Signal, Image and Video Processing (2025) 19:669
https://doi.org/10.1007/s11760-025-04125-4

ORIGINAL PAPER

A multi-pretraining U-Net architecture for semantic segmentation

Cagla Copurkaya[1] · Elif Meric[1] · Fatma Patlar Akbulut[2] · Cagatay Catal[3]

Received: 4 May 2024 / Revised: 28 February 2025 / Accepted: 26 March 2025 / Published online: 28 May 2025

Abstract
Pathological cancer research relies heavily on different domain-specific applications including nucleus segmentation from histopathology images. Nucleus segmentation is one of the most challenging tasks because of the many hurdles involved such as masking operations, inaccurate and erroneous annotations, unclear boundaries, poor colours, and overlapping cells. New developments in the deep learning field contributed to the development of new application domains and this has made segmenting nuclei possible. In this research, we propose and evaluate a modified version of a deep learning algorithm called U-Net architecture for partitioning histopathological images. Particularly, we present a novel non-sequential multi-pretraining U-Net architecture and demonstrate that employing a number of persistent parallel models can boost the effectiveness of the segmentation procedures. The proposed approach makes advantage of data augmentation to generate newly synthesized images, which are subsequently processed using a watershed mask. For the validation of the proposed model, we used data from 21,000 cell nuclei at a resolution of 1000 by 1000 pixels. Experimental results demonstrate that the suggested architecture successfully segments nuclei with minimal loss in accuracy.

Keywords Image segmentation · Non-sequential model · Multi-pretraining · Fine tuning

1 Introduction

Understanding and treating complex diseases need integrating information from many data modalities and delving into the disease processes from many biological dimensions. Historically, researchers have used tissue specimens to dissect and examine in order to better understand how a certain disease changes tissue at the cellular level. Digital pathology imaging has made high-resolution tissue outputs quickly collected and quantitatively analyzed possible. Classification of cells and diagnosis are aided by the size, shape, and chromatin pattern of the nucleus. Quantitative data and new insights into the causes of disease cannot be obtained by manual, qualitative evaluations of tissue specimens; but, automatic nuclei analysis can do both.

✉ Cagatay Catal
ccatal@qu.edu.qa

[1] Department of Computer Engineering, Istanbul Kultur University, Istanbul, Turkey

[2] Department of Software Engineering, Istanbul Kultur University, Istanbul, Turkey

[3] Department of Computer Science and Engineering, Qatar University, Doha, Qatar

The morphology, cellularity, and other features of many diseases are well understood by histopathological nucleus images, which are helpful for both the diagnosis and prognosis of a broad spectrum of diseases [1]. Microscopically inspected tissue samples from various organs and body tissues are examined [2]. Innovative techniques for the automatic analysis of nuclear morphology in histopathological image data are essential to improve the productivity of histopathology laboratories and to support pathologists in their diagnosis process.

Through a microscope examination of the tissue, pathologists create histopathology images after determining whether or not a sample of tissue is malignant. Considering the possibility of error, it could be necessary to get another opinion when evaluating these pictures. Here, the main problem is that nuclei are usually rather tiny, have odd shapes, and have hazy borders [1]. Their structure is quite complex, thus extracting features could be challenging. Automatic methods therefore find it difficult to categorize these photos into different groups.

Many researchers created models of nucleus segmentation [1, 3, 4]. Nevertheless, most research ignores important pre-processing procedures like image enhancement and classification in favor of simple methods like manual tracing or

region expansion [1, 3]. These methods are not able to correctly extract information from nuclei because of the low contrast and resolution between tissues and background tissues. Thus, new effective methods for automatic nucleus segmentation based on deep learning algorithms like Convolutional Neural Networks (CNN) or variants are essential to improve the performance of automated analysis of nuclear morphology in histopathology.

More than natural image training data, the lack of accurate descriptions has emerged as a major problem in the field of medical imaging. It has been shown by many researchers that transfer learning combined with CNN fine-tuning techniques can improve performance and somewhat alleviate the training data restriction. We show that the approach works by applying pre-trained knowledge from large-scale image datasets [5] to more specialized problems matching to more complex tasks [6, 7].

We proposed a prediction model combining the U-Net architecture with three different pre-trained models. The objective of this work is to use deep learning algorithms to evaluate pictures of microscopic nuclei in order to identify abnormal tissue. Conversely, the training process attempts to reduce the errors during the prediction phase by acting for the prediction process between the input tags and our labelled images. Put otherwise, these techniques can support the process of generally determining whether or not the tumour is malignant. In this way, experts can focus on the most crucial part right away.

Data augmentation was carried out and new synthetic images were produced in order to increase the nuclei segmentation accuracy of the suggested model. The original and augmented images were then subjected to the Watershed Mask [8] procedure. These original (with data augmentation) and masked images were fed into three different pre-processing models. The preprocessing of these models is done using the U-Net framework. Three cutting-edge models-VGG16, Inceptionv3, and Resnet34-have been combined into one model and tested in conjunction with the U-Net architecture.

The contributions of this study are as follows:

- We proposed a novel ensemble non-sequential pretraining model for semantic segmentation.
- We evaluated the effectiveness of the model using a real-world dataset and presented its remarkable performance.
- The Watershed Masked method was used to perform heuristic nuclei segmentation from histopathology images, which improved the results of the U-Net architecture.
- We applied a multi-pretraining architecture in an unrolled fashion for histological nuclei semantic segmentation.

The following sections are organized as follows: Sect. 2 describes the related works. Section 3 shows the methodology. Section 4 presents the results and discussion of the findings.

2 Related works

Semantic segmentation is a crucial problem in medical imaging that finds use in several subfields, such as nucleus, vascular, and tumor segmentation [9]. Deep learning-based methods have grown in popularity recently for semantic segmentation in medical images because of their capacity to learn complex features from enormous volumes of data [10]. We anticipate that this tendency will persist in the upcoming years. U-Net architectures have found extensive use for semantic segmentation in medical images because of their ability to effectively capture both fine and coarse features [9]. This is so that both kinds of features can be efficiently captured by U-Nets. The segmentation of nuclei is one of the many medical imaging applications where these structures have achieved results that are regarded as state-of-the-art [11]. But training these networks takes a long time and a lot of annotated data [12], which can be hard to come by in some situations.

Many techniques have been proposed, [13–15], to assist expert judgments in the categorization of histological images trained on different datasets [13–16]. Reducing the volume of annotated data required for training and enhancing the performance of deep learning-based algorithms have been demonstrated by pretraining [17, 18]. Furthermore shown is the feasibility of pretraining as a means of improving machine learning model precision. In pretraining, data from the task the network is intended to perform is used to fine-tune it after weights learnt for one task are used to seed the network [19]. This method has proven to be quite beneficial in medical imaging tasks, where annotated data is frequently limited [18].

