IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 May 2023, accepted 12 June 2023, date of publication 21 June 2023, date of current version 29 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3288422

RESEARCH ARTICLE

Automatic Scoring Method for Tumor IHC Images Based on Deep Learning and Its Application on P53 Protein

JINBO ZHANG[1], XIAOLEI GUO[1], WEN CAI[2], YUQI CAO[1], WEITING GE[2], PINGJIE HUANG[1], DIBO HOU[1], SHU ZHENG[2], AND GUANGXIN ZHANG[1]

[1]School of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
[2]Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education, Cancer Institute, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310000, China

Corresponding author: Yuqi Cao (yuqicao@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61873234, in part by the "Pioneer" and "Leading Goose" Research and Development Program of Zhejiang under Grant 2022C03002, and in part by the Key Technology Research and Development Program of Zhejiang Province under Grant 2021C03177.

ABSTRACT Immunohistochemical (IHC) assay is a commonly used auxiliary technique in pathological diagnosis. Compared to the conventional manual scoring methods that are complicated and time-consuming, automated scoring methods have been playing a more and more important role in the development of digital medicine due to their adaptability and consistency. This study proposes an automatic scoring model for tumor IHC images, which mainly consists of a module for extracting the regions of interest (ROI) and a feature fusion scoring network. The former module extracts the effective tissue regions and the nuclear regions as prior knowledge to exclude cytoplasmic staining interference. The feature fusion network includes two branches. The main branch network combines the structure of cross-block stitching feature maps and the frequency channel attention networks (FcaNet) to extract the features of the effective tissue region images. The other branch network extracts the color representation vector of the cell nucleus region images. The fully-connected layers combine the features from both branches to give a comprehensive final score as the result. We performed experiments on IHC images of P53 protein in colorectal cancer. The results show that the proposed P53Net achieves better classification results than the commonly used classification models, with 94.21% accuracy, 89.24% F1-Score, and 0.9136 kappa coefficient.

INDEX TERMS Immunohistochemical scoring, P53 protein, computer vision, DenseNet, attention mechanism.

I. INTRODUCTION

Statistics shows that there were about 4.57 million cancer cases and 3 million deaths in China in 2020 [1]. Timely diagnosis and treatment can greatly improve the survival rate of patients [2], [3]. The gold standard for clinical diagnosis of cancer is the pathological diagnosis [4]. However, there exist 5% to 10% of cases upon which a definitive morphological diagnosis is difficult to make with the conventional hematoxylin-eosin staining technique alone. Using Immunohistochemical (IHC) staining as an adjunct can noticeably increase the diagnostic accuracy of poorly differentiated or undifferentiated tumors [5]. Pathologists focus on the level of P53 protein expression in the nucleus rather than the cytoplasm and score IHC images based on the percentage and intensity of positive staining. Conventional IHC scoring is done manually, which is very time-consuming, therefore the shortage of medical equipment and specialized pathologists can affect the diagnosis and treatment severely. Therefore, it is important to develop an automated IHC scoring technique.

With the rapid development of artificial intelligence in recent years, intelligent and automatic detection has become an important field in digital medicine. Computer vision techniques have been applied to IHC image analysis, providing pathologists with efficient and reliable auxiliary

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Zhou.

interpretation information. The existing research can be divided into conventional image processing techniques and deep-learning-based methods.

Conventional techniques usually require manual selection of the ROI on the images [6], followed by the use of digital image processing [7] and machine learning techniques to calculate metrics such as positive staining intensity and staining range at the pixel or cellular level to obtain an IHC score after a comprehensive evaluation. Konsti et al. [8] used ImageJ image analysis software to evaluate Ki67 IHC images. They determined the segmentation threshold through the sample test, combined staining intensity and percentage for scoring, and achieved 87% evaluation accuracy with manual scoring. Roge et al. [9] used a virtual double staining technique to differentiate tissue types for Ki67 IHC images. They circumvented the effect of physical double staining on the analysis, delineated the tumor regions using a Bayesian classifier, and classified the stained cells to be negative or positive according to the set thresholds. A grid counting method was then used to calculate the number of positive and negative cells as well as the proliferation index. Conventional techniques have the advantages of good interpretability, good consistency and high repeatability. However, some division thresholds need to be determined, such as staining intensity threshold, positive percentage threshold, etc. Most current methods determine the threshold by testing on a small number of sample images. Some scholars also directly set the threshold based on experience. Whether the threshold division is accurate will greatly affect the scoring accuracy.

In recent years, intelligent medical prediction and evaluation models based on deep learning have been extensively studied [10]. Some scholars have applied deep learning technology to the field of IHC image analysis and have achieved excellent results for the ROI extraction [11], [12], feature extraction [13], and automatic scoring tasks [14]. Some of the existing deep-learning-based methods evaluate single nuclei to score the entire image. Feng et al. [15] developed an automatic scoring model for IHC images of ER, PR, Ki-67, and HER-2. They selected the fully convolutional network as the backbone network for nucleus detection and DenseNet as the classification backbone network for scoring, in combination with the nucleus and cell membrane staining intensity. Saha et al. [16] proposed the HscoreNet for IHC scoring of breast cancer-related estrogen and progesterone proteins. The encoder and decoder of the HscoreNet were used for nucleus segmentation. The scoring layer calculates the IHC score based on staining intensity, color expression, and the number of positive and negative nuclei, achieving an accuracy of 94.53%. These methods require adding training labels to each nucleus. The image annotation is time and labor consuming.

Some scholars regard IHC image scoring as an image classification task. They used convolutional neural networks to extract full-image features and predict IHC scores. Khosravi et al. [17] predicted negative and positive categories for the IHC images of breast cancer using the InceptionV3 network and achieved 96% accuracy. Xue et al. [14] established an automatic classification model for the expression level of proteins in IHC images. They compared the classification performance of deep neural networks, random forests, and support vector machines. The highest classification accuracy is 73.72%. These methods can greatly reduce the amount of image annotation, but do not incorporate clinical prior experience. However, there are problems that the extracted features do not have clear meanings.

In summary, conventional image processing techniques require manually set quantitative metrics [18], [20] that are usually not well adapted. Some deep-learning-based methods evaluate protein expression from the nucleus level, which requires a lot of time for data annotation. Other methods directly predict scoring results by building image classification models. It can greatly reduce the amount of image annotations. However, there are also problems that the extracted features do not have clear meanings and the accuracy is not high.

Inspired by image fusion techniques [21] and heterogeneous pulse-coupled neural network (HPCNN) [22], we propose a two-step feature fusion method for automatic IHC scoring to solve the above problems. Our method consists of two steps. The first step is to extract the effective tissue area and the nucleus area as the region of interest (ROI). In an IHC image, not all areas are tissue. The periphery of the tissue is the background of the slide and impurities such as broken cells. Effective tissue regions need to be extracted from the original IHC images to exclude the interference of irrelevant regions. In addition, pathologists pay attention to the expression of P53 protein in the nucleus region when interpreting the IHC images of P53 protein. P53 protein expression in the cytoplasm needs to be excluded when interpreting. Therefore, the second step is to construct an IHC scoring model based on feature fusion. The main branch extracts the features of the effective tissue region images using DenseNet and FcaNet mechanism. The branch of nuclear region feature extraction extracts the color representation vector of the nuclear region as the staining intensity feature. Finally, the classifier is trained to synthetically discriminate the IHC scores corresponding to the fused feature vectors. The main contributions of this paper are as follows:

1. Propose a method for segmenting nuclei in the hematoxylin staining channel to improve the segmentation accuracy.

2. Propose an improved FCM segmentation algorithm combined with the region merging strategy to solve the problem that negatively stained nuclei are easily lost.

