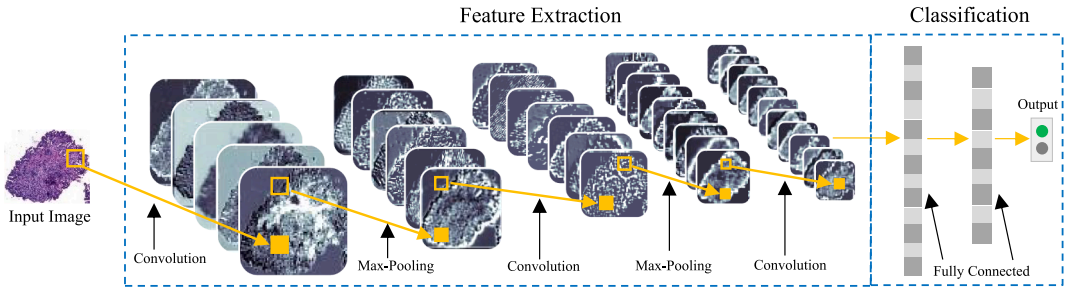

Fig. 2. A typical convolutional neural network architecture with three main layers of convolution layers, pooling layers, and fully connected layers.

component in recognition systems for detecting multi-scale objects [28].

3.2. Deep CNN feature extractors

3.2.1. Deep residual learning network

The Deep Residual Learning Network (ResNet) was proposed by He et al. [29]. This architecture won the ILSVRC classification task with good results on ImageNet and MS-COCO object detection competitions. This architecture introduced the concept of residual blocks. The main goal of residual blocks is to add a connection (instead of concatenation) from the input of the first block to the output of the next block in order to train a deeper network with better recognition ability. This architecture can solve the issues of vanishing gradients and parameter explosion by shortcut connection using the residual blocks.

3.2.2. VGG-Net

The Visual Geometry Group (VGG-Net) [30] was proposed by Karen Simonyan and Andrew Zisserman. This architecture obtained top performance on ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2014. VGG-Net provides better features extraction from input images by using $3\times$ 3 filter size. VGG16 and VGG19 are two versions of VGG-Net architecture with different depths and layers.

3.2.3. Densely connected convolutional networks

The Densely Connected Convolutional Networks (DenseNet) architecture proposed by Huang et al. [31] and is an extension for ResNet architecture. Summation operations are used to connect layers to each other in this architecture. The summation operation helps further improve generalization and addresses the problem of vanishing gradient better than the ResNet architecture. In this approach, the features extracted from each layer are reused as input for the next layers. Reusing feature maps helps to further improve the final results.

3.2.4. MobileNet

The MobileNet architecture [32] was designed by the Google research team for object recognition on mobile devices. This architecture introduced the depth-wise separable convolution and $1\times$ 1point-wise convolution layers. MobileNet architecture achieved an accuracy at the same level as VGG16 with 32 times fewer parameters than standard convolutions and won an ImageNet dataset competition while being 27 times less computationally intensive. Depth-wise convolution applies a single spatial filter to each input channel. Point-wise convolution is a standard convolution with the kernel size of (1×1) that computes a linear combination of different input channels. This operation can significantly reduce the dimensionality of the feature maps.

3.2.5. InceptionV3

The InceptionV3 architecture, proposed by Szegedy et al. [33] in 2014, won the ImageNet competition. InceptionV3 architecture introduced the concept of inception module. This architecture consists of 159 layers in total, and is the third generation of Inception model. In the Inception model, instead of using one type of kernel size (i.e. 3×3, or 5×5), different convolution size, i.e. 1×1, 3×3, and 5×5 filter sizes are used. Employing different convolution sizes allows for the extraction of multi-level features from the input image in every convolution operation. In this architecture, point-wise 1×1 convolution is also used to reduce the number of parameters. The point-wise convolution helps reduce the computational cost.

3.2.6. InceptionResNetV2

The InceptionResNetV2, network was proposed by Szegedy et al. [34]. Both Inception models and residual connections are introduced in InceptionResNetV2 [34]. The combination of these models helps develop deeper architecture, and achieves a better performance with relatively low computational cost.

3.2.7. ResNeXt

The ResNeXt CNN model, developed by Xie et al. [35]won the ImageNet ILSVRC 2016 competition. A split-transform-merge strategy was used to further improve the deep residual network incorporated in a ResNeXt architecture. The adapted residual units in ResNeXt, similar to those in the ResNet architecture, help develop a deeper model and tackle the issues of over-fitting and vanished gradients. In the experiments conducted by Szegedy et al., a new hyper-parameter, cardinality, was introduced to the network to define number of paths in the ResNeXt block.

3.2.8. SE-ResNet and SE-ResNeXt architectures

The squeeze-and-excitation networks (SE) were first proposed by Hu et al. [36] in 2017 for image classification. The main idea of the SE units as a self-gating mechanism, are to learn features from each image channels adaptively. A squeeze operation is used to squeeze the generated feature maps into channel-wise descriptors to represent global features of each channel. The excitation operation is 1×1 point-wise convolutional layers followed by a ReLU layer to adaptively strengthen learned features. The integration of SE blocks into ResNet and ResNeXt architectures allows to design more complex architecture with improved accuracy.

3.3. Transfer learning

Transfer learning is a common strategy in deep learning tasks where a large dataset from a source task is used for training of a target task to overcome the problem of small datasets, accelerate the learning process and improve accuracy. Previous studies showed that transfer learning also has the potential to prevent over-fittings [37–39]. The transfer learning approach enables us to adopt a pre-trained network that has already learned a rich set of low-level features from layers that are closer to the input image. Though the dataset is not the same, the low-level features produced by source CNN are mostly in general shapes, e.g. edges, contours and curves, which are similar to the low-level features of the target dataset. In contrast, high-level features at the final layers concentrate on complex class-level characteristics which are necessary to differentiate between classes. With the use of transfer learning,

training of large CNNs can be a more practical strategy with more promising results and significantly cost-effective by avoiding training a CNN from scratch. In training a network from scratch, the weights are initialized randomly, while in transfer learning strategy, the learned weights from another domain are transferred to the medical domain (see Fig. 3).

