A novel solution of deep learning for endoscopic ultrasound image segmentation: enhanced computer aided diagnosis of gastrointestinal stromal tumor

Sanira Tuladhar1 • Abeer Alsadoon1,2,3,4 • P. W. C. Prasad1,2,3 • Akbas Ezaldeen Ali5 • Ahmad Alrubaie6

Received: 12 August 2020 / Revised: 20 March 2021 / Accepted: 3 January 2022
© The Author(s) 2022

Abstract
Gastrointestinal stromal tumor is one of the critical tumors that doctors do not suggest to get frequent endoscopy, so there is a need for a diagnosis system which can process ultrasound images and figure out the tumor. Many gastrointestinal tumor diagnosis methods were developed, but all of these methods used manual contour rather than automatic segmentation. The research adopts enhanced automatic segmentation to improve the diagnosis of the gastrointestinal stromal tumor with deep convolutional neural networks. This solution’s proposed system is an enhanced automated segmentation methodology using multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution, which segments the ultrasound image automatically into the region of interest (the infected area). Convolutional Neural Network with Class Activation Mapping is done to diagnose an image with the tumor for Four datasets, namely (USS1, SH Hospital, SNUH, BUSI). This proposed system helps to get a clearer tumor image, and the accuracy has increased from 84.275% to 88.4%, and the processing time has reduced from 28.525% to 24.575%. The proposed solution enhanced Automatic Segmentation helped to get clearer tumor image which resulted in increased accuracy and decreased performance time compared to the state-of-the-art. Automatic segmentation overcomes the dependency on the expert for drawing the Region of Interest (ROI).

Keywords Convolutional neural network • Class activation mapping • Multi-scale Gaussian kernel fuzzy • Multi-scale vector field convolution • Region of interest • And automatic segmentation

Abeer Alsadoon alsadoon.abeer@gmail.com

Extended author information available on the last page of the article
1 Introduction

The traditional methods in diagnosing the gastrointestinal stromal tumor involve classifying the image into different sizes, shapes, and textures of the lesion handcrafted features were employed. These kinds of methods are sensitive to the image’s quality and generate poor diagnoses under different scenarios [13]. The classification of the image is highly dependent on the segmentation of the image. If the segmentation has an error, it may cause failure in subsequent classification. The traditional methods lack robustness because of the dependency with handcrafted features in ultrasound images [13].

The latest technology used to solve the traditional method’s limitation is a deep convolutional neural network for the classification of the image. Deep CNN extracts features of different levels that do not require handcrafted features, reduces the expert’s dependency, and decreases the computational time [13]. Automatic segmentation is one of the latest technologies to be implemented during the tumor or cancer diagnosis. It helps to increase the accuracy by extracting the Region of Interest [12].

Computer-aided diagnosis with a deep neural network is the latest technology used to diagnose the tumor [13]. This technology is mainly used in breast cancer and thyroid cancer for the detection and classification of tumors [19]. The exact mechanism with automatic segmentation and the convolutional neural network will be used in the proposed system for the diagnosis of the gastrointestinal stromal tumor. The segmentation of ultrasound images into the region of interest can be done in a proper manner so that many of the features can be extracted. The main improvement is to use multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution, which segments the image automatically into the region of interest.

The proposed system in this solution is an enhanced automatic segmentation methodology using multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution, which segments the image automatically into Region of Interest which is also the infected area. The diagnosis Convolutional Neural Network with Class Activation Mapping is done to figure out network Mt-Net and Sn-Net. Where Mt-Net gives the malignant and non-malignant tissues, and Sn-Net presents the solid nodules.