3. Propose an automatic scoring method for IHC images based on feature fusion network. Prior knowledge of pathologist scoring can be effectively exploited.

II. MATERIALS AND METHODS

A. IMAGE ACQUISITION

The IHC images were obtained from tissue microarrays produced by the Institute of Cancer Institute, the Second Affiliated Hospital, Zhejiang University School of Medicine.

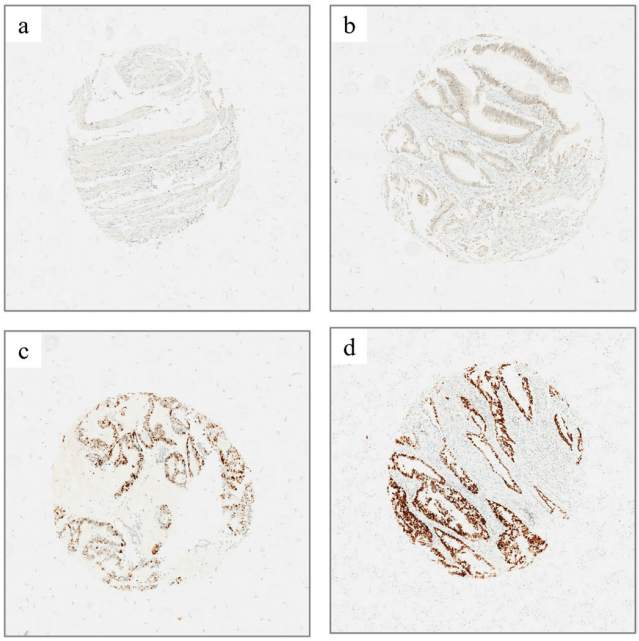


FIGURE 1. Original IHC images corresponding to different scores. (a) Negative. (b) Weakly positive. (c) Positive. (d) Strongly positive.

The tissues came from surgically resected specimens of colorectal cancer patients. The antibody is P53 (CST, item number 2527S) [23], [24]. The dataset contained 1168 whole-slide images (WSIs) with a resolution of 3000 × 3000 and were classified into four categories: negative (374), weakly positive (200), positive (117), and strongly positive (477), corresponding to a score of 0, 1, 2, and 3. The original images are shown in Fig. 1. Negative cell nuclei were stained blue by hematoxylin, and positive areas were stained brown at various degrees by DAB. The darker the color and the larger the positively stained area is, the higher the score should be.

B. FRAMEWORK

The method in this paper incorporates the scoring rules of pathologists. The procedures of the automatic IHC images scoring proposed in this paper is shown in Fig. 2. First, the ROI extraction part extracts the effective tissue regions to exclude the interference of irrelevant regions in recognition of the stained area. Meanwhile, it extracts the nucleus regions to exclude the interference from cytoplasmic staining, making the scoring network focus more on the nucleus regions. To address the problem of limited number of WSIs, which is prone to overfitting, data augmentation is used to enhance the number of data sets. Afterwards, the main branch of the feature fusion network uses a modified convolutional neural network to learn the features of the effective tissue region images. The feature extraction branch of the nuclear regions obtains the staining intensity feature vector. The feature vectors are stitched together as the input into the fully connected layers to calculate the IHC score.

C. EFFECTIVE TISSUE REGION EXTRACTION

We focus on the circular tissue area at the center of the images. In addition to the stained tissue sections in IHC images, there also exist meaningless regions, such as interstitial and slide backgrounds, which can influence the network scoring results. The process starts from graying out and enhancing the original image, followed by significance detection, and finally, the complete effective tissue region image is obtained by connected domain detection.

1) Image preprocessing: The color of the circular tissue region is very light and close to the background color of the slide, which leads to poor results using direct extraction. Therefore, the tissue region features need to be enhanced. Image graying can reduce the computational effort and increase the contrast visually at the same time. Then, the contrast at the dark regions in the image is enhanced through gamma transformation, which helps to better distinguish the image details at low gray levels. A gamma value of 0.5 is selected in our work.

2) Significance detection algorithm: The LC significance detection algorithm [25] is used to extract tissue regions from the contrast space. This method can increase the contrast at the tissue regions. The saliency mapping of the image is built on the grayscale contrast among the image pixels. The significance value is the sum of the distances between one pixel and all the other pixels in the graph in grayscale value.

3) Connected domain detection: In order to obtain the complete tissue region, it is necessary to remove the peripheral impurities and to fill the small areas missed among the tissues. The sizes of all the connected domains are detected by using the four-neighborhood connected domain detection algorithm. According to the set threshold, the small regions that are incorrectly detected are removed to obtain a more complete tissue region.

D. CELL NUCLEUS SEGMENTATION ALGORITHMS

In P53 IHC scoring, pathologists focus only on the degree of the expression of P53 protein in the nucleus. However, P53 protein is also expressed to some extent in the cytoplasm, resulting in some degree of positive staining in the cytoplasm and mesenchyme in the IHC images as well.

Various types of stains are used in IHC staining, among which the diaminobenzidine (DAB) stains P53 protein brown through antigen-antibody reaction, and the hematoxylin stains the nucleus blue. Therefore, we first obtain an image composed of the hematoxylin staining channel, the DAB staining channel, and the residual channel by color deconvolution, and then segment the nucleus region based on fuzzy C-mean clustering.

1) Color deconvolution: It is a common algorithm used for color separation in IHC images. The color deconvolution algorithm in [26] was used to transform the P53 IHC image from RGB space to stain space to obtain image that consists of the hematoxylin staining channel, the DAB staining channel, and the residual channel.

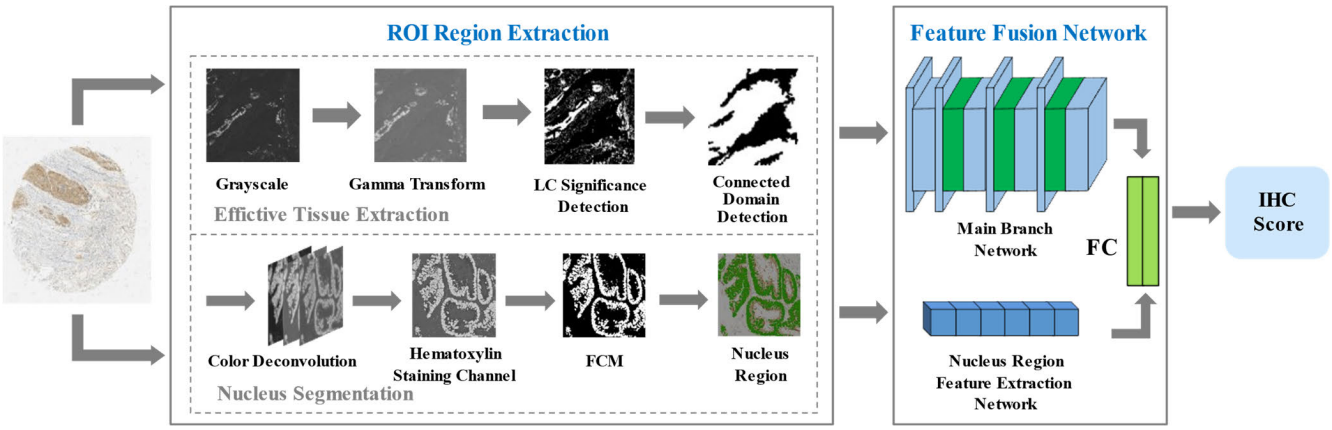


FIGURE 2. The structure of proposed IHC image scoring method.