3.4. Deep CNN architecture for CRC segmentation

The main objective of this work is to explore the impacts of varying modules on the performance of deep CNN models for the purpose of determining an optimal set of CNN modules and pre-trained architectures. Hence, the performance of multiple well-established deep CNN models with various layers or specifications is investigated. In this research, 51 combinations of pre-trained networks and segmentation architectures are designed in the down-sampling part of segmentation architecture backbones, and then a comparative study is conducted to report the results. To conduct this research, the impact of different CNN modules, e.g. squeeze-excitation (SE) incorporated into SE-ResNet (18, 34 and 50) and SE-ResNext50 models, residual blocks in ResNet (18, 34 and 50) models, dense modules in DenseNet (121, 169 and 201) models, inception modules in InceptionV3, InceptionResNetV2 model and standard convolutions in VGGNet (16 and 19) were assessed for automated CRC cancer tissue segmentation. Residual units were also incorporated in InceptionResNetV2, ResNeXt, SE-ResNeXt and SE-ResNet models.

The proposed deep learning framework is based on deep CNN and includes two parts: the first stage is the encoder part that incorporates pre-trained deep CNN architecture to extract high-level contextual feature representations automatically from the input image. The next part of the architecture is the decoder part that up-samples the encoded image feature representations into an output predicted mask. Fig. 4 shows the proposed fully convolutional-based feature extractor with InceptionV3 and LinkNet architecture for automatic CRC tissue segmentation and screening. The left side of this figure is the image preprocessing step, e.g. extracting patches from an input whole slide image (WSI). The right side of this figure demonstrates the proposed architecture. The encoder part of this architecture extracts features using inception modules and the decoder part of the proposed architecture produces the final mask. The designed generic framework allows extracting discriminative features based on the end-to-end learning process of the texture and shape of normal and tumor regions and finally delineate ROIs from CRC histology slide images. Additionally, the proposed model provides a new level of feature extractors by incorporating prior knowledge already trained on the ImageNet dataset using pre-trained deep modules into the segmentation framework.

As demonstrated in the Fig. 4, the proposed architecture is composed of two separate parts. The left part carries out feature extraction from the input layer with a resolution of 512 × 512 pixels. In contrast, the right part propagates the obtained extracted features to the left part to produce the final predicted mask. The 512 × 512 pixels of input images are large enough to cover the ROIs of the provided dataset with reasonable memory consumption. This approach allows to design a much deeper architecture, i.e. U-Net with 2 M parameters. In contrast, the combination of DenseNet201 with UNet with the total size of 26 M parameters, LinkNet with the number of 1 M parameters and LinkNet combined with DenseNet201 has 22 M parameters to successfully accomplish the segmentation task without the problem of vanishing gradient problem.

3.5. Dataset description

The image dataset used in this work for colorectal cancer segmentation is DigestPath [40] available at [41]. The dataset consisted of a total of 250 positive colonoscopy tissue slices containing both normal and tumor regions. The size of images ranges from 3538 × 5736 pixels to 16054 × 13821 pixels extracted from the high-resolution scans of anonymous patients to evaluate the performance of the segmentation model. The cancerous regions of whole slide images are annotated by the HistoPathology Diagnostic Center with their collaborating hospitals. The whole slide images were stained by hematoxylin and eosin and scanned at × 20. The provided H&E stained histology WSIs of colorectal tissue were highly heterogeneous in terms of shape, texture and appearance as the data were collected from 4 different medical centers in developing countries. Malignant lesions contained high-grade intraepithelial neoplasia and adenocarcinoma diseases, including papillary adenocarcinoma, mucinous adenocarcinoma, poorly cohesive carcinoma and signet ring cell carcinoma.

3.5.1. Patch extraction

To conduct a successful diagnosis, the magnification level of a WSI

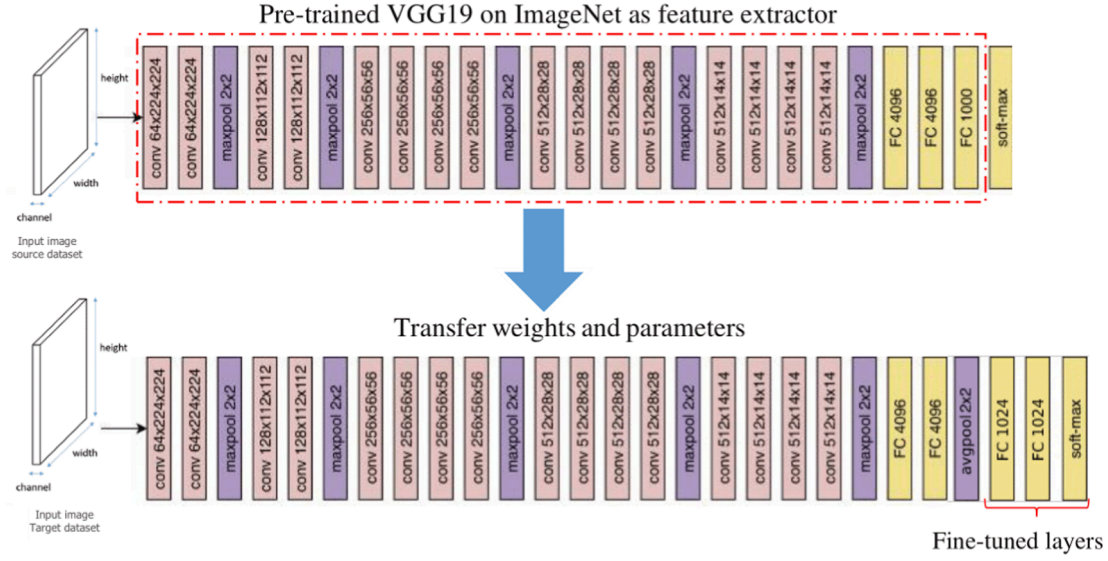

Fig. 3. Illustration of the transfer learning process. The top architecture in the figure is originally trained on the source dataset, while the bottom architecture is used to train the target dataset. The weights and parameters learned of the pre-trained CNN is transferred to the bottom architecture. For fine-tuning, the last fully connected (FC) layer, which can be seen as a classification layer of the pre-trained CNN networks is discarded. After that, two new FC layers, each of which with 1024 hidden units is included to the bottom architecture.