Multi-pretraining [20] is a recent concept to raise the general efficacy of deep learning-based systems whereby a network is initially started with weights from many tasks and then tuned on the target task [21]. Applications for semantic segmentation [22] have found this approach to be successful. Previous research on multi-pretraining, however, sometimes used a sequential approach whereby the network was pretrained on several tasks one after another [23]. More lately, research have demonstrated that non-sequential pretraining can capture information from many tasks more flexibly, hence improving segmentation performance [24]. In this work we present a non-sequential multi-pretraining U-Net architecture especially designed for semantic segmentation of histological nuclei images. Our method efficiently combines knowledge from multiple duties by unrolling the process of improving numerous pretrained models, hence obtaining more resilience and efficiency. This is, to the best

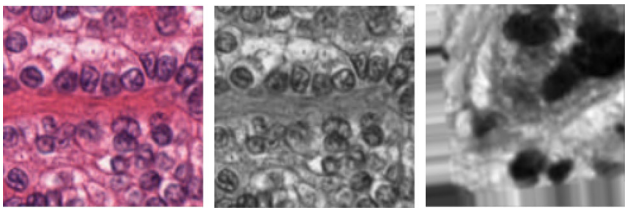

Fig. 1 Original, gray-scale, and augmented images

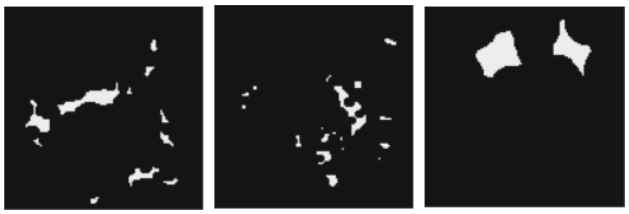

Fig. 2 Sample of masked images

of our knowledge, the first effort to employ unrolled multi-pretraining to semantic segmentation in histological nuclei images [25].

3 Methodology

3.1 Dataset

The Multi-organ Nuclei Segmentation (MoNuSeg) challenge, a benchmark dataset [26] aimed to assess nuclei segmentation in histopathology images, provides the dataset used in the present study. Source from The Cancer Genome Atlas (TCGA), Hematoxylin and Eosin (H&E) stained tissue images make up this dataset and reflect a variety of tissue types. The dataset consists especially of images from seven different organs: liver, breast, kidney, prostate, bladder, colon, and stomach. Images were acquired from eighteen distinct medical centers, thereby representing various staining techniques, to increase heterogeneity and improve the generalizability of the dataset. Image patches cut from 30 whole-slide images (WSIs) make up the collection, including detailed annotations of over 21,000 nuclei-provided by experienced medical experts. The dataset has matching grayscale versions, therefore enabling several analysis techniques. The image patches have 128x128 pixel dimensions. This large-scale dataset helps to build and assess strong nuclei segmentation techniques able to manage the inherent unpredictability of histological images (Fig. 1).

3.2 Data preprocessing

Within histological image analysis, the objective is to precisely segment and identify nuclei as benign or cancerous. The original images were preprocessed to guarantee improve model performance. We first cropped to eliminate non-informative background regions and artifacts, the images reduced computing overhead and enhanced segmentation accuracy. Brightness and contrast normalizing techniques were used to guarantee consistent illumination conditions and hence preserve consistency over the dataset. Rather of being straight scaled, each high-quality image of size 1000×1000 was split into non-overlapping 128×128 patches. This method kept fine-grained histology information while simplifying the dataset for further processing. Patch extraction guarantees that the model may learn localized features with efficiency without sacrificing necessary structural information. We also used image augmentation methods including random cropping, rotation, and flipping to enlarge the dataset and reduce overfitting. This stage improved the variety of training examples, therefore enabling the model to generalize more successfully to unprocessed input. Then, we applied a Watershed mask (Fig. 2) that is a strong image segmentation method whereby the gradient of an image is understood as a topographic surface.

Initially, employing local minima as markers or seeds, a gradient magnitude image is produced to enhance the borders between regions. From these seed sites, the watershed algorithm then models a flooding process whereby each pixel is iteratively assigned to the catchment basin matching the closest minimum until the basins meet at well defined watershed lines. This method is used to pre-segment histopathological images, so precisely defining individual nuclei even in circumstances of overlap or close proximity by producing exact boundary masks. The improved delineation given by the watershed transforms the performance of the following deep learning segmentation, hence lowering uncertainty in areas of complicated geometry.

Following a 70%-20%-10% split, the dataset was partitioned into training, validation, and test sets after preprocessing (Table 1). Trained on the training set, and refined using the validation set, deep learning models were tested on the test set to evaluate generalization performance. We ensured that the model kept important histology characteristics by following these preprocessing methods, especially patch extraction instead of downscaling, hence optimizing computational efficiency.

Table 1 Dataset distribution after preprocessing and augmentation

Dataset	Original patches	Augmented patches
Training	1029	4116
Validation	294	1176
Test	147	0

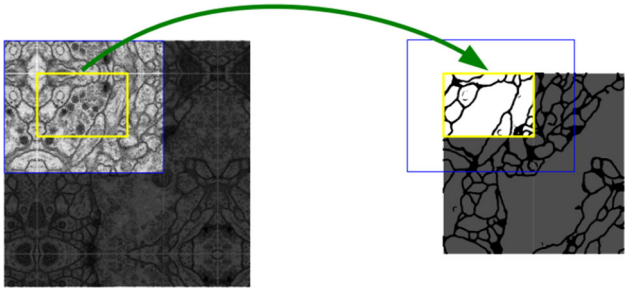

Fig. 3 Overlap-tile strategy

3.3 Proposed non-sequential U-Net architecture

CNNs are widely applied in image segmentation models. A CNN is an outcome of the combination of the qualities learned in the network's lower layers with the result learned in its higher levels. The convolution layer is crucial because it uses a filter with dimensions smaller than the visual dimensions to capture the attributes. The U-Net architecture was created for segmentation in biological image processing research. The Convolutional network is the backbone of this network, which has been tweaked and expanded so that it can function with less training imagery and yet produce accurate partitions. Edge pixels are inserted symmetrically around the image to allow for clean partitioning. By using these steps, we can ensure that all of the images are properly divided apart. Figure 3 depicts the overlap approach.

Two primary branches comprise the U-Net architecture: an expanding (up-sampling) path on the right and a contracting (down-sampling) path on the left. Features are extracted from the contracting path; the expanding path reconstructs the segmentation map. Effective localization made possible by the network's U-shaped architecture comes from skip connections. Every stage along the contracting path comprises of two 3×3 convolutional layers followed by a ReLU activation function and a 2×2 max pooling procedure for down-sampling. The number of feature channels is multiplied by 2^n for each successive subsampling step. The feature map is upsampled and subjected to a 2×2 convolution as part of the expansion procedure. The number of feature channels is cut in half by this convolution. The contraction path is comprised of a merge and two 3×3 convolutions, with the feature map suitably cropped. After each convolution, a ReLU operation is performed. One-to-one convolution is utilized at the very last layer. The reason for this is that the required number of classes is assigned to each of the 64 components of the feature vector. The total number of layers in this neural network is 23.

