RESEARCH ARTICLE

Automatic recognition and classification of tumor cells in medical images based on convolutional neural networks

Yanli Tan[1, 2, *], Yongqiang Zhao[1], Jin Wang[1, 2], Guoqin Li[1, 2]

[1]Department of Electronic Engineering, Taiyuan Institute of Technology, Taiyuan, Shanxi, China. [2]Faculty of Computer and Mathematical Sciences, University Teknologi MARA, Selangor Shah Alam, Malaysia.

Received: July 10, 2024; **accepted:** November 2, 2024.

Tumors are one of the main causes of global mortality, and early diagnosis and precise treatment affect the survival rate of patients. Traditional image analysis relies on the professional knowledge and experience of radiologists. Manually labeling and analyzing tumor cells is laborious and subjective, which may lead to variability and inconsistency in diagnostic results. This study proposed a convolutional model based on a single-stage object detector. To address the shortcomings of the single-stage object detector model, the receptive field was introduced to improve the depth of information extraction, and an attention mechanism was also introduced to enhance its focus. The model was used to recognize pathological tissue slice images and classify tumors. The results indicated that, when the dataset size reached around 200, the performance of each model reached its maximum and showed convergence. When the dataset size was 500, the accuracy values of the single-stage target detector model, the single-stage target detector model based on receptive field, and the proposed model were 0.89, 0.91, and 0.98 with root mean square errors of 0.15, 0.14, and 0.09, respectively. Among the three models, the recognition accuracies of the proposed model for breast cancer, lung cancer, brain cancer, kidney cancer, head and neck cancer, and liver cancer were 92.3%, 93.9%, 88.8%, 82.2%, 93.1%, and 83.2%, respectively. The average recognition time for tumors was within 3 seconds. The proposed model demonstrated outstanding performance across all types of tumors, thereby providing an effective solution for medical image recognition.

Keywords: tumor cells; convolutional neural networks; image recognition; single shot multi-box detector; attention mechanism.

[*]**Corresponding author:** Yanli Tan, Department of Electronic Engineering, Taiyuan Institute of Technology, Taiyuan 030008, Shanxi, China. Email: whytyl0310@126.com.

Introduction

With the rapid development of medical imaging technology, medical images are playing an increasingly important role in disease diagnosis and treatment. In particular, the accurate imaging analysis of tumor cells is of great importance for the early diagnosis of cancer and the subsequent development of effective treatment plans that will facilitate the patient's recovery. However, traditional manual analysis methods are not only time-consuming and labor-intensive but also susceptible to subjective factors, resulting in insufficient accuracy and consistency of diagnostic results [1, 2]. In recent years, Convolutional neural network (CNN) has made significant progress in computer vision. With its powerful feature extraction and pattern recognition capabilities, CNN has shown great potential in medical image analysis. By training on a large amount of medical image data, CNN can automatically learn and extract complex

features, thereby achieving efficient recognition and accurate classification of tumor cells.

Yilmaz found that the traditional synthetic variable ratio image fusion algorithm (SVRIFA) could not meet the requirements of image fusion. In response to this issue, researchers proposed genetic algorithms to improve traditional SVRIFA. The model demonstrated robust performance in both single-sensor and multi-sensor images, exhibiting consistent accuracy and reliability across different imaging modalities [3]. You *et al*. proposed a multi-focus image fusion method, which combined with the local standard deviation of the Laplacian image corresponding to the source image and further enhanced through a guided filter to fuse the image [4]. Long *et al*. found that the existing method for obtaining high-resolution images was to fuse low-resolution hyperspectral images with high-resolution multi-spectral images. However, this method required decomposing the target image into a multi-factor matrix or tensor before fusion and was not suitable for real-time scenes. In response to this issue, a new non-factorization model was proposed, which could directly estimate the resolution image of the target image, reduce the analysis process of intermediate variables, effectively save computational time, and improve the stability of fusion [5]. CNN is a popular technology in computer science. Zhang *et al*. found that road conditions were essential for traffic flow efficiency and proposed an automatic classification method for winter road conditions using CNN to ensure the safety of traffic operations. This method could effectively classify different road surfaces and had good performance [6]. Miao *et al*. found it difficult to extract fault features from multiple heterogeneous sensor data and therefore proposed a feature-enhanced channel CNN for robot fault diagnosis. It also introduced a feature enhancement layer that highlighted important features by adaptively weighting feature maps (Fmap). This method exhibited excellent accuracy and robustness in fault diagnosis [7]. Xie *et al*. found that the detection performance of LiDAR points could be severely affected when there was noise or damage. In response to this, a 3D object detection approach that fuses point-cloud with 2D semantic segmentation was proposed to enhance feature representation in difficult situations. The average accuracy of this model in 3D object detection had significantly improved by about 11.13% [8]. On pathological tissue slice images, the target of tumor cells is relatively small and difficult to detect. Small target detection (STD) that refers to detecting and locating smaller targets in an image can precisely address this problem. Compared to large object detection, STD presents greater challenges, mainly due to the smaller area occupied by small targets in the image, less pixel information, and the susceptibility to noise and background interference [9]. In addition, small targets typically have lower contrast and resolution in the image, making the detection process more complex.