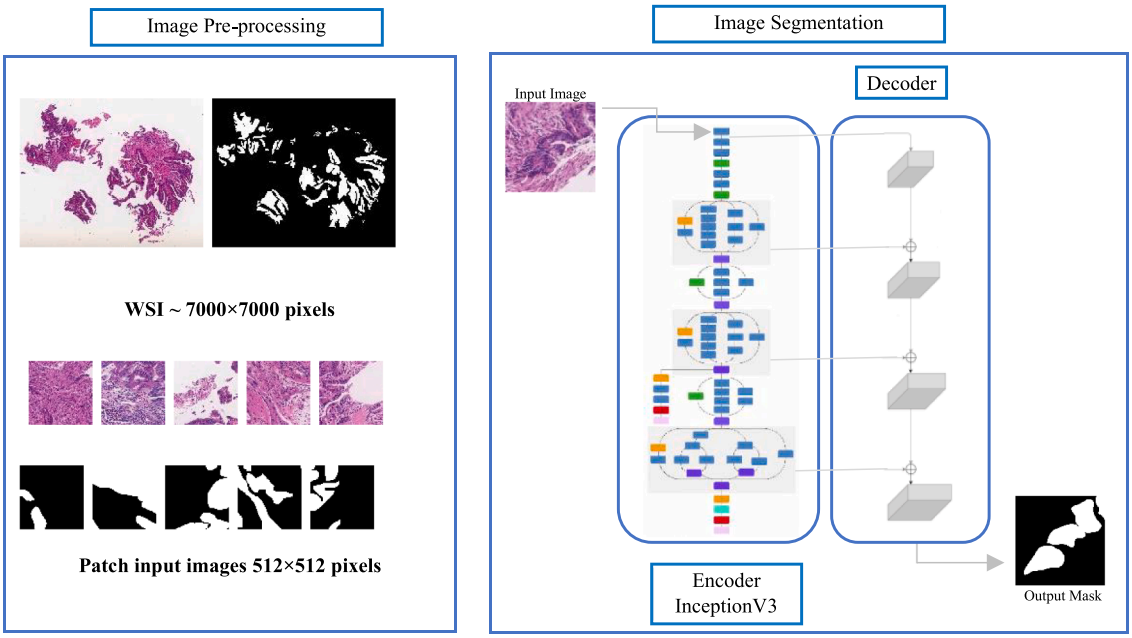

Fig. 4. The illustration of the proposed convolutional architecture with dense blocks.

should be adjusted to provide more detailed information about cancerous and/or healthy regions [42]. Due to the substantial size of WSIs and limitations in computational power, it is difficult to process the entire WSI at once [43]. A patch extraction method was therefore employed on the tissue slides containing both normal and tumor regions to generate the dataset. A non-overlapping window was used to crop patches of 512 × 512 pixels from each abnormal WSI (examples in Fig. 5).Another issue with WSI processing is that abnormal regions only occupy small proportion of some WSIs compared with the healthy regions. Also, patches with less than 25% cancerous tissue sections were discarded from the generated dataset. To this end, 1596 patches for the training set and 150 patches for the test set were selected from 250 positive WSIs in total. An advantage of the proposed patch-based model is the computational efficiency with which it allows to train very deep CNN architectures.

3.6. *Experimental settings*

Training is performed using Stochastic Gradient Descent (SGD) with momentum, proposed by [44]) as an optimizer with an initial learning rate 0.1, momentum 0.9 and decay rate 0.1. The learning rate is the most important hyper-parameter in an optimizer when training a deep model. To set the optimal value of learning rate, this study utilizes a learning rate scheduler to facilitate the optimal convergence and avoid overfitting during training. For this research, a grid search is employed to find the optimal values of optimizer and learning rate to train the

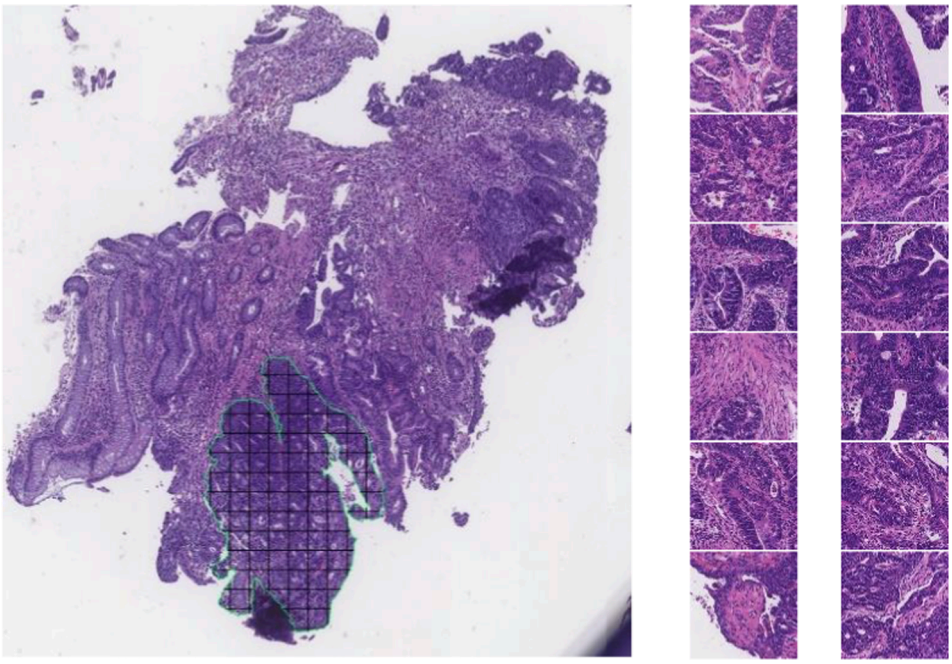

Fig. 5. The process of patch generation from a whole slide image of the provided dataset. The scale of the image is 15μm.

architectures. The learning rate drops every two epochs during the training procedure. The SGD accepts the parameter learning rate η (default value is set to 0.01), momentum as a parameter of μ, decay parameter to decay the learning rate over the weights updates and Nesterov parameter with the following formula:

$$\eta^{(t+1)} = \frac{\eta^{(t)}}{1 + decay} \tag{4}$$

The mini-batch size was set to 4 images due to the GPU memory limitations, and all models were trained for 50 epochs. The dataset has been divided into training, validation and test sets with portions of 70%, 20% and 10%, respectively. The weights for feature extractors were initialized by using pre-trained ImageNet initialization. The ImageNet weight initialization approach assists in faster convergence and speeds up the training process. All experiments were run on a PC with the following configuration: Intel(R) Core (TM) i7-8700 K 3.7 GHz processors with 32 GB RAM. The training and testing process of the proposed architecture for this experiment is implemented in Python using Keras package with Tensorflow as the deep learning framework backend and run on Nvidia GeForce GTX 1080 Ti GPU with 11 GB RAM.

3.7. Evaluation metrics

To measure the performance of the proposed method for the segmentation task, common segmentation evaluation metrics such as dice similarity coefficient, precision, recall, f1-score were adopted to quantitatively measure similarity and difference between the predicted mask from the proposed segmentation model and the ground-truth mask at the pixel level.