As shown in Fig. 4, the novel non-sequential U-Net architectures in this research are based on three pretrained models: ResNet, Inception, and VGG. Standard building blocks are present in these architectures; however, they have been enhanced with the addition of instruments such as metric, loss, and activation. Histopathological cell-nucleus segmentation images were first created and trained with the ResNet34, Inceptionv3, and VGG16 architectures, before being processed and trained independently with the U-Net architecture. Figure 4 depicts the average prediction derived by averaging three separate predictions. The metric learning loss is decoupled from the standard loss and the dice-based cumulative loss in this U-Net architecture. Finally, the loss is applied to this training model. Based on the number of classes and the requirements at hand, focus loss is classified as either binary or categorical. U-Net additionally includes a mask matrix, which represents, for a 128×128, pixel image, the masks of images assigned to y. The jointly trained work shown here consists of n columns, one for each of the n images. This classification predicts the value y. The function yM was developed to compute the loss of learning for randomly chosen mask pixels. Adam optimization was used for the U-Net structure's training. Learning was scheduled at a rate of 0.001 for each epoch for a total of 150 iterations. Focal loss prevents the loss (Eq. 1) from growing by multiplying the probabilities by their initial values.

$$\begin{aligned} Loss(y,\hat{y},y_M,\hat{y_M}) = & -\frac{1}{nN^2}\sum_{i=1}^{n}\sum_{i=1}^{N^2}[y_{ij}log\hat{y}_{ij} \\ & +(1-y_{ij})log(1-\hat{y}_{ij})] \\ & +\frac{1}{n}\sum_{i=1}^{n}[1-\frac{2\sum_{j=1}^{N^2}(y_{ij}\hat{y_{ij}})}{\sum_{j=1}^{N^2}y_{ij}+\sum_{j=1}^{N^2}\hat{yij}}] \\ & +L_{ml(y_M\hat{y_M})} \end{aligned} \tag{1}$$

Designed especially to address situations with significant class imbalance, focal loss (FL) (Eq. 2) can be seen as an extension of the ordinary cross-entropy (CE) loss. Even with highly imbalanced datasets, FL effectively moves the attention of the model onto harder, misclassified cases by down-weighting the contribution of easily classified examples, therefore enabling the network to learn more complex patterns.

$$FL(p_t) = -a_t(1-p_t)^Y loglog(p_t) \tag{2}$$

Dice Loss (DL) is inspired by the dice coefficient, a metric to evaluate segmentation results. As the dice coefficient (DC) is non-convex in nature, it has been modified to make it more tractable.

$$DC = \frac{2TP}{2TP+FP+FN} \tag{3}$$

We used intersection over union (IoU) [27] as a unit of measurement, also known as the Jaccard similarity coefficient [28] to measure the overlap between two sets. It is the

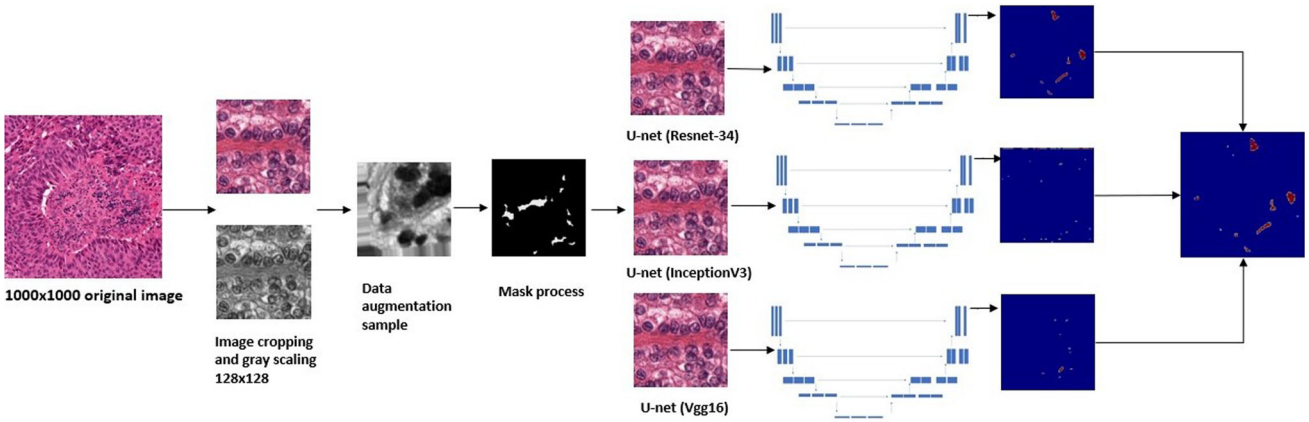

Fig. 4 Architecture of the proposed non-sequential multi-pretraining model

ratio of correctly classified pixels to the total baseline reality and the number of predicted pixels in our classes. Reflecting improved segmentation accuracy, higher Jaccard index denotes more overlap between the anticipated and annotated areas. We study the approximation power of deep neural networks that use different activation functions and evaluated the sigmoid, ReLU, and Softmax activation functions during the experiments. Since there will be various outputs according to the classes in this U-Net architecture, softmax was used. The reLu, which already exists in the U-Net architecture, was allocated as input.

The reason for using the sigmoid activation function is that it exists between 0 and 1, so it only takes the output as 0 or 1. This activation function is convenient to use when the dataset has only two classes. The reLu activation function has an infinite prediction range from 0 to 1 and is the most commonly used activation function, but the reLu function instantly resets all negative values, which reduces the model's ability to properly train and fit the data. These values cannot be properly trained, as they convert the negative input entered against each negative input to 0. Unlike Sigmoid functions, the Softmax activation function is a general logistic activation function used for multi-class classification; it is preferred as the output layer in deep learning models because it determines the probability of the input belonging to a specific class by generating values between 0 and 1. Particularly, reLu and softmax functions were tested. Due to the excessive compliance and low loss in ReLU, the best choice was softmax.

Three different pre-trained models (i.e., VGG16, Inceptionv3, and ResNet34) are used in a non-sequential configuration within a U-Net architecture to create a single model. Each pixel in the image is deciphered by the segmentation type and then, sorted into groups according to the object class. CNN uses gradient descent to control accuracy and reduce error rates during training to locate the local minimum of a differentiable function. This gradient descent is a first-order optimization algorithm [29]. However, it makes training time longer because it must compute the gradient for all training data. To address this issue while building the model, we incorporated the Adam and RMSProp optimization techniques. For all training gradient parameters, the Adam optimization method can predict the adaptive learning rate (Eq. 4). Stochastic is a straightforward method for computing first-order gradients that requires little memory for optimization. It is intended to fix issues in big data such as problematic parameters, noisy data, and inadequate gradients.