Many scholars have studied CNN and image recognition. However, these studies have adopted a single algorithm model and have not made improvements to address the shortcomings of the model. This study innovatively proposed a convolutional model based on single shot multi-box detector (SSD) to identify the tumor types confirmed by analyzing pathological tissue slice images. In response to the shortcomings of the SSD, a receptive field block (RFB) and an attention mechanism (AM) were introduced to improve their performance, aiming to provide a reliable tumor recognition and detection method. This research would provide an effective solution for medical image recognition, thereby promoting the development of intelligent healthcare and improving the accuracy and efficiency of tumor diagnosis.

Materials and methods

Automatic recognition model for tumor cells based on RFB

This study used SSD as the basic model. In the feature extraction network of the SSD model, the

shallow convolutional layer mainly extracted the detailed information of the image, while the deep layer mainly extracted abstract information. The SSD algorithm utilized feature information at different scales to combine detailed information with abstract information [10]. The SSD is based on the VGG16 visual geometry group network (VGG Net) as a whole. VGG Net is a deep CNN architecture that has a concise and deep structure, smaller convolutional kernels, and a deep network structure. The two common variants of VGG Net are VGG16 and VGG19, which contain 16 and 19 convolutional layers. The additional layers in VGG19 introduce more parameters, making the model larger and computationally demanding. In medical imaging, datasets are usually very large and involve high-resolution images, so training more complex models like VGG19 requires significantly more computing resources and time. Therefore, this study selected VGG16, which consisted of 16 convolutional layers and fully connected layers (FCLs) (Figure 1). The input size of VGG16 was shown in equation (1).

$$P = H_{in} \times W_{in} \times D_{in} \tag{1}$$

where H_{in} , W_{in} , and D_{in} were the height, width, and depth of the input image, respectively. P was the input size.

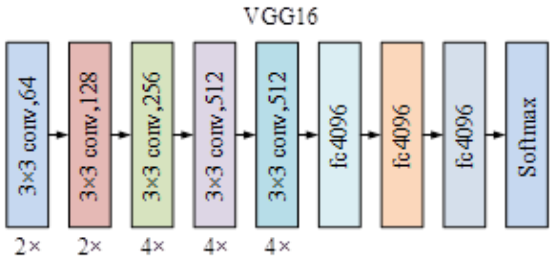


Figure 1. VGG16 network model diagram.

The output height of the convolution operation of the convolutional layer was given by equation (2).

$$H_{out} = \frac{H_{in} - F + 2P}{S} + 1 \tag{2}$$

where H_{out} was the output image's height. F was the convolution kernel's size. P was padding. S was the convolution's stride. The output width of the convolution operation was shown in equation (3).

$$W_{out} = \frac{W_{in} - F + 2P}{S} + 1 \tag{3}$$

where W_{out} was the width of the output image [11]. The pooling operation of the pooling layer was as follows.

$$\begin{cases} H_{out} = \dfrac{H_{in} - D}{H} + 1 \\ W_{out} = \dfrac{W_{in} - D}{H} + 1 \end{cases} \tag{4}$$

where D was the pooling window's size. H was the stride of pooling. The SSD algorithm sets prior boxes with different scales for different Fmaps, allowing each Fmap to focus on detecting targets at its specific scale. However, neural networks mainly detect large targets and ignore small objects, resulting in a lower success rate for detecting small objects [12]. In addition, due to the low-resolution of small targets, the image is relatively blurry and carries too little information, resulting in weak representation ability. Therefore, an RFB module based on multi-scale feature fusion was introduced to enhance the detection capability of the model. The proposed improved RFB module was mainly divided into a multi-branch convolution module and a dilated convolution module with the multi-branch convolution module expanding the RFB and enhancing the network's ability to capture features of diverse scales by introducing multiple parallel convolution branches, each using convolution kernels of distinctive sizes and shapes. By simultaneously extracting features through multiple convolutional branches, the

width of the network was increased, and each branch used different convolution kernels to get features of various scales that could be obtained by fusing the features of multiple branches through concatenation or weighted fusion. The dilated convolution module expanded the RFB of convolution operations by introducing holes in the convolution kernel, thereby increasing the RFB size without increasing the parameters and computational complexity [13]. Compared to regular convolution, dilated convolution could increase RFB while maintaining the same number of parameters, generating higher-resolution Fmaps. By adding different convolutional layers of RFB through the RFB module, the feature extraction network could obtain more contextual information (Figure 2).

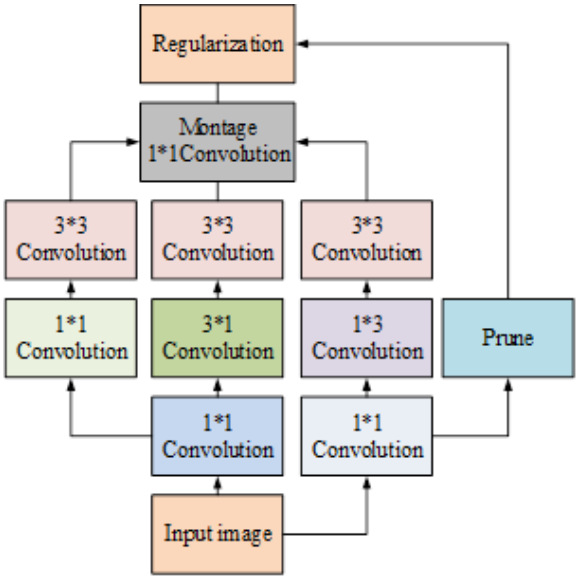


Figure 2. Network structure of RFB module.