The dice similarity coefficient measures the spatial overlap between the ground truth and predicted mask produced by the proposed architecture. These metrics are computed by the following:

$$Dice(A, B) = \frac{2 \times |A \cap B|}{|A| \cup |B|} \times 100 \tag{5}$$

where A represents the output binary mask, produced from the segmentation method, and B represents the ground-truth mask, $\cup$ represents union set between A and B, and $\cap$ represents the intersection set between A and B.

Accuracy metric is used to measure the overall accuracy of the segmentation models. Given the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{6}$$

Precision and Recall metrics are analyzed to measure the amount of over-segmentation and under-segmentation, respectively. Precision is sensitive to over-segmentation as it is associated with a small precision score. The recall is sensitive to under-segmentation as it is associated with results in low recall scores.

$$Precision = \frac{TP}{TP + FP} \times 100 \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{8}$$

F1-score is also computed as a harmonic mean of precision and recall between predicted and ground truth boundaries to evaluate the performance of the proposed approach.

$$F1 - Score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \times 100 \tag{9}$$

Mean squared error (MSE) is the average of the squared error and is widely used as the loss function to evaluate the quality of the models. In this study, MSE is also used to measure the predicted mask and ground truth masks as a standard of reference. It is an average squared difference between the predicted and actual target values.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{10}$$

where $\hat{y}_i$ represents the output binary mask produced from the segmentation method, y_i represents the ground-truth mask.

4. Comparative experimental results

The main objective of this experiment is to test the generalization ability of the proposed segmentation method via a convolutional feature extractor and transfer learning for early-stage colon tumor detection from small tissue slices. Different pre-trained CNN models were selected as feature extractors of the encoder part of different backbones, e.g. U-Net, LinkNet and FPN, for comparative analysis. These architectures are selected for feature extraction based on their (i) satisfactory performance in medical image processing, (ii) adaptation towards real-time (or near real-time) image diagnosis support system and, (iii) feasibility of transfer learning for different computer vision tasks such as detection, segmentation and classification [45].

Tables 1–3 report the patch-based tumor segmentation results of different approaches. Each entry in these tables, is in the format ($\mu \pm \sigma$) where μ is the average segmentation metric and σ is standard deviation. Overall, the results are in favor of the InceptionResNetV2 and DenseNet architectures in terms of dice similarity coefficient, accuracy and F1-score. InceptionResNetV2 combines the merits of Inception modules with residual connections to increases convergence speed and improve performance.

Analyzing Table 3, the topmost result of all combinations was obtained by DenseNet121 feature extractor on LinkNet segmentation architecture with a maximum of 82.74%±1.77 dice similarity coefficient and accuracy of 87.07±%1.56. It is also inferred from Table 1 that the second-best result from all combinations obtained by InceptionResNetV2 feature extractor and FPN backbone architecture with an overall dice similarity coefficient of 82.53±%1.56 and accuracy of 87.10%± 1.53. Based on the observations in Table 1, ResNet50, VGG16 and, VGG19 have the lowest segmentation accuracies, dice similarity index and F1-scores to all segmentation backbones of FPN, U-Net and LinkNet in this study. Conversely, MobileNet, and MobileNetV2 models consistently perform better with FPN than U-Net and LinkNet networks. Also, ResNext50 and squeeze-and-excitation networks (SE-ResNet18, 34, 50, ResNext50) models, when applied in FPN architecture, are more stable than counterparts.

Comparing the first and second winners among all combinations, the performance of dense modules in DenseNet architecture is slightly better (1%) than the rest of the feature extractors on the FPN backbone. Analyzing Table 2, among deep feature extractors, InceptionResNetV2 has the highest dice similarity score of 82.14%±1.79, and accuracy of 87.03%±1.51 as well as the highest F1-score of 82.27%±1.81, followed by DenseNet201 with overall dice similarity score of 82.07%±1.85, accuracy of 86.99%±1.52, as well as F1-scores of 82.12%±1.84. Though the lowest scores are obtained by Se-ResNeXt50 feature extractor, VGG19 has the worse MSE rate (0.1552).

In terms of comparing our results and current state-of-the-art methods based on the similarity of architecture and histopathology dataset, it can be clearly seen in Table 5 that our approach using transfer learning in the encoder part of segmentation architecture shows a significant improvement in Dice, accuracy and F1-score against other approaches.

The segmentation of tumor epithelium in histopathology slide images is a critical step for early diagnosis in colorectal cancer. In this research, we presented a comparative analysis of a wide variety of well-established deep CNNs as feature extractors as part of the FPN, U-Net and LinkNet architecture. The proposed framework based on encoder-

Table 1
Comparative analysis ($\mu \pm \sigma$) of different feature extractors and FPN segmentation architecture on the DigestPath challenge dataset. The bold value indicates the best result; underlined value represents the second-best result of the respective category.

FPN	Dice	Accuracy	Precision	Recall	F1-score	MSE
DenseNet121	82.48 ± 1.75	86.96 ± 1.48	82.93 ± 1.73	83.64 ± 1.76	82.62 ± 1.78	0.1123
DenseNet169	82.50 ± 1.85	87.08 ± 1.49	83.60 ± 1.80	82.97 ± 1.77	82.63 ± 1.74	0.1120
DenseNet201	82.34 ± 1.98	86.88 ± 1.58	84.27 ± 1.75	82.29 ± 1.76	82.46 ± 1.79	0.1148
InceptionV3	81.48 ± 1.75	86.70 ± 1.57	84.05 ± 1.72	80.90 ± 1.82	81.64 ± 1.81	0.1152
InceptionResNetV2	**82.53 ± 1.56**	**87.10 ± 1.53**	82.74 ± 1.66	84.10 ± 1.70	**82.73 ± 1.72**	0.1115
MobileNet	82.40 ± 1.58	86.85 ± 1.54	82.08 ± 1.70	84.24 ± 1.76	82.51 ± 1.75	0.1151
MobileNetV2	82.12 ± 1.65	86.59 ± 1.58	81.21 ± 1.79	84.82 ± 1.80	82.34 ± 1..78	0.1154
ResNet18	81.56 ± 1.74	86.21 ± 1.56	80.83 ± 1.75	84.10 ± 1.83	81.68 ± 1.84	0.1211
ResNet34	82.09 ± 1.84	86.57 ± 1.54	81.07 ± 1.82	84.84 ± 1.86	82.24 ± 1.84	0.1175
ResNet50	79.24 ± 1.72	84.90 ± 1.58	78.01 ± 1.70	83.76 ± 1.82	80.05 ± 1.87	0.1309
ResNeXt50	81.39 ± 1.94	86.34 ± 1.30	82.96 ± 1.73	81.42 ± 1.78	81.52 ± 1.81	0.1182
SE-ResNet18	81.63 ± 1.83	86.24 ± 1.54	81.34 ± 1.77	83.58 ± 1.79	81.78 ± 1.79	0.1199
SE-ResNet34	81.48 ± 1.79	86.30 ± 1.60	82.22 ± 1.73	82.39 ± 1.81	81.62 ± 1.82	0.1194
SE-ResNet50	81.39 ± 1.84	86.86 ± 1.61	85.94 ± 1.69	78.87 ± 1.88	81.55 ± 1.78	0.1128
SE-ResNeXt50	81.05 ± 1.89	86.60 ± 1.60	85.85 ± 1.72	77.97 ± 1.87	81.19 ± 1.79	0.1151
VGG16	80.55 ± 2.05	83.78 ± 1.71	73.86 ± 1.91	90.54 ± 1.60	80.62 ± 1.86	0.1451
VGG19	80.68 ± 2.11	83.85 ± 1.69	73.38 ± 1.94	91.59 ± 1.53	80.80 ± 1.87	0.1448