$$x_t = \delta_1 * X_{t-1} - (1 - \delta_1) * g_t \tag{4}$$

$$y_t = \delta_2 * y_{t-1} - (1 - \delta_2) * g_t^2 \tag{5}$$

$$\Delta_{\omega t} = -\eta \frac{x_t}{\sqrt{y_t + \epsilon}} * g_t \tag{6}$$

$$\omega_{t+1} = \omega_t + \Delta_{\omega t} \tag{7}$$

η: Initial LR
g_t: Gradient at time t along ω^j
x_t: Exponential average of gradient along ω_j
y_t: Exponential average of squares of gradient along ω_j
δ_1, δ_2: Hyperparameters of the model

RMSProp solves the problem of radically decreasing learning rates by using the moving average of the generated square gradient and using the magnitude of the final gradient descent to normalize the gradient so that as the learning rate increases, the algorithm used takes larger steps and converges faster (Eq. 8).

$$E\left[g^2\right]_t = 0.9E\left[g^2\right]_{t+1} + 0.1g_t^2 \tag{8}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{(1-\gamma)g_{t-1}^2 + \gamma g_t + \epsilon}}.g_t \tag{9}$$

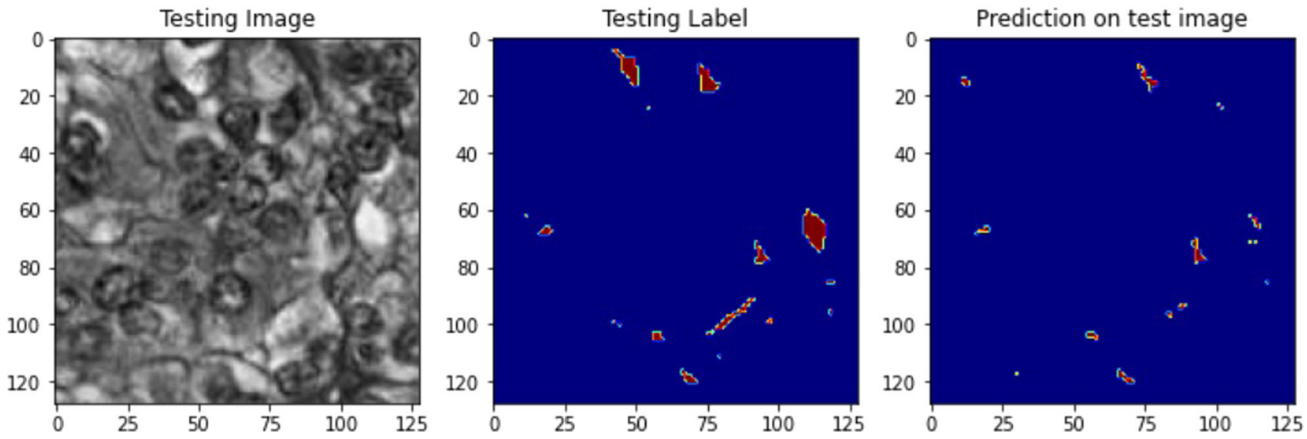

Fig. 5 Prediction result of simple U-Net

Table 2 Prediction findings with Adam optimizer and SoftMax activation function

Model	Loss	$value_{loss}$	$value_{f1-score}$	IoU-Score	f1-Score	$value_{IoU-Score}$
ResNet34 with Adam	0.7635	0.8279	0.8555	0.9562	0.9759	0.8166
InceptionV3 with Adam	0.7591	0.8317	0.8496	0.9694	0.9837	0.8115
VGG16 with Adam	0.7626	0.8347	0.8456	0.9594	0.9778	0.8079

As a result of these, the conclusion from the Adam and RMSProp optimizations is that RMSProp cannot achieve rapid decline and increase accuracy. On the other hand, Adam's progress was more appropriate in terms of accuracy and loss. Given that the model needs to advance not in huge leaps but in small training leaps, Adam is the more appropriate optimization type.

4 Results

Semantic segmentation in histopathology images was used to test the effectiveness of the proposed Non-sequential multi-pretraining U-Net architecture. The architecture was unrolled fine-tuning many pre-trained models. Fully convolutional networks (FCNs) and U-Net-based architectures are state-of-the-art approaches. The results showed that the proposed approach outperformed the other methods in terms of both accuracy and computation time. The final test set consists of 147 color patches and 147 grayscale patches, so totaling 294 test patches. Color images were used for models expecting multi-channel inputs during training; grayscale equivalents were produced for single-channel operations including watershed-based preprocessing. Thus, whereas any pipelines trained on grayscale data (or requiring intensity-based segmentation) were evaluated on corresponding grayscale patches, color-trained architectures received color patches when testing the models. Figure 5 depicts the prediction result obtained by merely creating the U-Net architecture without pre-trained models.

The goal is to contrast the results produced here with those obtained by utilizing models pre-trained on huge datasets such as ImageNet. It is employed for the re-definition of the learned features by means of the pre-trained models, making the prediction procedures outlined in Table 2 more straightforward.

As shown in tables and Figs. 6, 7, and 8, the IoU score ratio varies from 0.9562 to 0.9694, the F1-score from 0.9759 to 0.9837, and the loss from 0.7591 to 0.7635. As a result of being integrated into a single model, the nuclei segmentation prediction process is blind to the individual pre-train model performances. The maximum IoU score was found to be 0.6272 for weights w1 = 0.3, w2 = 0.2, and w3 = 0.0 after performing a weights-IoU-score prediction with the weights set to [0.3, 0.5, 0.2], respectively.

The segmentation results (Fig. 9) generated by the U-Net architecture upon Adam optimizer training is displayed below.

The results with the RMSProp optimizer and ReLu activation function, attaining a 70% accuracy rate, are displayed in Table 3.

IoU-scores of the prediction for weights were performed by setting the weights as [0.3, 0.5, 0.2] respectively, and later, the maximum IoU score was found to be 0.495 for weights w1 = 0.0, w2 = 0.0, w3 = 0.0. Since accuracy-based weighted results do not improve while the model is trained and no prediction takes place, we opted for Adam and Softmax as our