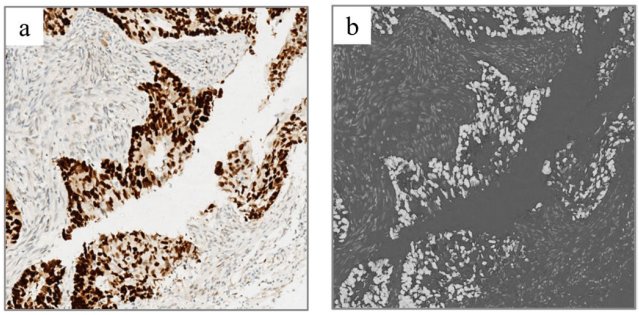


FIGURE 3. Hematoxylin channel image obtained by color deconvolution. (a) Original image. (b) Hematoxylin channel image with white color in the nucleus region.

2) Fuzzy C-mean clustering segmented cell nuclei region: Although the P53 protein expressed in positive cell nuclei is stained brown by DAB, hematoxylin staining is still present. As shown in Fig. 3, the pixels in the nucleus region of the hematoxylin channel have a high gray value and appear white. This feature can be used to segment the nucleus region in the hematoxylin channel. The above analysis shows that the pixels of the hematoxylin staining channel can be divided into 3 categories: positive nuclei, negative nuclei, and other pixels. Using the Fuzzy C-means (FCM) algorithm [27], the segmentation results of the nucleus region are obtained by combining the positive and negative nucleus regions. When positive nuclei are not present in the image, the following merging strategy is developed by observing the values of the centroids of each cluster to prevent unknown classification results leading to merging errors:

$$I = \begin{cases} 0, & I \in \{C_k \,|c_k - \min c_i <= 0.1\} \\ 1, & I \in \{C_k \,|c_k - \min c_i > 0.1\} \end{cases} \quad (1)$$

where C_i is the cluster center. Observing the value of each cluster center point, it is found that the value of the cluster center point of positive nuclei and negative cell nuclei is usually about 0.2 higher than that of other regions. Therefore, a threshold value of 0.1 is set, and the class with the value of the cluster centroid higher than the minimum value of 0.1 is divided into cell nuclei or is divided into other regions. The cell nucleus region segmentation image is successfully obtained by using this merging strategy.

E. P53Net

As mentioned above, we have extracted valid tissue region images and cell nucleus region images. Data enhancement is first performed, and then the improved feature fusion network P53Net makes full use of image features to obtain more accurate IHC scores.

1) Data enhancement: The number of full-field IHC images is limited, so the trained network model is prone to overfitting. In this paper, four data enhancement methods are employed, including vertical flip, horizontal flip, random pan, and random crop. Moreover, we propose a data enhancement method based on region swapping. Since the key feature is the color of the IHC image, swapping regions of the same size in the image can enrich the morphological features and suppress overfitting while preserving the color features. The specific steps of image region exchange are as follows: divide the image into 16 blocks of the same height and width, then randomly swap the positions of them to form a new image.

2) Main branch network: A FcaNet [28] module is inserted after each transition layer of DenseNet as the backbone structure for effective tissue region feature extraction. On the one hand, the score for P53 IHC images is not sensitive to high-level semantic features such as the shape and the contour of the images but is more concerned with low-level features such as pixel color and staining area. The densely connected structure of DenseNet enables the reuse of feature maps and optimizes the propagation path of low-level features, making it more suitable for the scoring task of P53 IHC images. The main branch uses the DenseNet169 network as the basic model, which contains four dense convolutional blocks and three transition layers. The four dense convolutional blocks contain 6, 12, 32, and 32 convolutional blocks, respectively; each convolutional block has the same structure. They all

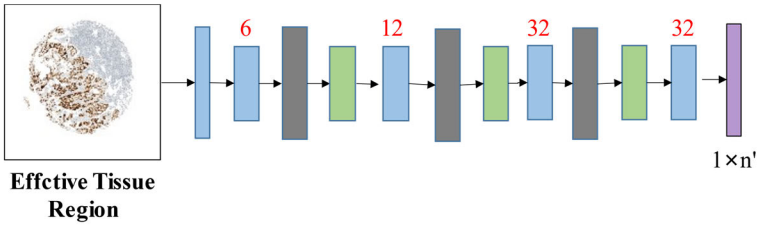


FIGURE 4. Backbone network for effective tissue region feature extraction. The convolutional layer is blue, and the dense convolutional block is shorter. The transition layer is gray. The FcaNet structure is green. The feature vector is purple.

consist of a 1 × 1 convolutional layer, a 3 × 3 convolutional layer, and a batch normalization layer. On the other hand, the number of feature map channels increases with the number of network layers. In order to enhance the key features and weaken the other features in the feature map, we insert a FcaNet module after each transition layer of DenseNet, so that the network can learn the weights of each channel. The performance of the whole network can be affected by the choice of the frequency component in the Discrete Cosine Transform (DCT). The analysis shows that the neural network is more concerned with the low-frequency information in the images [29]. So we introduce 32 low-frequency components to obtain the best performance based on the experimental comparisons.

The structure of the main branch network is shown in Fig. 4. A feature map with 1024 channels is obtained after going through all the convolutional, pooling, and batch normalization layers [30] of the network. Each channel of the feature map can be regarded as a feature of the input image.

3) Nuclear region feature extraction branch: The degree of positivity is directly related to the pixel color. Therefore, a vector can be used to record the distribution of image pixels in the nucleus region as a characterization of the image color distribution features. The first step is the grayscale value division. In order to compress the information and avoid the high dimensionality after vectorization, we divide the gray values of each channel into 26 degrees, starting from 0, for every 10 gray values. Then, we arrange the compressed R, G, and B values from high to low in 26 to map the points in the three-dimensional coordinate system to one-dimensional. The length of the constructed feature vector is 17576. Next, we count the number of pixels, so that the value of the corresponding position of the vector is equal to the number of pixels mapped to this position. Finally, the extracted vectors are normalized to the range 0 to 1 and sent into the feature extraction network. The network consists of an input layer containing 17576 neurons and a fully connected layer containing 1024 neurons. The nuclear region feature extraction branching network can learn the staining intensity features of the cell nuclear region and output a feature vector of a length of 1024. The branching structure is shown in Fig. 5.