Table 2
Comparative analysis ($\mu \pm \sigma$) of different feature extractors and U-Net segmentation architecture on the DigestPath challenge dataset. The bold value indicates the best result; underlined value represents the second-best result of the respective category.

U-Net	Dice	Accuracy	Precision	Recall	F1-score	MSE
DenseNet121	81.90 ± 1.84	86.92 ± 1.54	85.12 ± 1.77	80.31 ± 1.86	81.95 ± 1.83	0.1189
DenseNet169	81.85 ± 1.80	86.77 ± 1.56	84.82 ± 1.74	80.50 ± 1.88	81.89 ± 1.81	0.1188
DenseNet201	82.07 ± 1.85	86.99 ± 1.52	83.55 ± 1.79	82.23 ± 1.82	82.12 ± 1.84	0.1179
InceptionV3	81.84 ± 1.86	86.76 ± 1.54	83.12 ± 1.83	81.97 ± 1.84	81.91 ± 1.82	0.1189
InceptionResNetV2	**82.14 ± 1.79**	**87.03 ± 1.51**	83.53 ± 1.75	82.42 ± 1.87	**82.27 ± 1.81**	0.1175
MobileNet	80.43 ± 1.88	85.38 ± 1.57	77.49 ± 1.81	85.58 ± 1.83	80.63 ± 1.82	0.1279
MobileNetV2	81.20 ± 1.83	86.36 ± 1.52	80.16 ± 1.79	83.60 ± 1.85	81.31 ± 1.84	0.1212
ResNet18	82.16 ± 1.82	86.76 ± 1.56	80.57 ± 1.80	85.30 ± 1.79	82.23 ± 1.86	0.1200
ResNet34	81.86 ± 1.87	86.64 ± 1.58	80.52 ± 1.79	84.65 ± 1.84	81.93 ± 1.84	0.1217
ResNet50	81.76 ± 1.84	86.62 ± 1.53	83.28 ± 180	81.68 ± 1.86	81.81 ± 1.87	0.1223
ResNeXt50	81.50 ± 1.92	86.26 ± 1.56	83.09 ± 1.78	81.23 ± 1.81	81.56 ± 1.81	0.1247
SE-ResNet18	81.66 ± 1.87	86.53 ± 1.55	80.82 ± 1.80	83.97 ± 1.89	81.73 ± 1.83	0.1219
SE-ResNet34	82.07 ± 1.90	87.04 ± 1.52	82.24 ± 1.81	83.13 ± 1.82	82.15 ± 1.85	0.1166
SE-ResNet50	80.43 ± 1.94	86.32 ± 1.59	87.43 ± 1.77	75.78 ± 2.01	80.56 ± 1.89	0.1214
SE-ResNeXt50	79.18 ± 2.11	85.65 ± 1.64	86.47 ± 1.75	74.49 ± 1.99	79.34 ± 1.92	0.1267
VGG16	81.48 ± 2.02	85.00 ± 1.70	75.59 ± 1.92	90.05 ± 1.76	81.56 ± 1.87	0.1355
VGG19	80.05 ± 2.16	83.02 ± 1.72	72.13 ± 1.94	92.05 ± 1.71	80.14 ± 1.86	0.1552

Table 3
Comparative analysis ($\mu \pm \sigma$) of different feature extractors and LinkNet segmentation architecture on the DigestPath challenge dataset. The bold value indicates the best result; underlined value represents the second-best result of the respective category.

LinkNet	Dice	Accuracy	Precision	Recall	F1-score	MSE
DenseNet121	**82.74 ± 1.77**	**87.07 ± 1.56**	82.78 ± 1.77	84.03 ± 1.83	**82.79 ± 1.79**	0.1176
DenseNet169	81.95 ± 1.82	86.94 ± 1.59	85.13 ± 1.74	80.58 ± 1.89	81.99 ± 1.86	0.1195
DenseNet201	81.44 ± 1.80	86.69 ± 1.54	85.14 ± 1.73	79.72 ± 1.	81.51 ± 1.82	0.1209
InceptionV3	81.07 ± 1.74	86.45 ± 1.57	83.42 ± 1.70	80.23 ± 1.81	81.16 ± 1.80	0.1212
InceptionResNetV2	81.02 ± 1.78	86.25 ± 1.53	83.92 ± 1.76	80.01 ± 1.83	81.09 ± 1.80	0.1242
MobileNet	80.88 ± 1.78	85.68 ± 1.58	77.72 ± 1.78	86.30 ± 1.79	81.07 ± 1.81	0.1266
MobileNetV2	79.61 ± 1.85	85.09 ± 1.58	78.44 ± 1.75	83.05 ± 1.82	79.84 ± 1.83	0.1312
ResNet18	81.21 ± 1.81	86.33 ± 1.49	81.90 ± 1.74	82.28 ± 1.85	81.28 ± 1.79	0.1252
ResNet34	77.52 ± 2.20	86.09 ± 1.52	79.55 ± 1.82	84.24 ± 1.76	81.21 ± 1.81	0.6358
ResNet50	80.93 ± 1.79	86.13 ± 1.57	82.23 ± 1.80	81.28 ± 1.1.88	80.98 ± 1.84	0.1272
ResNeXt50	81.25 ± 1.75	86.03 ± 1.54	82.30 ± 1.79	81.97 ± 1.90	81.33 ± 1.79	0.1259
SE-ResNet18	80.73 ± 1.82	85.98 ± 1.61	80.20 ± 1.78	83.08 ± 1.86	80.82 ± 1.82	0.1275
SE-ResNet34	81.37 ± 1.81	86.53 ± 1.60	80.84 ± 1.76	83.35 ± 1.84	81.46 ± 1.80	0.1223
SE-ResNet50	80.78 ± 1.85	86.50 ± 1.55	84.95 ± 1.79	78.39 ± 1.92	80.88 ± 1.81	0.1210
SE-ResNeXt50	81.68 ± 1.89	86.93 ± 1.59	85.22 ± 1.78	79.58 ± 1.98	81.80 ± 1.80	0.1148
VGG16	79.12 ± 2.06	81.98 ± 2.06	71.06 ± 2.54	91.62 ± 1.71	79.20 ± 1.84	0.1648
VGG19	73.48 ± 2.37	74.56 ± 2.25	61.01 ± 3.11	95.47 ± 0.96	73.56 ± 2.12	0.2399

decoder architectures of FPN, U-Net and LinkNet integrated with pre-trained feature extractors has the potential to overcome the current challenges of conventional segmentation methods, reducing subjectivity and the daily workload of pathologists with decent speed and accuracy.