The loss function selection of the model added the position error and confidence error, which matched the prior box with the annotation box. When the intersection-to-union ratio between the two was greater than or equal to 0.5, it could be determined that they were matched and expressed as follows.

$$L(x,c,l,g)=\frac{1}{N}(L_{conf}(x,c)+\alpha L_{loe}(x,l,g)) \quad (5)$$

where L_{conf} and L_{loc} were the error loss of confidence and position. N was the number of prior boxes. c and x were real and predicted category labels. l and g were the predicted and actual bounding box positions. α was the weight factor of position error loss and confidence error loss, usually defaulted to 1. The expression for position loss was given below.

$$L(x,l,g)=\sum_{i\in pos}^{N}\sum_{m\in\{cx,cy,w,h\}}x_{i,j}^{p}Smooth_{L1}(l_i^m-\hat{g}_j^m) \quad (6)$$

where $x_{i,j}^{p}$ was whether the i-th predicted box matched the j-th true box concerning category k. pos was all positive samples. m was the overall range of the default box. l_i^m was the prediction box. $\hat{g}_j^m$ was the position of the real box. The purpose of using a loss function was to correct the significant discrepancy between the predicted and the true boxes. The confidence loss part was calculated in equation (7).

$$L(x,c)=\sum_{i\in pos}^{N}x_{i,j}^{p}\log(\hat{c}_i^p)-\sum_{i\in neg}\log(\hat{c}_i^0) \quad (7)$$

where $x_{i,j}^{p}\log(\hat{c}_i^p)$ was the same as $x_{i,j}^{p}$. $(\hat{c}_i^0)$ representing the situation where there was no real target object in the prediction box. The final algorithm process was shown in Figure 3. The neural network consisted of multiple components including convolution, batch normalization, and feature fusion modules. The input of the network went through the first set of convolutional layers Conv4_3, and then features were enhanced through three RFB modules. These modules could increase the RFB and improve the recognition capacity of objects in images. The Fmap was then spatially transformed through ConvID Interp layers to meet the input requirements of subsequent layers. Afterward, the Fmaps were fed into another set of convolutional layers, which further processed the features enhanced by the RFB module. The subsequent feature fusion was achieved through

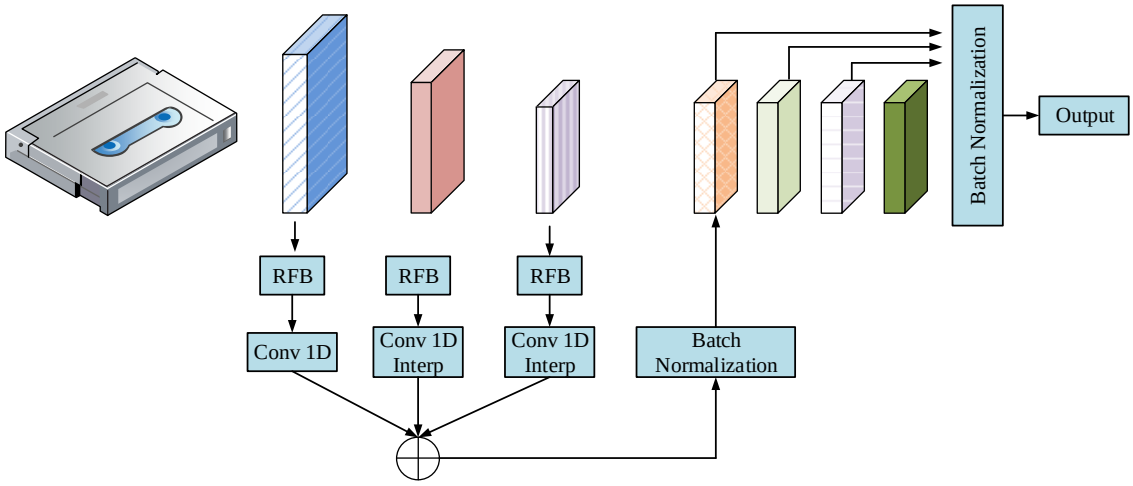


Figure 3. Overall structure of the algorithm.

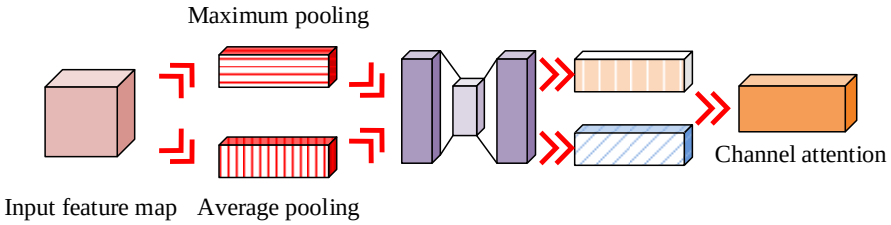


Figure 4. Structure of channel attention module.

connection operations, which combined Fmaps from different layers to capture more levels of information and details. The processed Fmap was fed into the prediction layer, which was responsible for outputting the final recognition result.