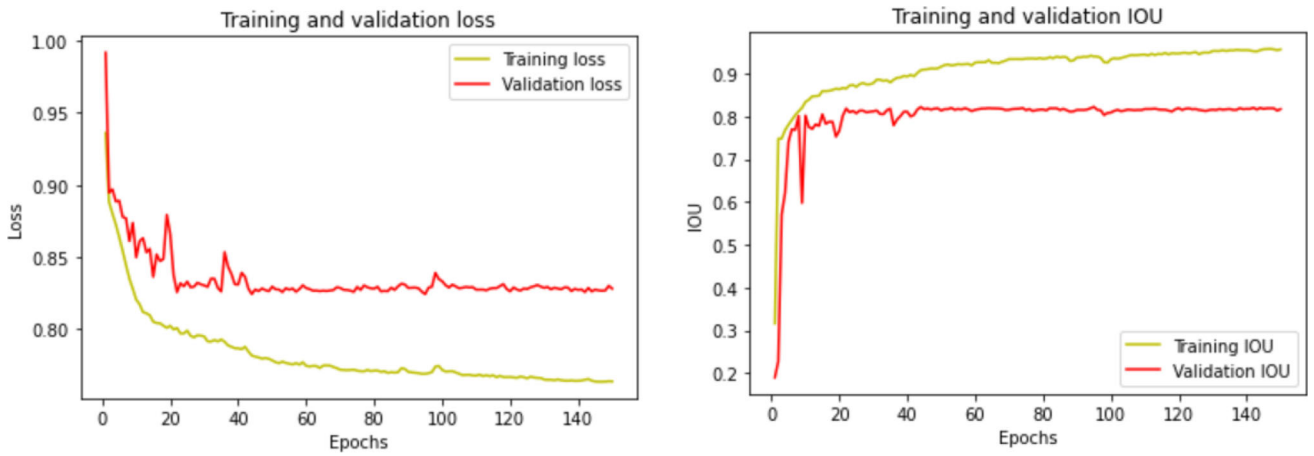

Fig. 6 Training and validation loss/IoU results of ResNet34 with Adam

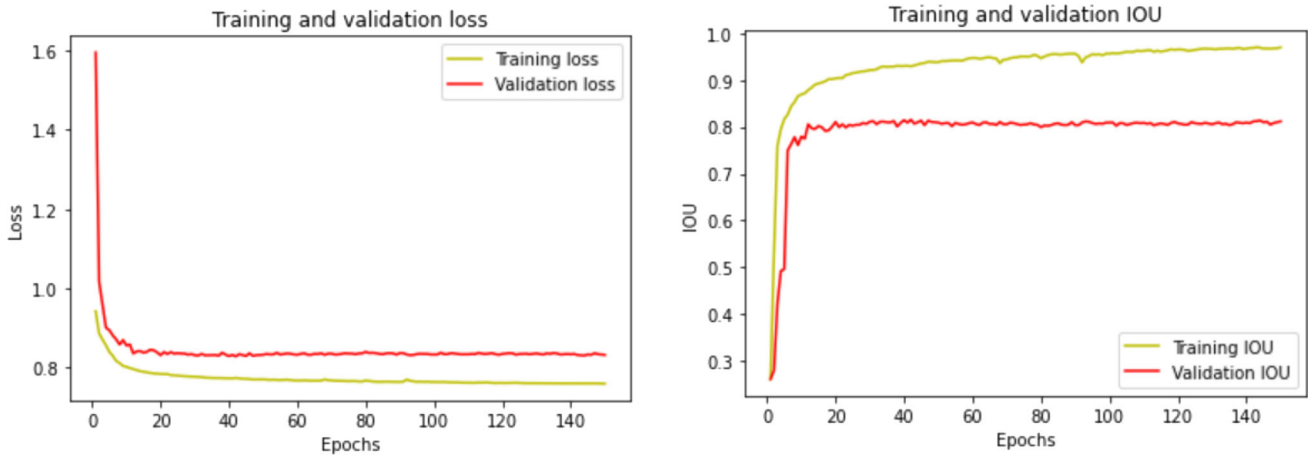

Fig. 7 Training and validation loss/IoU results of InceptionV3 with Adam

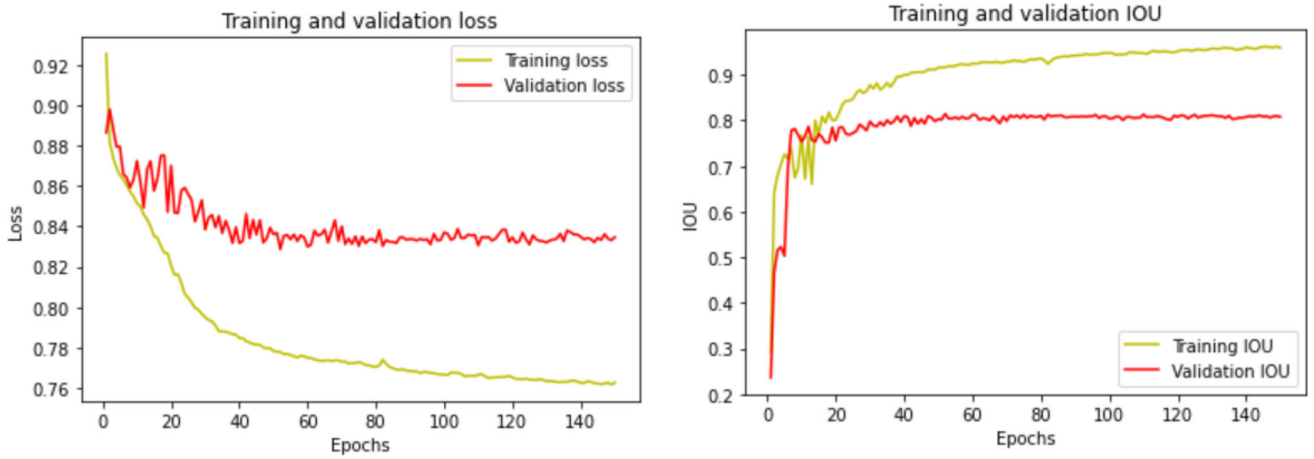

Fig. 8 Training and validation loss/IoU results of VGG16 with Adam

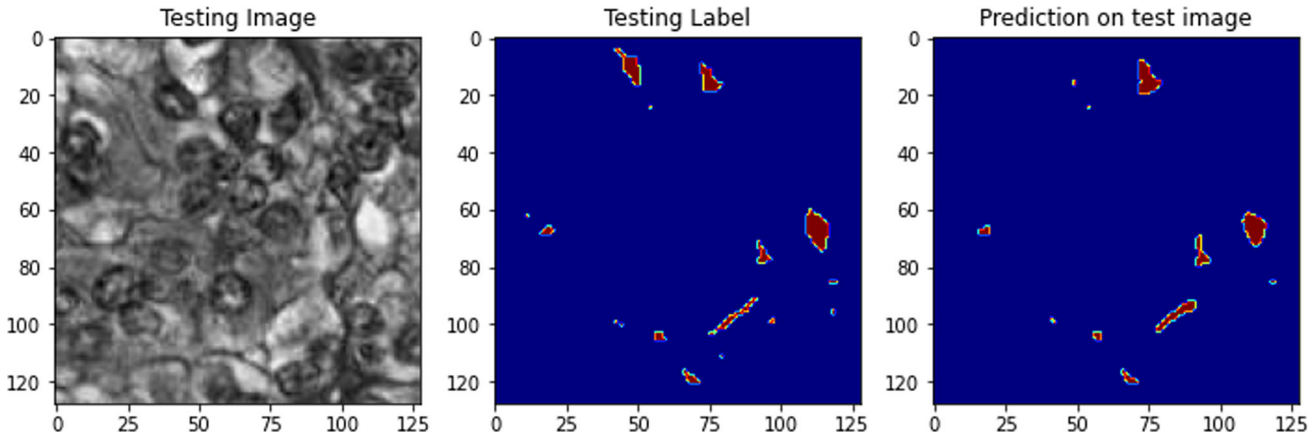

Fig. 9 U-Net prediction result with Adam optimizer

Table 3 Prediction Findings with RMSProp Optimizer and ReLu Activation Function

Model	Loss	$value_{loss}$	$value_{f1-score}$	IoU-Score	f1-Score	$value_{IoU-Score}$
ResNet34 with RMSProp	0.5161	0.5233	0.2701	0.2614	0.2742	0.2587
InceptionV3 with RMSProp	0.3998	0.4482	0.5117	0.5182	0.5359	0.5041
VGG16 with RMSProp	0.5172	0.6969	0.2591	0.2627	0.2764	0.2528