Several conclusions can be summarized based on the obtained results: i) it can be observed from the experimental results that the proposed tumor segmentation method can exploit deep convolution features and learn discriminative activation maps from the representative patches, which is

less computationally expensive. Another advantage of the patch-based approach is that all the regions of interest present in an image can be cropped and be used as the input while discarding non-informative regions such as the white background. In this way, a patch-based method can significantly decrease the processing time of both the training and validation set. ii) It is important to be noted that the hyper-parameters such as the network depths and network widths can substantially impact the performance and generalizability of the networks. Table 4 presents the number of parameters and layers of each architecture examined in this study. As listed in Table 4, developing very deep feature extractors with millions of parameters (e.g. InceptionResNetV2 integrated into U-Net with 62 million parameters) can improve the deep CNN results. However, deep models with skip connections (e.g. ResNet50), which discard some layers, can decrease the performance. The trade-off curve between the number of parameters and model performance can be adjusted with strategies such as skip connection is dense and residual modules. For instance, the combination of inception module with shortcut path of residual units in InceptionResNetV2 architecture achieved the second-best result. Training of very deep architectures remains an open problem due to its adverse effect on the ability to generalize unseen test data. First, as the number of layers increases, model performance increases too fast, but after a few iterations, the performance decreases due to the undesirable gradient degradation. A major contributing factor in resolving the vanishing gradient issue and taking full advantage of performance gain of training deep models is to introduce shortcut paths that allow flowing the gradient throughout the very deep networks. Such techniques are crucial to enable gradient-based training of very deep models in an end-to-end manner, which means effective utilization of deep features, therefore better generalization capability. In contrast, using just standard convolutions (e.g. VGG16 or VGG19) oversimplifies the architecture that can adversely affect the performance of the automatic CRC tissue segmentation as results shown in Table 3 and Fig. 6.

5. Model explainability

Although the proposed method achieved overall good performance on most testing images, the performance can be decreased as the histological structures of the cancerous area in some malignant cases are more challenging and severely irregular. A careful study of the obtained results showed that irregular structures in malignant regions might reduce the ability of feature extractor modules in differentiating malignant and healthy regions. For example, a deep learning model may fail in cases with irregular and high-dense structures, resulting from high proliferation. This is a significant and long-standing limitation for developing robust deep learning models to segment more challenging cases in histology images. It is worthwhile to note that the careful analysis of the errors in the results showed cases with considerable variation in tissue with irregular shapes, noisy background and vague edge resolution, which are resulted from acquisition images from different scanners, different staining protocols, specimens acquired from patients with a different stage of disease or at different time slots. To address the issue of over- or under-segmentation of tissue segmentation tasks and further improve the overall performance, pre-processing methods can be a potential solution. Furthermore, the performance of the deep learning architectures should be evaluated in real clinical study trials.

Table 4
Total number of parameters and layers of each deep CNN architecture. #P represents the number of parameters in million and #L represents the number of layers and sub-layers.

Model	FPN		U-Net		LinkNet	
	# P	# L	# P	# L	# P	# L
DenseNet121	9.9	474	12.1	468	8.3	483
DenseNet169	15.7	642	19.5	636	15.6	651
DenseNet201	21.2	754	26.3	748	22.5	763
InceptionV3	25	358	29.9	352	26.2	367
InceptionResNetV2	57.5	827	62	821	57.8	836
MobileNet	6.1	134	8.3	128	4.5	143
MobileNetV2	5.2	202	8	196	4.1	211
ResNet18	13.8	133	14.3	127	11.5	142
ResNet34	23.9	205	24.4	199	21.6	214
ResNet50	26.9	237	32.5	231	28.7	246
ResNeXt50	26.4	1263	32	1257	28.2	1272
SE-ResNet18	13.9	189	14.4	183	11.6	198
SE-ResNet34	24	317	24.6	311	21.7	326
SE-ResNet50	29.4	350	35.1	344	31.3	359
SE-ResNeXt50	28.9	1374	34.5	1368	30.8	1383
VGG16	17.5	66	23.7	66	20.3	81
VGG19	22.8	69	29	69	25.6	84

Table 5
Performance comparison of the proposed method and with current state-of-the-art methods.

Studies	Dice %	Accuracy %	F1-Score %
Figueira et al. [16]	-	86.61	-
Šarić et al. [14]	-	75.41	-
Sun et al. [15]	69.4	-	-
Khanagha and Kardehdeh [18]	71.24	-	-
Koyun and Yildirim [17]	-	-	71.6
Zeng et al. [20]	80.08	-	82.78
Mahmood et al. [19]	72.1	-	-
Proposed (LinkNet + Denset121)	82.74	87.07	82.79

6. Conclusion

This research presents a detailed comparative analysis of a wide variety of state-of-the-art deep CNNs in the encoder part of three segmentation backbones. The method is fast in analyzing batches of images as it is based on transfer learning strategy and patch-wise extraction method from colorectal cancer histology WSIs with heterogeneous shape and texture. Transfer learning strategy helps accelerate the learning process and further improve the performance of the proposed network. The extensive comparative evaluation demonstrated the state-of-the-art performance achieved by Densenet121 integrated by the LinkNet model with dice similarity score of 82.74 ± 1%.77, accuracy of 87.07 ± 1%.56 as well as the highest F1-score of 82.79%±1.79. The second-best result is obtained by InceptionResNetV2 pre-trained model with dice similarity score of 82.53%±1.56, accuracy of 87.10 ± 1%.53 as well as the highest F1-score of 82.73 ± 1%.72. Overall, comparing all of the feature extractors and segmentation backbone models, FPN and U-Net models tend to produce more stable results and are (almost) equivalent. Compared to conventional methods, where extensive pre-processing methods are used to increase the performance, the proposed approach avoids task-specific pre-processing method or data augmentation in order to improve the generalization ability. Furthermore, the proposed framework could be adopted to analyze complex problems with laboratory-dependent staining protocols, heterogeneous textures, and scanner-dependent intensity inhomogeneity. For our future research direction, our approach could be extended to different segmentation tasks on histology slide images. Another interesting direction could be applications of deep CNNs on temporal histology data with long short-term memory (LSTM) or recurrent networks. Moreover, the performance of the proposed framework could further be improved by reducing the noise using specific stain normalization techniques.