Tumor cell automatic recognition model based on AM and CNN

The lack of surrounding objects in weak lateral areas of the observed image made it difficult to identify and highlight the features of small targets. Therefore, AM was introduced into the RFB-based tumor cell automatic recognition model to simulate the human visual attention process, allowing CNN to selectively concentrate on specific parts of the input [14]. Usually, AM could be divided into two types including channel AM and spatial AM. Channel AM mainly focused on the channel dimension of Fmaps to dynamically adjust the importance of different channels [15, 16]. This method learned a set of channel weights or activations to enable the network to adaptively and selectively enhance or suppress the response of each channel. The structure of the Channel Attention (CA) module included a global pooling layer, an FCL, and an activation function. The global pooling layer was used to capture global information, while the FCL was used to generate CA weights, and the activation function introduced nonlinear transformations. In the workflow, each input Fmap was globally pooled, and the CA weights were obtained through an FCL. Then, an activation function was utilized for nonlinear transformation. The CA weights were multiplied by the original map to gain an enhanced feature (Figure 4). The final Spatial Attention (SA) feature obtained through the activation function was shown in equation (8).

$$M_C(F) = Sig\,\mathrm{mo}\,id(F_{\max}^{c\prime} + F_{av}^{c\prime}) \quad (8)$$

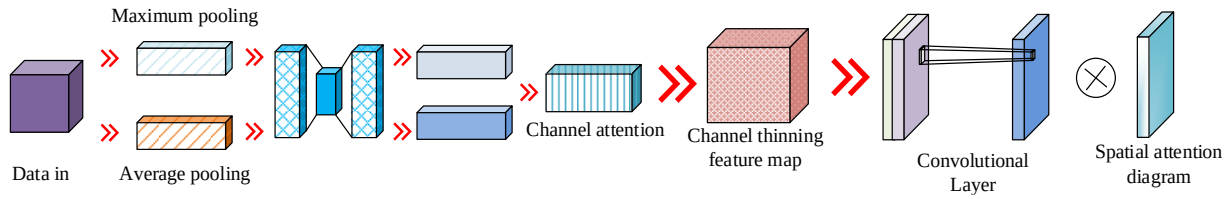


Figure 5. CBAM network.

where $M_C(F)$ was the final CA feature. $F_{\max}^{c}{}'$ was the Fmap obtained after global pooling. $F_{av}^{c}{}'$ was the Fmap obtained after average pooling. The attention features obtained through the CA module were expressed in equation (9).

$$F_C = M_C(F) \otimes F \tag{9}$$

where F was the input Fmap. Spatial AM was a method used to enhance the attention and understanding ability of neural networks on the spatial structure of input data, which mainly focused on the spatial dimension of input data. In spatial AM, the model learned to dynamically adjust the importance of different spatial positions to selectively focus on or ignore information in specific regions. The structure of the SA module was to input feature values for global maximum pooling (GMP) and global average pooling (GAP) in the channel dimension, respectively, to obtain max pooling feature values and average pooling feature values. The two features were concatenated to obtain an Fmap, which was then input into a convolutional layer for dimensionality reduction to obtain an output Fmap. Then, the Sigmoid was utilized to activate, and the final CA feature $M_S(F)$ was calculated as shown in equation (10) [17].

$$M_S(F) = Sigmoid(F_S{}') \tag{10}$$

where $F_S{}'$ was the Fmap obtained through dimensionality reduction. The attention features obtained through the SA module were calculated in equation (11).

$$F_S = M_S(F) \otimes F_C \tag{11}$$

where F_c was the feature gained through the SA module. This study adopted a convolutional block attention module (CBAM) network. The data were first input into the model, and two feature descriptions were generated through GAP and GMP, respectively. The features after GAP and GMP were processed through a shared FCL to generate two CA maps. The final CA weights were obtained by adding the attention maps of two channels and using the Sigmoid. The CA weight was multiplied by the input Fmap channel by channel to gain the weighted Fmap [18, 19]. GAP and GMP were performed on the channel-weighted Fmaps, compressing along the channel dimension, and two spatial Fmaps were generated. These two maps were connected along the channel dimension and processed through a convolutional layer to generate an SA map [20] (Figure 5).

To obtain the final SA weight through Sigmoid, the SA weight with the channel-weighted Fmap element was multiplied by the element to obtain the final output Fmap. The structure of the model introducing AM was shown in Figure 6. The network structure started from the input image and sequentially extracted features through multiple convolutional layers, batch normalization, and ReLU functions. The feature learning was enhanced, focusing on important regions through RFB and attention modules. The features were fused through connection operations, and the final output was the processed image.