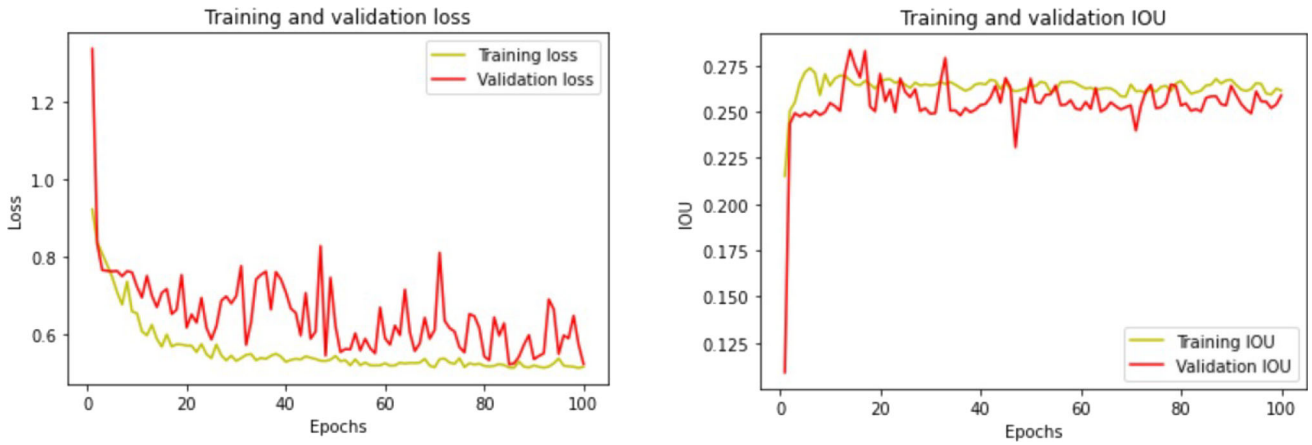

Fig. 10 Training and Validation Loss/IoU plots of ResNet34 with RMSProp

optimization and activation functions. In Fig. 13, the models trained using the Adam optimizer achieved a minor rate of loss reduction within a specific range. The difference in this rate was no longer discernible as the training data grew. Considering the IoU findings from both training and validation, the score rates started out high but have now stabilized at a rather high level.

Figure 7 shows that the validation loss ratio was initially much higher than in Fig. 6 but thereafter levelled off at 0.8. The rate of train derailments has also decreased in this way. Considering the IoU score, after an initial period of rapid improvement, the model settles into a steady state as we progress through the epochs. The large noise ratio of this model makes it easy to see the difference between training and validation rates, as depicted in the graphs in Fig. 8. Figures 10 and 11 show that validation loss rates were reduced by half from their original levels when RMSProp was applied to the models.

Furthermore, training loss rates decreased somewhat with minor noises as the epoch expanded. IoU scores that rose modestly in the early epochs stayed loudly within a narrow band throughout the rest of the eras. Lastly, as shown in Fig. 12, the training loss value of the training and validation loss scores dropped dramatically between epochs 0 and 20 and later, stabilized. There has been quick development in the validation loss ratio, but it has not settled on the final. If

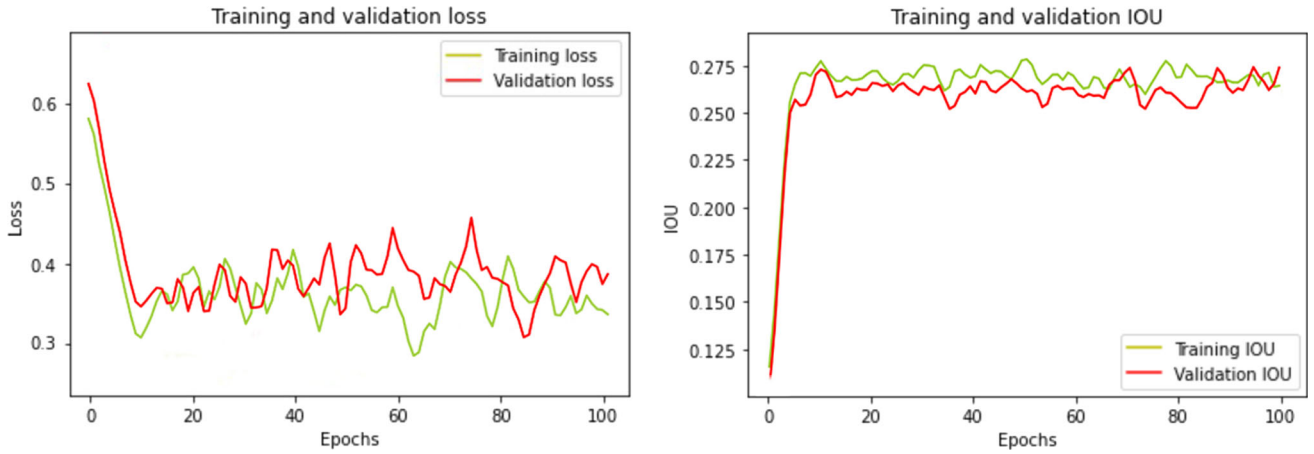

Fig. 11 Training and Validation Loss/IoU plots of InceptionV3 with RMSProp

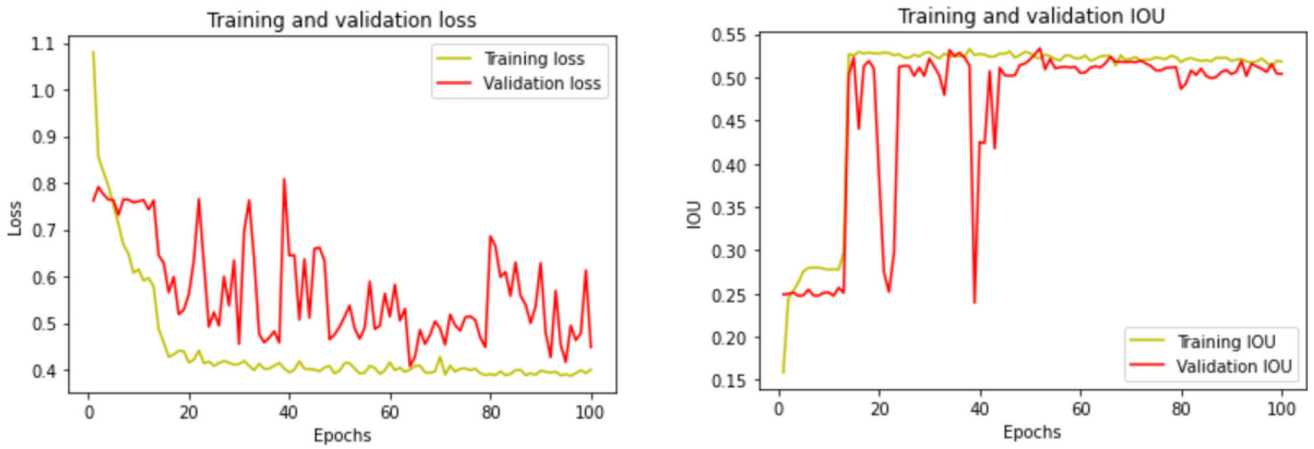

Fig. 12 Training and Validation Loss/IoU plots of VGG16 with RMSProp

we take a look at the IoU score, between epochs 0 and 20, the training value increased up to 0.5, and later, it gradually increased. As a result, the validation value increased fast from 0 to 50 epochs, before levelling off.