Author contributions statement

S.H.K. implemented the deep neural networks and classification methods development as well as prepared the related figures and drafted the manuscript text. P.H.K and M.J.W reviewed and edited the manuscript. R.D and K.A.S supported project supervision.

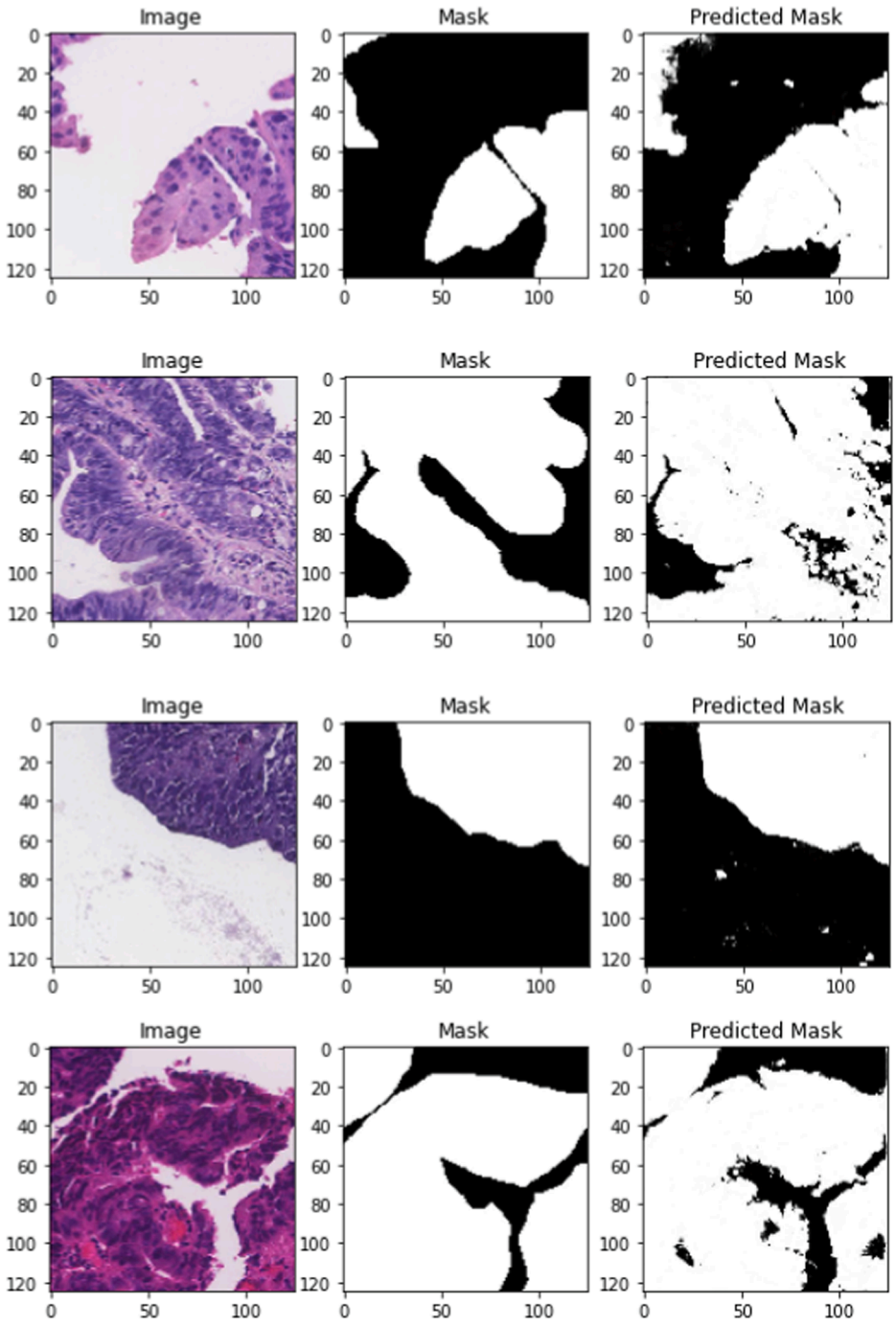

Fig. 6. Sample segmentation results: Different samples of the best proposed model on LinkNet and DenseNet121.

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.

Acknowledgments

Authors thank grand-challenge for providing the Digestpath dataset used in this research. We are also thankful to the reviewers in advance for their comments and suggestions.

References

[1] Yves-Rémi Van Eycke, Cédric Balsat, Laurine Verset, Olivier Debeir, Isabelle Salmon, Christine Decaestecker, Segmentation of glandular epithelium in colorectal tumours to automatically compartmentalise ihc biomarker quantification: A deep learning approach, Medical Image Anal. 49 (2018) 35–45.
[2] Key statistics for colorectal cancer. https://www.cancer.org/cancer/colon-rectal-cancer/about/key-statistics.html.
[3] Colorectal cancer. https://www.cancer.org/cancer/colon-rectal-cancer/about/what-is-colorectal-cancer.html.
[4] Rebecca L. Siegel, Kimberly D. Miller, Ahmedin Jemal, Cancer statistics, CA: A Cancer J. Clinicians 69 (1) (2019) 7–34.
[5] Talha Qaiser, Yee-Wah Tsang, Daiki Taniyama, Naoya Sakamoto, Kazuaki Nakane, David Epstein, Nasir Rajpoot, Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features, Medical Image Anal. 55 (2019) 1–14.