Hardware and data resources

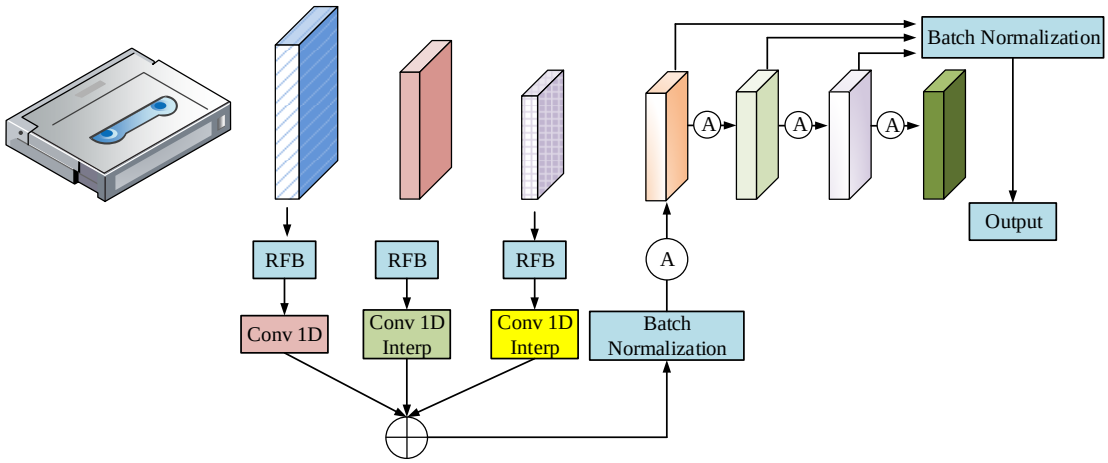


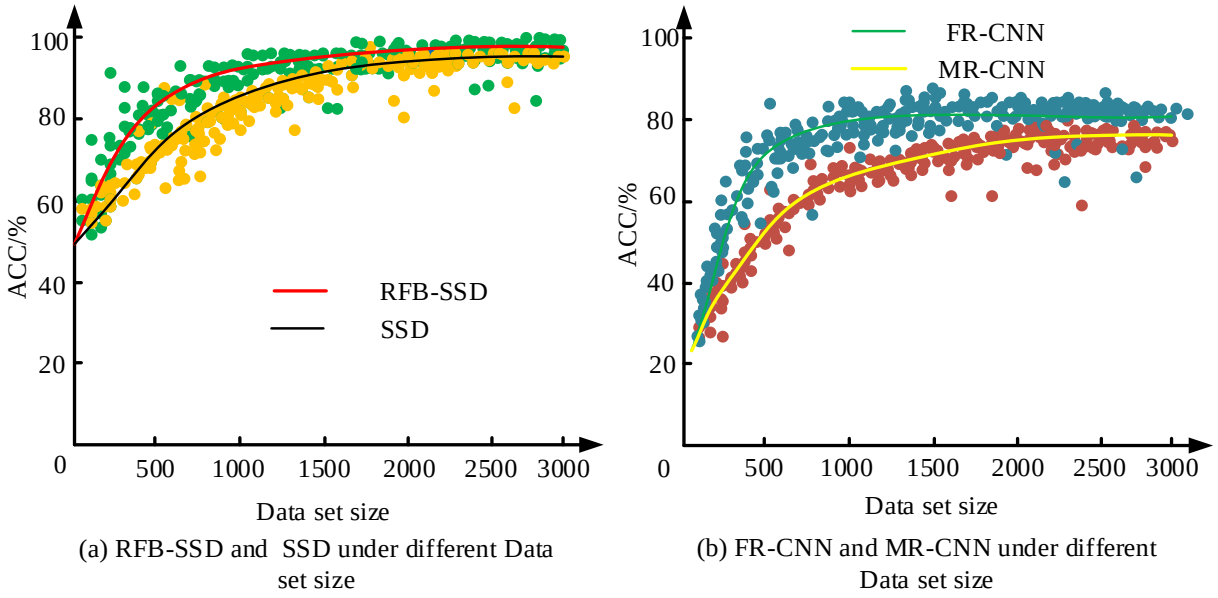
Figure 6. Overall structure of algorithm model.

The hardware used in this study included Intel Core i5-13600KF CPU, NVIDIA Geforce RTX4090 GPU, 24 GB graphics memory, and 32 GB memory. The employed datasets were The Cancer Imaging Archive (TCIA) (https://www.cancerimagingarchive.net/) and the Cameron Grand Challenge (CGC) (https://camelyon16.grand-challenge.org/). TCIA contained data from different cancer types and covered cancer cases at different stages and types. Many image data were accompanied by detailed annotations from radiologists and other medical experts including tumor localization, size, and sometimes pathological information. CGC was a public dataset focusing on pathological image analysis of breast cancer, which contained a lot of cell-level details. The image contained marked tumor areas by experts. To validate the proposed model in this study, the Faster Region CNN (FR-CNN) (https://paperswithcode.com/method/faster-r-cnn) and Mask Region CNN (MR-CNN) (https://github.com/matterport/Mask_RCNN) were employed to compare with the proposed model.