4) Comprehensive scoring: The two feature vectors extracted above are stitched into one vector and fed into

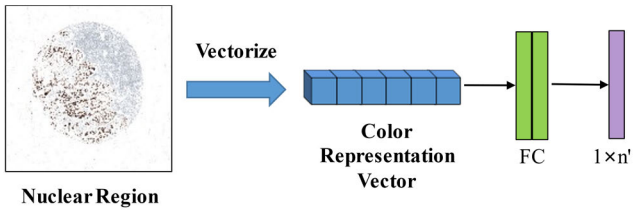


FIGURE 5. Extraction branch network of the nuclear region features.

two fully connected layers. The first fully-connected layer contains 1024 neurons, and the second output layer contains 4 neurons. The dropout is set to be 0.5 for each fully connected layer. The output layer uses the softmax activation function to normalize the output of each category to the range 0 to 1, indicating the confidence level. The category with the highest confidence level is the prediction result of the model. This feature fusion network structure extracts features from the images of effective tissue region while paying higher attention to the staining intensity features in the cell nucleus region.

III. RESULTS AND DISCUSSION

A. EFFECTIVE TISSUE REGION EXTRACTION RESULTS AND ANALYSIS

The three significance detection algorithms, RBD [31], SR [32], and FT [33] are selected to compare with the LC algorithm used in this study, and the results are shown in Fig. 6. It can be seen that although the tissue regions extracted by the SR algorithm are more complete, a very large number of incorrectly detected regions are generated. There are some tissue regions undetected when using RBD and FT algorithms. The tissue regions extracted by the LC algorithm are relatively complete and the number of false detection regions is less.

We compare the tissue region extraction algorithm used in this paper with the OTSU algorithm and the watershed algorithm, which are commonly used in medical image segmentation. The results are shown in Fig. 7. The cell nuclei with darker color have a larger grayscale difference from the tissue regions, as a result, the OTSU algorithm selects a larger threshold and classifies the tissue regions as background. The tissue regions extracted by the watershed algorithm are

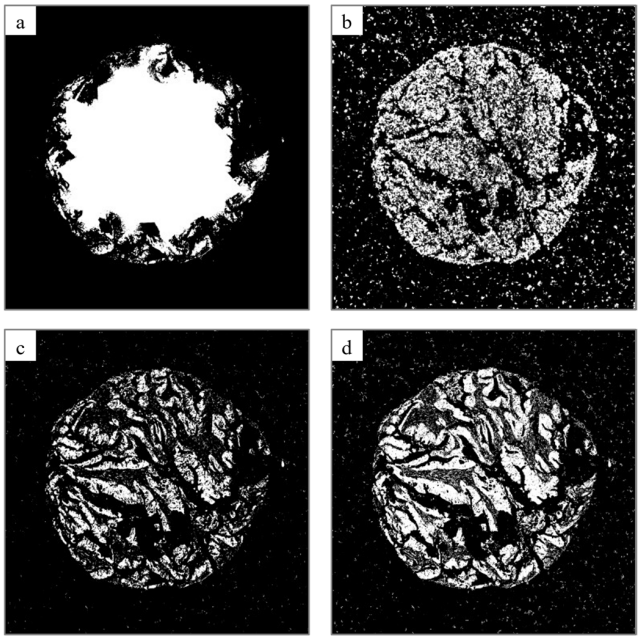


FIGURE 6. Comparison of the results of different algorithms for effective tissue region extraction. (a) RBD algorithm detection result. (b) SR algorithm detection result. (c) FT algorithm detection result. (d) detection results of LC algorithm used in this study.

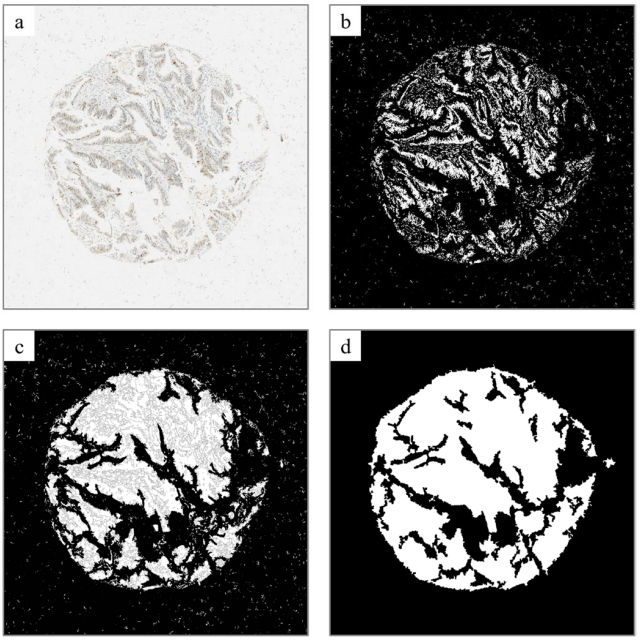


FIGURE 7. Comparison of the method in this study with the OTSU algorithm and the Watershed algorithm. (a) The original image. (b) The OTSU algorithm result. (c) The Watershed algorithm result. (d) Our result.

TABLE 1. Comparison of tissue region extraction results.

Method	DSC(%)	PA(%)
The OTSU	76.80	84.55
The Watershed	95.69	96.83
Proposed in this paper	98.23	98.79

partially missing and some regions in the middle are incorrectly detected due to over-segmentation. Compared with the above methods, the method used in this paper extracts the most complete tissue region and also identifies the vacant part in the middle of the tissue region better. To quantitatively compare the performance of segmentation, dice similarity coefficient (DSC) and pixel accuracy (PA) were used as evaluation metrics [34]. Comparison experiments were performed on 20 IHC images. Results are shown in Table 1. It can be seen that the proposed method in this paper can provide more accurate tissue region segmentation results than other methods.

B. CELL NUCLEUS SEGMENTATION RESULTS AND ANALYSIS

Fig. 8 shows the nucleus segmentation results. The nucleus regions in the original images that are white in the mask are kept, and the other regions are filled with a gray value of 255. It can be seen that the nucleus segmentation algorithm can acquire the nucleus region properly in the case where there is no cytoplasm stained brown. For the case with cytoplasm stained brown, the algorithm can extract only the nucleus regions, thus the influence of cytoplasm staining on the subsequent scoring calculation is excluded. It can be seen that the obtained nuclei are complete, and both the positive nuclei stained in brown and the negative nuclei stained in blue are correctly segmented.

The proposed nuclei segmentation method was compared with OTSU and FCM algorithm. Fig. 9 shows the comparison results of cell nucleus segmentation using three algorithms. The threshold calculated by OTSU algorithm was not suitable. As a result, some areas of negative nuclei with brighter colors were missed. When directly using FCM algorithm to segment the nuclei, some bright-colored negative nuclei will also be clustered into the background. The improved FCM algorithm clustered pixels into 3 categories. Positive nuclei were then merged with negative nuclei classes by formulating a class merging strategy. After adding the category merging strategy proposed in this paper, both positive and negative nuclei were completely segmented, and better nuclei segmentation performance was achieved.