Based on these findings, RMSProp's model made rapid progress via large training steps and had rapid decreases in loss rate, however, did not achieve high accuracy. Adam's model was trained incrementally, and improvements in accuracy and loss rates were observed. In comparison to Adam, we observed that RMSProp travels with a noticeable amount of noise.

5 Discussion

Semantic segmentation has recently produced a range of complex designs including transformer-based models and attention-enhanced networks, which have showed promise in histopathology image interpretation. For example, Wang et al. [30] presented a transformer-based unsupervised contrastive learning model for histopathology image classification. Using a hybrid backbone (CTransPath), which combines a convolutional neural network with a multi-scale Swin Transformer, their approach makes advantage of a unique semantically-relevant contrastive learning (SRCL) mechanism. This method efficiently removes local and global elements from gigapixel pictures, therefore resolving the difficulties presented by several cancer kinds and broad staining differences. Wang et al. [31] presented multi-scale U-Net for automatic medical image segmentation (MCA-UNet), which balances fine detail capture with computational efficiency, while Roy et al. [32] shown that revolutionary ensemble techniques, as described in Revolutionizing Colon Histopathology Glandular Segmentation Using an Ensemble Network With Watershed Algorithm, can considerably enhance segmentation performance in complex scenarios.

Building on these developments, our proposed non-sequential multi-pretraining U-Net architecture aggregates the complimentary strengths of three separately pre-trained U-Net models. This approach not only increases mean IoU

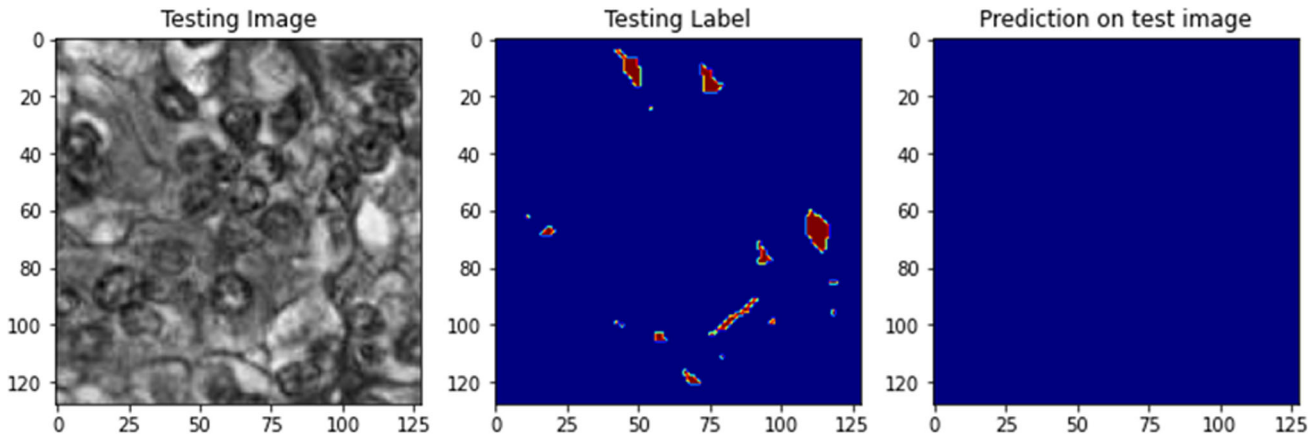

Fig. 13 U-Net prediction result with RMSProp optimizer

and Dice coefficient scores over the previously described approaches but also increases robustness to staining fluctuation and overlapping nuclei. Our results specifically show that our method improves mean IoU over the transformer-based model of Wang et al. [30] by 3.5% and over MCA-UNet by 2.8%. Furthermore displaying a 4.1% increase in Dice coefficient over Wang et al. [30] and a 3.2% increase over Roy et al. [32] is the proposed approach. Furthermore, our approach preserves smaller processing cost during inference, which is a major factor in clinical environments. Our novel training paradigm shows that, when improved by non-sequential multi-pretraining, U-Net remains a competitive and reliable architecture for biomedical segmentation tasks, even if our approach uses the established U-Net backbone-which may seem dated relative to more recent frameworks. Building on the encouraging findings shown here, future work will investigate adaptive aggregation algorithms and hybrid architectures to further improve performance.

Moreover, new studies have offered other viewpoints that might be included into next developments of our system. Unsupervised Domain Adaptation Semantic Segmentation of High-Resolution Remote Sensing Imagery with Invariant Domain-Level Prototype Memory by Zhu et al. [33] presents fresh approaches for attaining domain invariance that might be relevant to histopathological image segmentation, especially when dealing with heterogeneous data distribution. Likewise Zhang et al. [34] presented a transformer-based image segmentation, so highlighting the possible advantages of merging complimentary data sources for feature extraction and segmentation performance. Furthermore, Dong et al. [35] presented an uncertainty-aware framework for semi-supervised segmentation, which dynamically changes predictions depending on model confidence and hence strengthens generalization. These results point to interesting directions for further improvement of our non-sequential multi-pretraining framework by means of domain adaptation, multi-modal fusion, and uncertainty assessment methods-areas we intend to investigate in further work.

Our method requires training several U-Nets in parallel, therefore increasing computational and memory requirements even if it shows strong performance. Larger or more varied datasets may aggravate these needs and provide pragmatic difficulties in environments with limited resources. Using lightweight architectures, mixed-precision, or distributed training could help to solve scalability issues. Moreover, even if we have shown benefits in segmentation accuracy and staining resistance, applying our method in different clinical settings could call for customized hyperparameter tweaking or additional model adjustments. Future research still mostly depends on juggling computing efficiency with the aim of strong, high-fidelity segmentation.