Results and discussion

Performance analysis of tumor cell automatic recognition model based on RFB

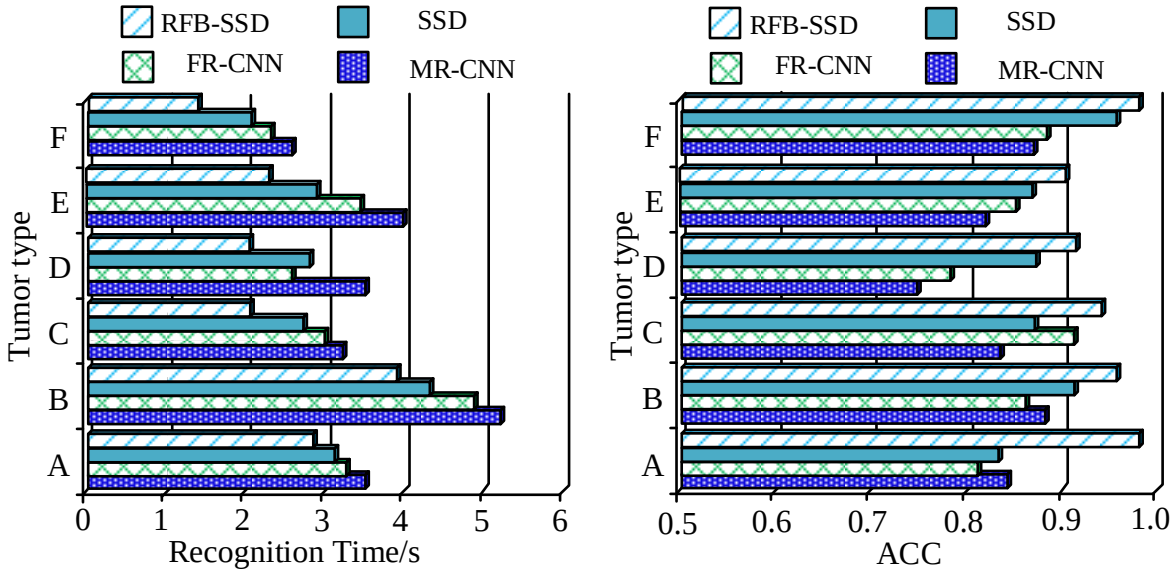
The results of simulation recognition accuracy (ACC) of RFB-SSD and SSD, FR-CNN and MR-CNN demonstrated that, as the dataset increased, the accuracy also increased. When the dataset reached 1,000 and 2,000, the performance of the model tended to be stabilized. When the dataset was 3,000, the ACC of RFB-SSD and SSD models were 97% and 94%, respectively (Figure 7a), while the ACC of FR-CNN and MR-CNN models were 81% and 78%, respectively (Figure 7b). The results indicated that RFB-SSD had excellent ACC. The comparison of identifying different types of tumors in the TCIA dataset showed that, among the various models, RFB-SSD model had the lowest recognition time for various tumors (Figure 8a). The recognition time for lung cancer was the longest one with the recognition times of 3.7 s, 4.2 s, 4.8 s, and 5.3 s for RFB-SSD, SSD, FR-CNN, and MR-CNN, respectively. The recognition time for liver cancer was the shortest with 1.5 s, 2.2 s, 2.4 s, and 2.7 s for RFB-SSD, SSD, FR-CNN, and MR-CNN, respectively. RFB-SSD showed the highest recognition ACC for various tumors. Among them, the recognition ACC for liver cancer was the highest one with the recognition ACCs of 0.98, 0.94, 0.89, and 0.87 for RFB-SSD, SSD, FR-CNN, and MR-CNN, respectively (Figure 8b). The results confirmed that the performance of RFB-SSD model was the most outstanding one. In the CGC dataset, the comprehensive performance of each model was analyzed, and the results



(a) RFB-SSD and SSD under different Data set size

(b) FR-CNN and MR-CNN under different Data set size

Figure 7. Comparison of accuracy of different models.



(a) Comparison of Recognition Time

(b) Comparison of ACC

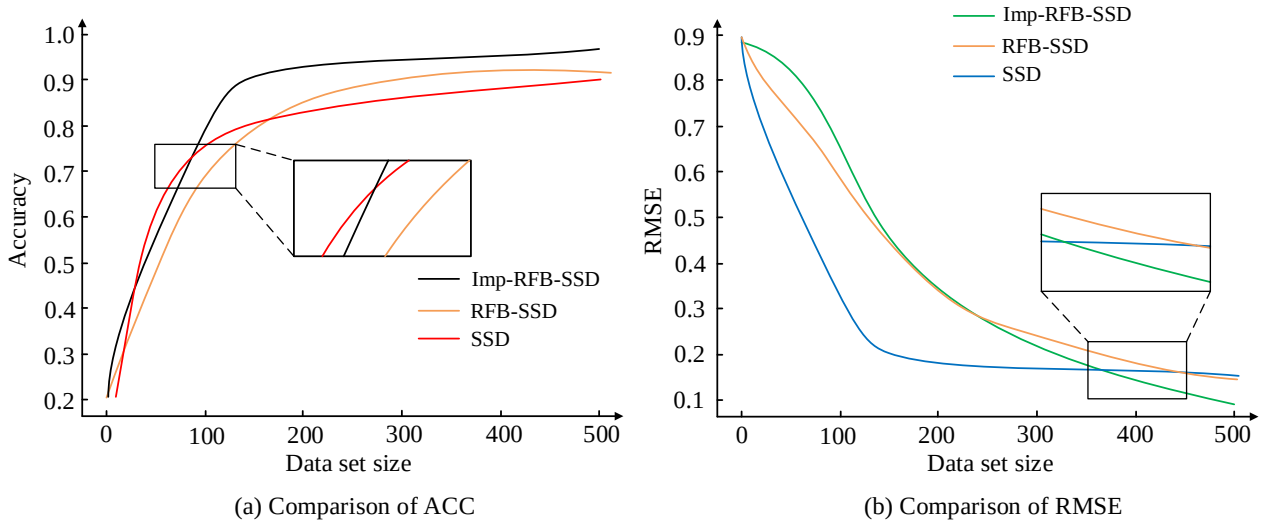
Figure 8. Recognition time and accuracy of various models. A to F represented breast, lung, brain, kidney, head and neck, and liver cancers, respectively.

showed that, among the six types of tumors, each model had a higher recognition rate for head and neck cancer and the lowest recognition rate for renal cancer. RFB-SSD demonstrated excellent model performance with recognition ACC of 92.6%, 94.6%, 89.4%, 82.8%, 93.6%, and 83.8% for breast, lung, brain, kidney, head and neck, and liver cancers, respectively. Among various methods, MR-CNN had the lowest recognition ACC, the highest root mean square error (RMSE) value, and the longest detection time. RFB-SSD had the highest recognition ACC, the lowest

Table 1. Comprehensive performance comparison of various models.