[6] Simon Graham, Hao Chen, Jevgenij Gamper, Qi Dou, Pheng-Ann Heng, David Snead, Yee Wah Tsang, Nasir Rajpoot, Mild-net: minimal information loss dilated network for gland instance segmentation in colon histology images, Medical Image Anal. 52 (2019) 199–211.
[7] Mohammed M. Abdelsamea, Alain Pitiot, Ruta Barbora Grineviciute, Justinas Besusparis, Arvydas Laurinavicius, Mohammad Ilyas, A cascade-learning approach for automated segmentation of tumour epithelium in colorectal cancer, Expert Syst. Appl. 118 (2019) 539–552.
[8] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Michal J. Wesolowski, Kevin A. Schneider, Ralph Deters, Automatic polyp segmentation using convolutional neural networks, 2020.
[9] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Michal J. Wesolowski, Kevin A. Schneider, Ralph Deters, A hybrid deep learning architecture for leukemic b-lymphoblast classification, in: 2019 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, 2019, pp. 271–276.
[10] Pavel Yakubovskiy, Segmentation models. https://github.com/qubvel/segmentation_models, 2019.
[11] Kemeng Chen, Ning Zhang, Linda Powers, Janet Roveda, Cell nuclei detection and segmentation for computational pathology using deep learning, in: Proceedings of the Modeling and Simulation in Medicine Symposium, Society for Computer Simulation International, 2019, pp. 12.
[12] Philipp Kainz, Michael Pfeiffer, Martin Urschler, Semantic segmentation of colon glands with deep convolutional neural networks and total variation segmentation. arXiv preprint arXiv:1511.06919, 2015.
[13] Aicha Bentaieb, Jeremy Kawahara, Ghassan Hamarneh, Multi-loss convolutional networks for gland analysis in microscopy, in: Proceedings - International Symposium on Biomedical Imaging, 2016.
[14] Matko Šarić, Mladen Russo, Maja Stella, Marjan Sikora, Cnn-based method for lung cancer detection in whole slide histopathology images, in: 2019 4th International Conference on Smart and Sustainable Technologies (SpliTech), IEEE, 2019, pp. 1–4.
[15] Shujiao Sun, Huining Yuan, Yushan Zheng, Haopeng Zhang, Zhiguo Jiang, Cancer sensitive cascaded networks (csc-net) for efficient histopathology whole slide image segmentation, in 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 476–480.
[16] Gonçalo Figueira, Yaqi Wang, Lingling Sun, Huiyu Zhou, Qianni Zhang, Adversarial-based domain adaptation networks for unsupervised tumour detection in histopathology, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 1284–1288.
[17] Onur Can Koyun, Tulay Yildirim, Adversarial Nuclei Segmentation on H&E Stained Histopathology Images. 2019.
[18] Vahid Khanagha, Sanaz Aliari Kardehdeh, Context aware lung cancer annotation in whole slide images using fully convolutional neural networks, in: International Conference on Image Analysis and Recognition, Springer, 2019, pp. 345–352.
[19] Faisal Mahmood, Daniel Borders, Richard Chen, Gregory N. McKay, Kevan J. Salimian, Alexander Baras, Nicholas J. Durr, Deep Adversarial Training for Multi-Organ Nuclei Segmentation in Histopathology Images, IEEE Trans. Medical Imaging (2019).
[20] Zitao Zeng, Weihao Xie, Yunzhe Zhang, Yao Lu, RIC-Unet: An Improved Neural Network Based on Unet for Nuclei Segmentation in Histology Images, IEEE Access (2019).
[21] Marc Macenko, Marc Niethammer, James S. Marron, David Borland, John T. Woosley, Xiaojun Guan, Charles Schmitt, Nancy E. Thomas, A method for normalizing histology slides for quantitative analysis, in: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE, 2009, pp. 1107–1110.
[22] Hao Chen, Xiaojuan Qi, Lequan Yu, Pheng Ann Heng, DCAN: Deep Contour-Aware Networks for Accurate Gland Segmentation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.
[23] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, Nong Sang, Learning a Discriminative Feature Network for Semantic Segmentation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2018.
[24] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassasni, Michal J. Wesolowski, Kevin A. Schneider, Ralph Deters, Automatic detection of coronavirus disease (covid-19) in X-ray and ct images: A machine learning-based approach. arXiv preprint arXiv:2004.10641, 2020.
[25] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Michal J. Wesolowski, Kevin A. Schneider, Ralph Deters, Breast cancer diagnosis with transfer learning and global pooling, in: 2019 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, 2019, pp. 519–524.
[26] Olaf Ronneberger, Philipp Fischer, Thomas 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] Wenlong Deng, Yongli Mou, Takahiro Kashiwa, Sergio Escalera, Kohei Nagai, Kotaro Nakayama, Yutaka Matsuo, Helmut Prendinger, Vision based pixel-level bridge structural damage detection using a link aspp network, Automat. Construct. 110 (2020) 102973.
[28] Shakiba Moradi, Mostafa Ghelich Oghli, Azin Alizadehasl, Isaac Shiri, Niki Oveisi, Mehrdad Oveisi, Majid Maleki, Jan Dhooge, Mfp-unet: A novel deep learning based approach for left ventricle segmentation in echocardiography, Physica Med. 67 (2019) 58–69.
[29] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.
[30] Karen Simonyan, Andrew Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[31] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, in: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017.
[32] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications Andrew. Reports of Practical Oncology and Radiotherapy, 2009.
[33] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, Rethinking the Inception Architecture for Computer Vision, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016.
[34] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, in: 31st AAAI Conference on Artificial Intelligence, AAAI 2017, 2017.
[35] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1492–1500.
[36] Jie Hu, Li Shen, Gang Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[37] Rajesh Mehra, et al., Breast cancer histology images classification: Training from scratch or transfer learning? ICT Express 4 (4) (2018) 247–254.
[38] Joey Tianyi Zhou, Sinno Jialin Pan, Ivor W. Tsang, A deep learning framework for hybrid heterogeneous transfer learning, Artif. Intell. 275 (2019) 310–328.
[39] Yusuf Celik, Muhammed Talo, Ozal Yildirim, Murat Karabatak, U. Rajendra Acharya, Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images, Pattern Recognit. Lett. (2020).
[40] Jiahui Li, Shuang Yang, Xiaodi Huang, Qian Da, Xiaoqun Yang, Zhiqiang Hu, Qi Duan, Chaofu Wang, Hongsheng Li, Signet Ring Cell Detection with a Semi-supervised Learning Framework, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2019.
[41] DigestPath. https://digestpath2019.grand-challenge.org/.
[42] Sai Chandra Kosaraju, Jie Hao, Hyun Min Koh, Mingon Kang, Deep-hipo: Multi-scale receptive field deep learning for histopathological image analysis, Methods (2020).
[43] Kaushiki Roy, Debapriya Banik, Debotosh Bhattacharjee, Mita Nasipuri, Patch-based system for classification of breast histology images using deep learning, Comput. Med. Imaging Graph. 71 (2019) 90–103.
[44] Sebastian Ruder. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747, 2016.
[45] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, A comparative study of deep learning architectures on melanoma detection, Tissue Cell 58 (2019) 76–83.