C. P53Net SCORING RESULTS AND ANALYSIS

In this work, a NVIDIA RTX2060 graphics card was used to train the network and the Karas framework was used for programming. The data set is divided into training sets and test sets by a ratio of 2:1. Due to the high resolution of the original images, all images are uniformly reduced to 672×672 in order to reduce the computational effort without losing too much image detail. The idea of transform learning is used in the training process. The network uses the weights from the DenseNet169 model pre-trained on the ImageNet. The training process consists of two stages. We freeze the

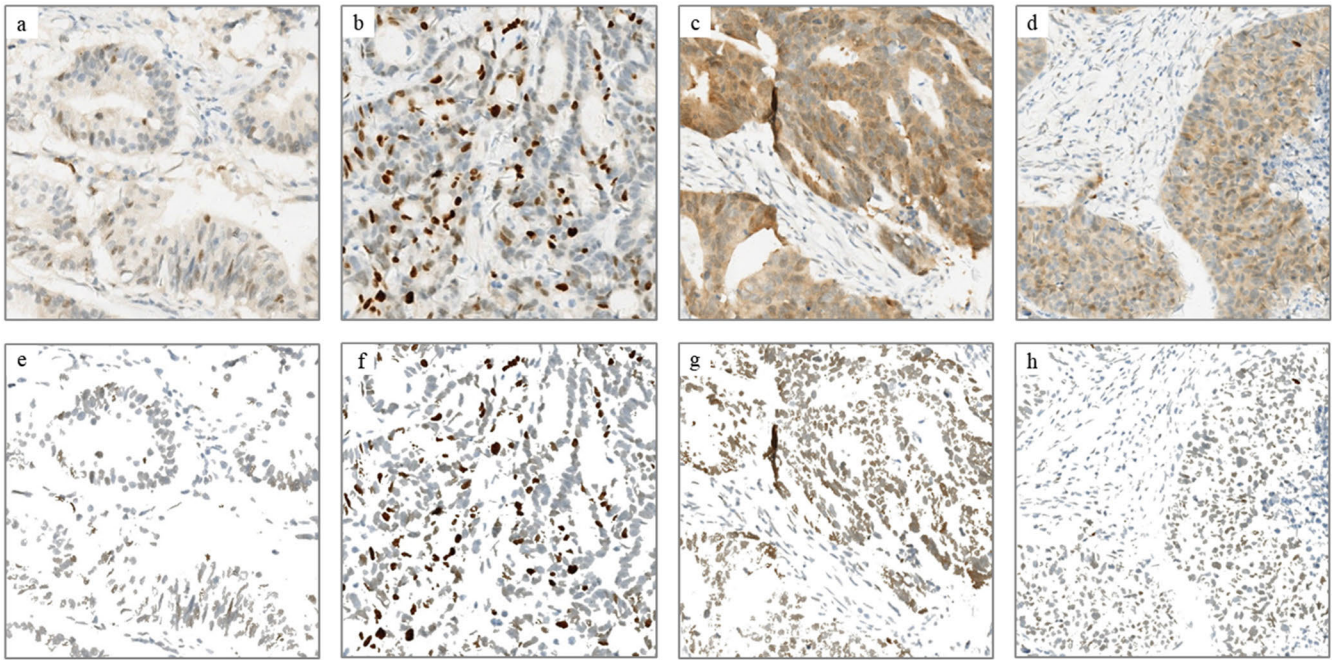


FIGURE 8. Cell Nucleus Segmentation Results. (a)(b)(c)(d) The original image. (e)(f)(g)(h) Segmentation of the proposed method.

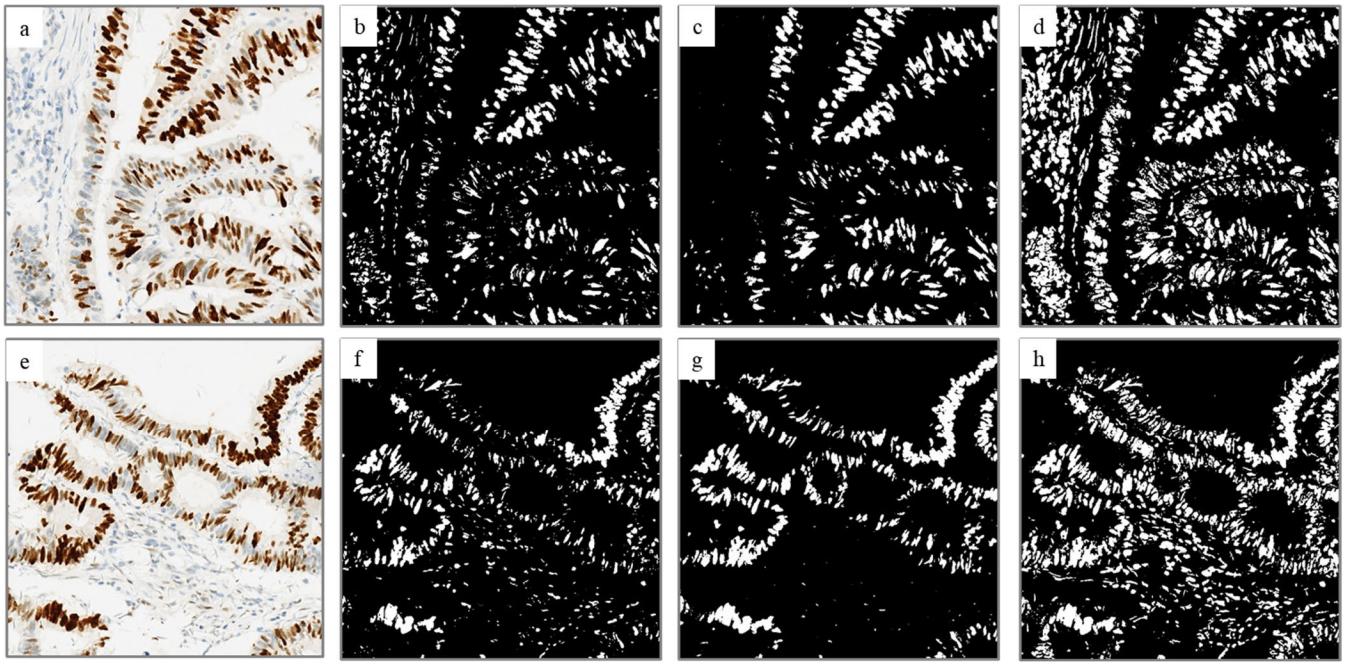


FIGURE 9. Comparison results of cell nucleus segmentation algorithms. (a)(e) Original image. (b)(f) Segmentation of OTSU algorithm (c)(g) Segmentation of unimproved FCM algorithm (d)(h) Segmentation of the proposed method.

convolutional layers of the pre-trained model and train only the newly added layers and the fully connected layers in the first stage, and all layers of the model are added to the training in the second stage. This two-stage training strategy takes use of the pre-trained weights to speed up the learning efficiency and shortens the time needed for training. The Adam optimizer is used in the training process, and the learning rate is decayed by cosine annealing. The initial learning rate is set to a large value and the termination learning rate is set to a small value. It decays in a descending cosine function. The accuracy cannot be improved after 30 epochs of iteration in the first stage, which is designed to determine a suitable initial weight for these layers. It is the second stage of training that determines the final performance of the network, so it is

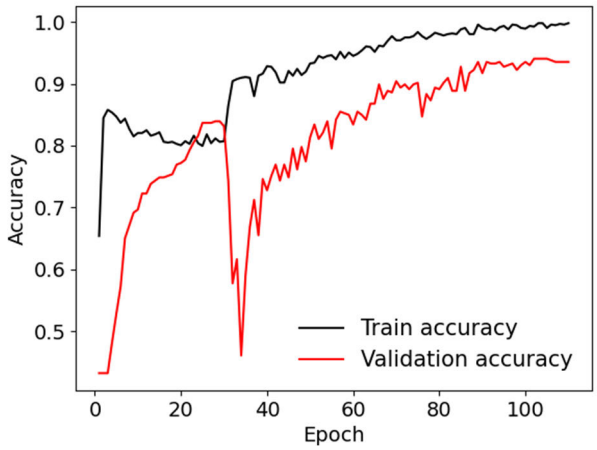


FIGURE 10. Model performance at different iteration epochs.