Type	MR-CNN			FR-CNN			SSD			RFB-SSD		
	ACC	RMSE	Time (ms)	ACC	RMSE	Time (ms)	ACC	RMSE	Time (ms)	ACC	RMSE	Time (ms)
A	78.7	37.2	317	81.3	31.1	243	87.4	18.1	162	92.6	15.8	88
B	80.1	38.5	309	82.6	32.4	235	88.7	19.4	154	94.6	17.1	80
C	75.5	34.4	378	78.1	27.9	304	84.2	14.9	223	89.4	12.6	149
D	67.7	26.2	400	71.5	20.1	326	77.6	12.9	245	82.8	4.8	171
E	78.5	37.6	328	82.3	30.9	254	88.4	17.9	173	93.6	15.6	99
F	68.7	27.2	315	72.5	21.1	237	78.6	14.5	209	83.8	5.8	156

Note: A to F represented breast, lung, brain, kidney, head and neck, and liver cancers, respectively.



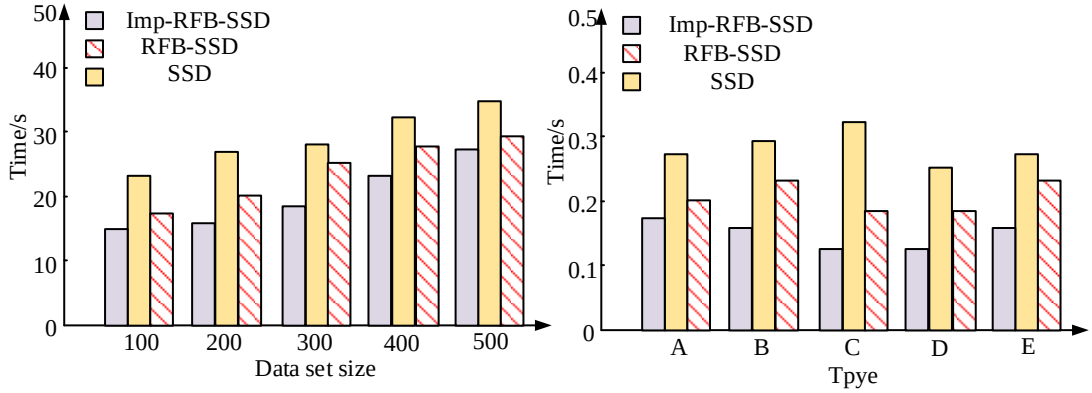
(a) Comparison of ACC (b) Comparison of RMSE

Figure 9. Comparison of ACC and RMSE of various models.

RMSE, and the shortest detection time (Table 1). The results indicated that RFB-SSD model performed the best among the four algorithm models and also had good performance for different types of tumors.

Analysis of tumor cell recognition model based on AM and CNN

The performance of the improved model was validated using TCIA, which was segmented into training and validation sets in a 4:1 ratio. The ACC and RMSE values of each model under different dataset sizes showed that, as the dataset grew, the accuracy of each model also increased, while the RMSE value decreased. When the dataset size reached around 200, the performance of each model reached its maximum and showed convergence. When the dataset size was 500, the ACC values of SSD, RFB-SSD, and Imp-RFB-SSD were 0.89, 0.91, and 0.98, respectively (Figure 9a), while the RMSE values of each model were 0.15, 0.14, and 0.09, respectively (Figure 9b). The Imp-RFB-SSD model demonstrated excellent accuracy and lower RMSE. The comparison results of the response times of various models in the TCIA dataset demonstrated the model training time under distinctive datasets. Imp-RFB-SSD showed the lower training time that was much shorter than that of the other two algorithms (Figure 10a). The recognition time for different tumor types showed that Imp-RFB-SSD had the shorter processing time for different tumors (Figure 10b), which indicated that Imp-RFB-SSD had good recognition efficiency in



(a) Training time under different dataset sizes (b) Recognition time for different types of tumors

Figure 10. Comparison of efficiency of different algorithms.

Table 2. Comprehensive performance comparison of various models in CGC dataset.

Type	SDD		RFB-SSD		Imp-RFB-SSD	
	ACC (%)	Time (s)	ACC (%)	Time (s)	ACC (%)	Time (s)
A	81.1	5.9	85.4	3.8	92.3	2.6
B	82.4	5.8	87.3	3.7	93.9	2.5
C	77.9	6.5	81.8	4.4	88.8	3.2
D	70.1	6.7	75.2	4.7	82.2	3.4
E	80.9	6.2	86.7	3.9	93.1	2.7
F	71.1	5.8	76.2	3.8	83.2	3.0

Note: A to F represented breast, lung, brain, kidney, head and neck, and liver cancers, respectively.

various datasets and different types of tumors. The analysis of the comprehensive performance of each model using CGC dataset showed that Imp-RFB-SSD had better retrieval accuracy and lower retrieval time among the three models. Its recognition ACCs for six types of tumors were 92.3%, 93.9%, 88.8%, 82.2%, 93.1%, and 83.2% for breast, lung, brain, kidney, head and neck, and liver cancers, respectively, with the average recognition time for tumors within 3 s, (Table 2) which confirmed that Imp-RFB-SSD had the best accuracy and recognition efficiency.