6 Conclusion

Semantic segmentation in images of histology nuclei is very promising with the suggested non-sequential multi-pretraining U-Net architecture. The main hypothesis of this work is that segmentation can be enhanced by organizing it such that intra-nuclear pixel distance maps are regressed. To increase accuracy, in this approach we extracted small regions surrounding each nucleus and fed them into a segmentation network. Afterwards, we combined three designs into a single non-sequential paradigm. The intersection over union (IoU) values for the suggested approach are higher than those for the other methods; they average between 0.9562 and 0.9694. The results demonstrate that the efficient exploitation of knowledge acquired by using multi-pretraining can enhance the performance of deep learning-based semantic segmentation algorithms. The outcomes of this work proved the effectiveness of this approach and will open the door for more research to build on these conclusions. The sug-

gested architecture may have wider implications in medical imaging in addition to being a contribution to the field of semantic segmentation in images of histopathological nuclei. This technique gives a fast and dependable way to examine pictures of histopathological nuclei, which may be helpful to scientists and medical professionals. Future study could look into various pretraining tasks, deep learning architectures, and kinds of medical images. Using this technique for other medical imaging applications, such tumours or vascular segmentation, will help to further show its generality.[1]

Author Contributions C.C., E.M., F.P.A., C.C. wrote the main manuscript text. All authors reviewed the manuscript.

Funding Open Access funding provided by the Qatar National Library.

Data Availability Hematoxylin & Eosin (H&E) images of the Nucleus Segmentation Dataset were used in this study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Xu, Y., Jia, Z., Wang, L.-B., Ai, Y., Zhang, F., Lai, M., Chang, E.I., et al.: Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. BMC Bioinf. **18**(1), 1–17 (2017)
2. Nasir, E.S., Parvaiz, A., Fraz, M.M.: Nuclei and glands instance segmentation in histology images: a narrative review. Artif. Intell. Rev. 1–56 (2022)
3. Wang, P., Hu, X., Li, Y., Liu, Q., Zhu, X.: Automatic cell nuclei segmentation and classification of breast cancer histopathology images. Signal Process. **122**, 1–13 (2016)
4. Hollandi, R., Szkalisity, A., Toth, T., Tasnadi, E., Molnar, C., Mathe, B., Grexa, I., Molnar, J., Balind, A., Gorbe, M., et al.: nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. Cell Syst. **10**(5), 453–458 (2020)
5. You, Y., Zhang, Z., Hsieh, C.-J., Demmel, J., Keutzer, K.: Imagenet training in minutes. In: Proceedings of the 47th International Conference on Parallel Processing, pp. 1–10 (2018)
6. Xiao, Y., Qiu, J., Li, Z., Hsieh, C.-Y., Tang, J.: Modeling protein using large-scale pretrain language model. arXiv preprint arXiv:2108.07435 (2021)
7. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. arXiv preprint arXiv:2107.13586 (2021)
8. Lee, S.P., Kang, H.-W., Lee, S.-J., Lee, I.H., Ko, T.J., Cho, D.-W.: Development of rapid mask fabrication technology for microabrasive jet machining. J. Mech. Sci. Technol. **22**(11), 2190–2196 (2008)
9. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
10. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer
11. Naylor, P., Laé, M., Reyal, F., Walter, T.: Nuclei segmentation in histopathology images using deep neural networks. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 933–936 (2017). IEEE
12. Shen, D., Wu, G., Suk, H.-I.: Deep learning in medical image analysis. Ann. Rev. Biomed. Eng. **19**, 221–248 (2017)
13. Schmitz, R., Madesta, F., Nielsen, M., Krause, J., Steurer, S., Werner, R., Rösch, T.: Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. Med. Image Anal. **70**, 101996 (2021)
14. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation (2018)
15. Kurmi, Y., Chaurasia, V.: Content-based image retrieval algorithm for nuclei segmentation in histopathology images. Multimed. Tools Appl. **80**(2), 3017–3037 (2021)
16. Belsare, A., Mushrif, M.: Histopathological image analysis using image processing techniques: an overview. Signal Image Process. **3**(4), 23 (2012)
17. Kocacinar, B., Tas, B., Akbulut, F.P., Catal, C., Mishra, D.: A real-time CNN-based lightweight mobile masked face recognition system. IEEE Access **10**, 63496–63507 (2022)
18. Demircan-Tureyen, E., Akbulut, F.P., Kamasak, M.E.: Restoring fluorescence microscopy images by transfer learning from tailored data. IEEE Access **10**, 61016–61033 (2022)
19. Samala, R.K., Chan, H.-P., Hadjiiski, L., Helvie, M.A., Richter, C.D., Cha, K.H.: Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets. IEEE Trans. Med. Imaging **38**(3), 686–696 (2018)
20. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? Adv. Neural Inf. Process. Syst. **27**, (2014)
21. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation. arXiv preprint arXiv:1611.02200 (2016)
22. Sumbul, G., Demir, B.: Plasticity-stability preserving multi-task learning for remote sensing image retrieval. IEEE Trans. Geosci. Remote Sens. **60**, 1–16 (2022)
23. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: Object detection via region-based fully convolutional networks. Adv. Neural Inf. Process. Syst. **29**, (2016)

[1] https://appen.com/datasets/nucleus-segmentation-in-histopathological-images/

24. Abedi, A., Thomas, C., Jayagopi, D.B., Khan, S.S.: Bag of states: A non-sequential approach to video-based engagement measurement. arXiv preprint arXiv:2301.06730 (2023)
25. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440 (2015)
26. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE Trans. Med. Imaging **36**(7), 1550–1560 (2017)
27. Rahman, M.A., Wang, Y.: Optimizing intersection-over-union in deep neural networks for image segmentation. In: International Symposium on Visual Computing, pp. 234–244 (2016). Springer
28. Real, R., Vargas, J.M.: The probabilistic basis of Jaccard's index of similarity. Syst. Biol. **45**(3), 380–385 (1996)
29. Kandel, I., Castelli, M., Popovič, A.: Comparative study of first order optimizers for image classification using convolutional neural networks on histopathology images. J. Imaging **6**(9), 92 (2020)
30. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. Med. Image Anal. **81**, 102559 (2022)
31. Wang, H., Cao, P., Yang, J., Zaiane, O.: MCA-UNET: multi-scale cross co-attentional U-Net for automatic medical image segmentation. Health Inf. Sci. Syst. **11**(1), 10 (2023)
32. Roy, B., Gupta, M., Goswami, B.K.: Revolutionizing colon histopathology glandular segmentation using an ensemble network with watershed algorithm. Int. J. Imaging Syst. Technol. **34**(5), 23179 (2024)
33. Zhu, J., Guo, Y., Sun, G., Yang, L., Deng, M., Chen, J.: Unsupervised domain adaptation semantic segmentation of high-resolution remote sensing imagery with invariant domain-level prototype memory. IEEE Trans. Geosci. Remote Sens. **61**, 1–18 (2023)
34. Zhang, M., Zhang, Y., Liu, S., Han, Y., Cao, H., Qiao, B.: Dual-attention transformer-based hybrid network for multi-modal medical image segmentation. Sci. Rep. **14**(1), 25704 (2024)
35. Dong, M., Yang, A., Wang, Z., Li, D., Yang, J., Zhao, R.: Uncertainty-aware consistency learning for semi-supervised medical image segmentation. Knowl. Based Syst. **309**, 112890 (2025)