TABLE 2. Results of ablation experiment I.

Ablation part	Test set Accuracy (%)
Region exchange method	93.35
FcaNet	93.52
Nothing	94.21

TABLE 3. Results of ablation experiment II.

Ablation part	Test set Accuracy (%)
The main-branch	90.41
The nucleus feature extraction branch	93.43
Nothing	94.21

necessary to design a comparison experiment of the network performance under different iterations. As shown in Fig. 10, the accuracy rate tends to be stable when the number of iterations reaches 80 in the second stage.

We have designed some ablation experiments based on the above study to demonstrate the necessity of the proposed method. The first ablation experiment is used to verify the effectiveness of the region exchange data enhancement method and the FcaNet module. The second experiment ablates two network branches separately to verify the necessity of each branch. The average accuracy on the test set in three training sessions is taken as the result.

It can be seen in Table 2 that the accuracy on the test set decreases if we ablate either the region exchange data enhancement or the FcaNet module. The result proves that both the proposed data augmentation method for region swapping and the FcaNet module used in the network structure are indeed helpful to improve the network performance.

As shown in Table 3, when the main branch network is ablated, the accuracy of the model on the test set drops a lot. This is because the main branch network is used to extract features from the effective tissue region images, which contain more comprehensive feature information. In addition, the nuclear region feature extraction branch is used to extract the staining intensity features of the nucleus region, which is a supplement to the former. Using the nuclear region feature extraction branch alone causes loss of other information and cannot obtain features related to staining extent, tissue region area, and other positive staining percentage indicators. There is also a decrease in accuracy on the test set after ablating the nuclear region feature extraction branch. This is because the nuclear region feature extraction branch provides a more complete set of features of the staining intensity in the nuclear region of the cell, so ablating this branch would reduce the effective information in the feature vector. The ablation experiments demonstrate that both the main branch and the nuclear region feature extraction branch contribute to the improvement of network performance. Combining the feature vectors extracted from the two branches through the structure of the feature fusion network can increase the information dimension and improve the model performance.

TABLE 4. Performance comparison of different methods on The test set.

Method	Accuracy (%)	F1-Score (%)	Kappa
VGG16	89.38	81.57	0.8439
InceptionV3	90.50	84.02	0.8552
ResNet-50	91.36	84.80	0.8699
DenseNet-169	91.71	85.79	0.8746
EfficientNet-b4	92.49	86.65	0.8838
P53Net	94.21	89.24	0.9134

The above two sets of ablation experiments demonstrate the effectiveness of the method proposed in this study. Next, we compare the method with five generic methods in the field of general image recognition to verify the effect of this method. These 5 methods are VGG16, InceptionV3, ResNet-50, DenseNet-169 and EfficientNet-b4. Accuracy, F1-Score, and Kappa coefficient are chosen as the evaluation indices. The parameters and optimizers are kept the same during the training process. The average of the results of three experiments is also taken for reducing the random factors such as dataset division and initial weights.

The experimental results are shown in Table 4. It can be seen that the generic image recognition methods also give good results. However, the IHC scoring task has its special features, which depend on the staining intensity and staining percentage and need to focus on the positive expression location. Therefore, it is difficult to learn the correct features with the limited amount of WSIs. In this study, we use the feature fusion method and introduce the FcaNet mechanism to improve the performance of the network to better perform the task of P53 IHC image scoring for colorectal cancer.

The challenges in implementation mainly came from how to effectively use the prior knowledge of pathologists and the large GPU memory required for model training. On the one hand, we analyzed the scoring process of pathologists.

Features were extracted separately from tissue regions and nuclei that are of interest to pathologists. Image texture features and color features were fused to improve the scoring accuracy of IHC images. On the other hand, to reduce the GPU memory usage of model training, we optimized the kernel region feature extraction branch. We reduced the size of the color feature vectors by setting a grayscale threshold. In addition. We also adjusted the hyperparameters such as the image resolution of the input model and the batch size used in the training process through experiments.

The algorithm in this paper outputs the score with the highest confidence level, showing good results on images of all scoring levels. The confidence level of the prediction results for the negative and strong positive images is high. Most of them are above 98%. A small number of weakly positive and positive images have low confidence in the prediction results because their features are not distinct enough and can be easily confused with the images of adjacent scoring levels. The low confidence level in this case suggests that the algorithm may give incorrect score. In practical applications, appropriate thresholds can be set to filter out images with low confidence for manual interpretation. Such combination of algorithmic scoring and manual verification approach can greatly reduce the required time for pathologists to read the images while ensuring the accuracy of the scoring results.

IV. CONCLUSION

With the development of computer technology and whole slide digital pathological section technology, computer-assisted IHC interpretation methods have shown great application prospects in improving the consistency of IHC scoring results and shortening the interpretation time. In this paper, we propose an automatic scoring method for tumor IHC images based on deep learning. The tissue regions and the cell nuclei regions are extracted as the ROI. The image features of tissue regions and the staining intensity features of cell nuclei are learned through a feature fusion network. The features are then combined to give the IHC score. We perform experiments on IHC images of P53 protein and verify the efficiency of this study. The ablation experiments show that the accuracy of scoring can be further improved by fusing the color features of the cell nucleus region, compared with the image features extracted by convolutional neural network alone. Our method can obtain accurate scores without extensive manual labels, further demonstrating the potential of artificial intelligence for automated IHC scoring. For the problem that weakly positive and positive images are occasionally confused, we propose that images with confidence levels below a threshold can be filtered out for manual scoring. This work can greatly save the time of pathologists on reading images while ensure the accuracy of the scoring results.

ACKNOWLEDGMENT

(Jinbo Zhang and Xiaolei Guo contributed equally to this work.)

REFERENCES

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA, Cancer J. Clinicians*, vol. 71, no. 3, pp. 209–249, Feb. 2021, doi: 10.3322/caac.21660.

[2] M. S. Hossain, H. Karuniawati, A. A. Jairoun, Z. Urbi, D. J. Ooi, A. John, Y. C. Lim, K. M. K. Kibria, A. K. M. Mohiuddin, L. C. Ming, K. W. Goh, and M. A. Hadi, "Colorectal cancer: A review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies," *Cancers*, vol. 14, no. 7, p. 1732, Mar. 2022, doi: 10.3390/cancers14071732.