Conclusion

This study developed a convolutional model built on SSD to address the subjective and error-prone issues of manual tumor cells identification. This model addressed the shortcomings of the SSD model by introducing the depth of RFB information extraction and AM to enhance the focus of attention. The results proved that, when the dataset was 3,000, the accuracy of RFB-SSD and SSD were 97% and 94%, respectively, while the accuracy of FR-CNN and MR-CNN were 81% and 78%, respectively. When identifying different types of tumors, RFB-SSD had the lowest recognition time for each type of tumor. The recognition time of RFB-SSD, SSD, FR-CNN, and MR-CNN for lung cancer was 3.7 s, 4.2 s, 4.8 s, and 5.3 s, respectively. The recognition time of each model for liver cancer was the shortest as 1.5 s, 2.2 s, 2.4 s, 2.7 s and the recognition accuracy of 0.98, 0.94, 0.89, 0.87 for RFB-SSD, SSD, FR-CNN, MR-CNN, respectively. When the

dataset size was 500, the ACC values of SSD, RFB-SSD, Imp-RFB-SSD were 0.89, 0.91, 0.98, and the RMSE values were 0.15, 0.14, 0.09. The results confirmed that the Imp-RFB-SSD model had better accuracy and recognition efficiency for different types of tumor images.

Acknowledgements

The research was supported by the Teaching Reform and Innovation Project of Colleges and Universities in Shanxi Province (Grant No. J20231298).

References

1. Bhosle K, Musande V. 2023. Evaluation of deep learning CNN model for recognition of devanagari digit. Artif Intell Appl. 1(2):114-118.
2. Zhang SH, Zhan QY, Li WY, Wang Q. 2021. Optimal fusion of multi-focus image: Integrating WNMF and focal point analysis. Int J Mod Phys B. 35(11):17-32.
3. Yilmaz SGO. 2021. Genetic algorithm-based synthetic variable ratio image fusion. Geocarto Int. 36(12):1247-1258.
4. You CS, Yang SY. 2022. A simple and effective multi-focus image fusion method based on local standard deviations enhanced by the guided filter. Displays. 72(41):102146-102149.
5. Long J, Peng Y, Zhao L, Zhou T, Li J. 2022. Semi-blind hyperspectral and multispectral image fusion based on a non-factorization model. Infrared Phys Technol. 125(74):547-567.
6. Zhang C, Nateghinia E, Miranda-Moreno LF, Sun L. 2022. Winter road surface condition classification using convolutional neural network (CNN): Visible light and thermal image fusion. Can J Civil Eng. 49(4):569-578.
7. Miao Z, Zhou F, Yuan X, Xia Y, Chen K. 2022. Multi-heterogeneous sensor data fusion method via convolutional neural network for fault diagnosis of wheeled mobile robot. Appl Soft Comput. 129(35):254-261.
8. Xie H, Zheng W, Chen Y, Shin H. 2022. Camera and LiDAR-based point painted voxel region-based convolutional neural network for robust 3D object detection. J Electron Imaging. 31(5):5301-5322.
9. Guilloteau C, Oberlin T, Berné O, Habart E, Dobigeon N. 2020. Simulated JWST datasets for multispectral and hyperspectral image fusion. Astron J. 160(24):74-82.
10. Guo R, Shen X, Zhang X. 2020. Comprehensive measure for evaluating image fusion algorithm. J Electron Imaging. 29(1):1546-1568.
11. Li H, Zhang L, Jiang M, Li Y. 2020. Multi-focus image fusion algorithm based on supervised learning for fully convolutional neural network. Pattern Recogn Lett. 141(1):45-53.
12. Wang Y, Yan J, Jing Q, Qi Z, Wang J, Geng Y. 2021. A novel adversarial transfer learning in deep convolutional neural network for intelligent diagnosis of gas-insulated switchgear insulation defect. IET Gener Transm Distrib. 15(23):3229-3241.
13. Sian HW, Kuo CC, Lu SD, Wang M. 2023. A novel fault diagnosis method of power cable based on convolutional probabilistic neural network with discrete wavelet transform and symmetrized dot pattern. IET Sci Meas Technol. 17(2):58-70.
14. Huang H, Hu X, Li X, Xiong X. 2022. An efficient loop tiling framework for convolutional neural network inference accelerators. IET Circuits Devices Syst. 16(1):116-123.
15. Wang X, Bai S, Sui Y, Tao J. 2021. The PAN and MS image fusion algorithm based on adaptive guided filtering and gradient information regulation. Inf Sci. 545(32):381-402.
16. Garciaa ABL, Garciab AO, Sentic VE. 2024. Thermal images pre-processing for early detection of breast cancer: A progressive review. Acta Inform Malaysia. 8(1):26-31.
17. Jing Y, Peng H. 2020. Fast and precise orbit determination of satellite GPS data: an image fusion algorithm. J Coast Res. 106(1):592-596.
18. Jia J, Sun J, Zhu Z. 2021. A multi-scale patch-wise algorithm for multi-exposure image fusion. Optik. 248(21):1680-1692.
19. Wang S, Shen Y. 2020. Multi-modal image fusion based on saliency guided in NSCT domain. IET Image Process. 14(13):3188-3201.
20. Pal S, Roy A, Shivakumara P, Pal U. 2023. Adapting a swin transformer for license plate number and text detection in drone images. Artif Intell Appl. 1(3):145-154.