[3] I. Bozic, J. G. Reiter, B. Allen, T. Antal, K. Chatterjee, P. Shah, Y. S. Moon, A. Yaqubie, N. Kelly, D. T. Le, E. J. Lipson, P. B. Chapman, L. A. Diaz, B. Vogelstein, and M. A. Nowak, "Evolutionary dynamics of cancer in response to targeted combination therapy," *eLife*, vol. 2, Jun. 2013, Art. no. e00747, doi: 10.7554/eLife.00747.

[4] M. Cregger, A. J. Berger, and D. L. Rimm, "Immunohistochemistry and quantitative analysis of protein expression," *Arch. Pathol. Lab. Med.*, vol. 130, no. 7, pp. 1026–1030, Jul. 2006, doi: 10.5858/2006-130-1026-IAQAOP.

[5] F. D. C. Bernardi, M. D. C. Bernardi, T. Takagaki, S. A. C. Siqueira, and M. Dolhnikoff, "Lung cancer biopsy: Can diagnosis be changed after immunohistochemistry when the based morphology corresponds to a specific tumor subtype?" *Clinics*, vol. 73, p. e361, 2018, doi: 10.6061/clinics/2018/e361.

[6] M. Braun, R. Kirsten, N. J. Rupp, H. Moch, F. Fend, N. Wernert, G. Kristiansen, and S. Perner, "Quantification of protein expression in cells and cellular subcompartments on immunohistochemical sections using a computer supported image analysis system," *Histol. Histopathol.*, vol. 28, no. 5, pp. 605–610, May 2013, doi: 10.14670/HH-28.605.

[7] G. Stålhammar, N. F. Martinez, M. Lippert, N. P. Tobin, I. Mølholm, L. Kis, G. Rosin, M. Rantalainen, L. Pedersen, J. Bergh, M. Grunkin, and J. Hartman, "Digital image analysis outperforms manual biomarker assessment in breast cancer," *Modern Pathol.*, vol. 29, no. 4, pp. 318–329, Apr. 2016, doi: 10.1038/modpathol.2016.34.

[8] J. Konsti, M. Lundin, H. Joensuu, T. Lehtimäki, H. Sihto, K. Holli, T. Turpeenniemi-Hujanen, V. Kataja, L. Sailas, J. Isola, and J. Lundin, "Development and evaluation of a virtual microscopy application for automated assessment of Ki-67 expression in breast cancer," *BMC Clin. Pathol.*, vol. 11, no. 1, p. 3, Jan. 2011, doi: 10.1186/1472-6890-11-3.

[9] R. Røge, R. Riber-Hansen, S. Nielsen, and M. Vyberg, "Proliferation assessment in breast carcinomas using digital image analysis based on virtual Ki67/cytokeratin double staining," *Breast Cancer Res. Treatment*, vol. 158, no. 1, pp. 11–19, Jun. 2016, doi: 10.1007/s10549-016-3852-6.

[10] Z. Lv, Z. Yu, S. Xie, and A. Alamri, "Deep learning-based smart predictive evaluation for interactive multimedia-enabled smart healthcare," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 18, no. 1s, pp. 1–20, Jan. 2022, doi: 10.1145/3468506.

[11] L. B. Mahanta, E. Hussain, N. Das, L. Kakoti, and M. Chowdhury, "IHC-Net: A fully convolutional neural network for automated nuclear segmentation and ensemble classification for allred scoring in breast pathology," *Appl. Soft Comput.*, vol. 103, May 2021, Art. no. 107136, doi: 10.1016/j.asoc.2021.107136.

[12] P. Ghahremani, Y. Li, A. Kaufman, R. Vanguri, N. Greenwald, M. Angelo, T. J. Hollmann, and S. Nadeem, "Deep learning-inferred multiplex immunofluorescence for immunohistochemical image quantification," *Nature Mach. Intell.*, vol. 4, no. 4, pp. 401–412, Apr. 2022, doi: 10.1038/s42256-022-00471-x.

[13] P. Huang, X. Tan, X. Zhou, S. Liu, F. Mercaldo, and A. Santone, "FABNet: Fusion attention block and transfer learning for laryngeal cancer tumor grading in P63 IHC histopathology images," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1696–1707, Apr. 2022, doi: 10.1109/JBHI.2021.3108999.

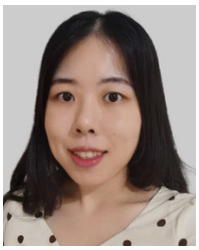
[14] Z.-Z. Xue, C. Li, Z.-M. Luo, S.-S. Wang, and Y.-Y. Xu, "Automated classification of protein expression levels in immunohistochemistry images to improve the detection of cancer biomarkers," *BMC Bioinf.*, vol. 23, no. 1, pp. 1–13, Nov. 2022, doi: 10.1186/s12859-022-05015-z.

[15] M. Feng, J. Chen, X. Xiang, Y. Deng, Y. Zhou, Z. Zhang, Z. Zheng, J. Bao, and H. Bu, "An advanced automated image analysis model for scoring of ER, PR, HER-2 and Ki-67 in breast carcinoma," *IEEE Access*, vol. 9, pp. 108441–108451, 2021, doi: 10.1109/ACCESS.2020.3011294.

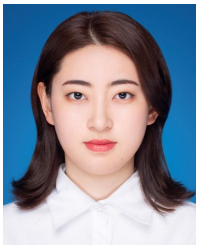
[16] M. Saha, I. Arun, R. Ahmed, S. Chatterjee, and C. Chakraborty, "*HscoreNet*: A deep network for estrogen and progesterone scoring using breast IHC images," *Pattern Recognit.*, vol. 102, Jun. 2020, Art. no. 107200, doi: 10.1016/j.patcog.2020.107200.
[17] P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, and I. Hajirasouliha, "Deep convolutional neural networks enable discrimination of heterogeneous digital pathology images," *EBioMedicine*, vol. 27, pp. 317–328, Jan. 2018, doi: 10.1016/j.ebiom.2017.12.026.
[18] F. Varghese, A. B. Bukhari, R. Malhotra, and A. De, "IHC profiler: An open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples," *PLoS ONE*, vol. 9, no. 5, May 2014, Art. no. e96801, doi: 10.1371/journal.pone.0096801.
[19] V. Tollemar, N. Tudzarovski, E. Boberg, A. Törnqvist Andrén, A. Al-Adili, K. Le Blanc, K. G. Legert, M. Bottai, G. Warfvinge, and R. V. Sugars, "Quantitative chromogenic immunohistochemical image analysis in cellprofiler software," *Cytometry A*, vol. 93, no. 10, pp. 1051–1059, Oct. 2018, doi: 10.1002/cyto.a.23575.
[20] S. Ram, P. Vizcarra, P. Whalen, S. Deng, C. L. Painter, A. Jackson-Fisher, S. Pirie-Shepherd, X. Xia, and E. L. Powell, "Pixelwise H-score: A novel digital image analysis-based metric to quantify membrane biomarker expression from immunohistochemistry images," *PLoS ONE*, vol. 16, no. 9, Sep. 2021, Art. no. e0245638, doi: 10.1371/journal.pone.0245638.
[21] X. Qin, Y. Ban, P. Wu, B. Yang, S. Liu, L. Yin, M. Liu, and W. Zheng, "Improved image fusion method based on sparse decomposition," *Electronics*, vol. 11, no. 15, p. 2321, Jul. 2022, doi: 10.3390/electronics11152321.
[22] H. Liu, M. Liu, D. Li, W. Zheng, L. Yin, and R. Wang, "Recent advances in pulse-coupled neural networks with applications in image processing," *Electronics*, vol. 11, no. 20, p. 3264, Oct. 2022, doi: 10.3390/electronics11203264.
[23] D. Huang, W. Sun, Y. Zhou, P. Li, F. Chen, H. Chen, D. Xia, E. Xu, M. Lai, Y. Wu, and H. Zhang, "Mutations of key driver genes in colorectal cancer progression and metastasis," *Cancer Metastasis Rev.*, vol. 37, no. 1, pp. 173–187, Mar. 2018, doi: 10.1007/s10555-017-9726-5.
[24] L. J. Hernández Borrero and W. S. El-Deiry, "Tumor suppressor P53: Biology, signaling pathways, and therapeutic targeting," *Biochimica Biophysica Acta-Rev. Cancer*, vol. 1876, no. 1, Aug. 2021, Art. no. 188556, doi: 10.1016/j.bbcan.2021.188556.
[25] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th ACM Int. Conf. Multimedia*, Oct. 2006, pp. 815–824.
[26] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quant. Cytol. Histol.*, vol. 23, no. 4, pp. 291–299, Aug. 2001.
[27] J. C. Bezdek, W. Full, and R. Ehrlich, "FCM: The fuzzy *c*-means clustering algorithm," *Comput. Geosci.*, vol. 10, nos. 2–3, pp. 191–203, 1984, doi: 10.1016/0098-3004(84)90020-7.
[28] Z. Qin, P. Zhang, F. Wu, and X. Li, "FcaNet: Frequency channel attention networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 763–772, doi: 10.1109/ICCV48922.2021.00082.
[29] K. Xu, M. Qin, F. Sun, Y. Wang, Y.-K. Chen, and F. Ren, "Learning in the frequency domain," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1740–1749.
[30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.
[31] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2814–2821, doi: 10.1109/CVPR.2014.360.
[32] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8, doi: 10.1109/CVPR.2007.383267.
[33] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604, doi: 10.1109/CVPR.2009.5206596.
[34] S. F. Qadri, L. Shen, M. Ahmad, S. Qadri, S. S. Zareen, and S. Khan, "OP-convNet: A patch classification-based framework for CT vertebrae segmentation," *IEEE Access*, vol. 9, pp. 158227–158240, 2021, doi: 10.1109/ACCESS.2021.3131216.

JINBO ZHANG received the B.S. degree in automation from Zhejiang University, Hangzhou, Zhejiang, in 2020, where he is currently pursuing the master's degree majoring in control science and engineering. His research interests include image processing and machine vision.

XIAOLEI GUO received the B.S. degree in automation from Hunan University, Changsha, Hunan, in 2022. He is currently pursuing the master's degree majoring in control science and engineering with Zhejiang University, Hangzhou, Zhejiang. His research interests include image processing and machine vision.



WEN CAI received the Medical Doctor degree in clinical medicine (gastroenterology) from Zhejiang University, in 2022. She is currently with The Second Affiliated Hospital, Zhejiang University, as an Oncologist, specializing in the gastrointestinal carcinomas and neuroendocrine neoplasm. She is a young member of the Chinese Society of Clinical Oncology and China Anti-Cancer Association. At present, more than ten articles were published as the first author or co-first author.



YUQI CAO received the Ph.D. degree from the College of Control Science and Engineering, Zhejiang University, Hangzhou, China, in 2022. She is currently an Associate Research Fellow with the College of Control Science and Engineering, Zhejiang University. She has authored or coauthored 13 journal articles and eight conference papers. Her research interests include advanced transducer and measurement, medical image analysis, and THz-based human cancer detection.

WEITING GE received the Ph.D. degree in bioinformatics from Zhejiang University, in 2006. He is currently with The Second Affiliated Hospital, Zhejiang University, specializing in the colorectal cancer research and cancer bioinformatics. He is a member of the China Medical Biotech Association. He has published over 20 journal articles and authorized five invention patents.

PINGJIE HUANG is currently a Professor with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University. His research interests include advanced transducer and measurement, industrial and environmental information processing and event detection, and computer control system design and development. He is a Key Member of Zhejiang University, research group, for tissue, base on THz-TDS, early warning techniques and systems for urban water quality assurance, NDT&E of conductive structures, food products, water samples, UV-Vis, and ECT methods. In recent years, he has authored/coauthored more than 40 peer-reviewed papers and conference papers. In September 2017, he was supported by CSC. He was once a Visiting Scholar with the Prof. Xi-Cheng Zhang's Research Group, The Institute of Optics, University of Rochester, working on terahertz science and technology.

DIBO HOU received the B.E. degree in industrial automation and the Ph.D. degree in control science and engineering from Zhejiang University, China, in 1999 and 2004, respectively. He has been the Associate Director of the Institute of Smart Sensing and Measurement, Zhejiang University, since 2014. He is currently a Full Professor and the Deputy Dean of the College of Control Science and Engineering, Zhejiang University. He is also an Affiliate Member of the State Key Laboratory of Industrial Control Technology and a member of the National Technical Committee on Industrial Process Measurement and Control of Standardization Administration of China. He was selected as the "New Century 151 Talent" of Zhejiang Province, in 2015. His research interests include smart sensing and measurement, environmental monitoring and early-warning, non-destructive testing, and intelligent information processing. He is the author or coauthor of over 80 articles in refereed journals. He has served as the principal investigator of over 20 research grants from the National Science Foundation of China, the National Science and Technology Major Project of China, the Key Technology Research and Development Program of Zhejiang Province, and other agencies. He was awarded the first prize of Scientific and Technological Award of Zhejiang Province, the China Industry-University-Research Cooperation Innovation Award, and the Science and Technology Award of China Instrumentation Society.

SHU ZHENG is currently the Director of the Academic Committee, Cancer Institute, Zhejiang University. She has participated and afforded the sixth, seventh, and eighth Five-Year National Medical Strategic Science and Technology Plan. She is the author or coauthor of more than 400 original articles and book chapters and an editor or a co-editor of several text books. Her research interests include the surgical and adjuvant therapy of breast and colorectal cancers, and tumor diagnosis with protein fingerprint by SELDI-TOF-MS. She is a member of numerous national and international scientific societies, a member of numerous boards of scientific journals, including *The Chinese-German Journal of Clinical Oncology*, and advisory boards of research funding agencies. For her accomplishments, she received numerous awards and was bestowed with many honors, including the honorary surgical professorship of The Chinese University of Hong Kong.

GUANGXIN ZHANG received the B.E. degree and the Ph.D. degree in industrial automation from Zhejiang University, in 1991 and 2002, respectively. He is currently a Professor with the State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University. He is the Executive Director of the China Association of Automation and the Deputy Director of the Education Working Committee, China Association of Automation. His research interests include non-destructive testing, terahertz science and technology, water quality early warning technology, and computer control systems.