2 Literature review

2.1 Related work

In this section, an overview of previous studies is presented. Li, et al. have enhanced the radionics method with image segmentation, feature selection, feature extraction, and classification [9]. This enhancement provides a preoperative computer-aided diagnosis system that can differentiate between the low-risk gastrointestinal stromal tumor and the high-risk gastrointestinal stromal tumor from the endoscopic ultrasound image. This solution provides a characteristic operating curve of 0.839, an accuracy of 0.823, and a sensitivity of 0.813, and a specificity of 0.826. As a result, the solution has improved accuracy and performance with good diagnosis classification of the gastrointestinal stromal tumor risk level. However, better image quality should be used to gain as much information as possible. In the solution, the number of negative samples was more wherein further research, the number of positive samples will be added and use the previous marked data to create an automatic segmentation rather than the manual contour. Bonmati, et al. enhanced to develop anatomical landmark
locations to maximize robustness and accuracy, to estimate the target registration errors [2]. The optimal planes were selected on the 90th percentile of target registration error to achieve accuracy and robustness for the endoscopic ultrasound image and the calculation of the target registration error. However, target registration error for different surfaces must be determined in order to work properly to capture the image without the requirement of skilled personnel. E. Rubinstein et al. implemented medical image analysis of PET image using a learning approach to detect prostate cancer [15]. The diagnosis of cancer which ultrasound and MRI could not do, was made possible by this solution. AUC = 0.658 with 256 × 256 resolution to the AUC = 0.812 with 64 × 64 resolution. However, this solution can be better used in an environment with a massive database. It is identified that the main reason for the tumor inferior is the small to large ones; for this resolution, a proper algorithm is required to achieve better performance.

Li, et al. implemented LinSEM technology to maintain the uniformity and consistency of the image segmentation when dealt with different metrics and different objects [10]. All three metrics – Dice coefficient (DC), Jaccard index (JI), and Hausdorff Distance (HD) have been lined into ideal curves acceptability. Dice efficient and Haus Dorff distance reached 8–25% linear, whereas Jaccard index (JI) has a more linear relationship with acceptability. However, the solution is quite specific to the regular shapes when performed with an irregular shape such as a tumor, and then it has to be classified as geometric and morphological attributes. This area needs to be researched more. L. Panigrahi, et al. used the MsGKFCM and MsVFC technology to improve the accuracy of the segmentation and remove the noise and shadowing of the image [12]. this solution tried to provide an accurate lesion segmentation for the computer-aided diagnosis, which was the main problem. The methods used helped in avoiding the manual tracing of the contour and saved the computational time. The improved accuracy in terms of segmentation relative to the methods is 12.797%, 7.864%, 1.735%, 15.084%, 7.531%, and 37.148% compared with K-Means, FCM, GKFCM, NLM, Graph-based and watershed methods, respectively. However, in the future, the ground truth design should be improved, and if possible, radiologists should be made able to score so that it can be used as the base value. The dataset used in the current solution is limited. We should impose the solution to be used in a large mass. The image result is represented in 2D, while we can look further for the 3D images. K. Robinson et al. enhanced the two-state methods for the robustness of the radionics [14]. He has proposed this technology because the images used in radionics aren’t that useful in the context be referred by doctors or radiologists. For this problem, the author proposed the method to figure out the robustness of the image using the features and filtering them into deep that only the best top features were selected. Then using the QDA model and clustering of GE and Hologic, the calculation of robustness with intra vendor and inter vendor is figured out. Intra vendor performance has not much significant difference with ComBat while Inter vendor performance was significantly high than ComBat. However, the future work is mainly targeted for the research and investigation on feature selection and classifier construction method.

Henry, et al. proposed an automatic trimap that can work without the human interaction in order to overcome the limitation of time consumption and dependency figured out from the previous work [6]. The solution is offered by estimating the foreground cue, computing with the lazy snapping algorithm for the foreground segmentation where the background and foreground were differentiated, eliminating the problem of overlapping by labelling the pixel with binary value of 0 and 1 and estimating the unknown region for the mixed pixel and finally generate an automatic trimap. This solution has been found to be more accurate than the previously proposed system and the processing time required is also very less. However, further in future the research can be done in foreground object segmentation which works with the critical images and real time automatic trimap generation can be processed further for the video matting. Z. Shi, et al. enhanced
O-RAW Open source ontology guided radiomics analysis workflow to overcome the problem of standard methodology, universal lexicon to make feature semantically equivalent, proper measure to combine the extracted features into strong effective values in existing radiomics [16]. This solution is mainly dealing about having a standard format and extracting of the features. First the images are parsed through an array and binary mask. This binary mask is converted using CERR and PyRadiomics extension to gain a mask in NRRD format and uploading this format into data store which can be extracted by the SPARQL query. The efficiency was tested in four datasets and the results are (RIDER) 407.3, (MMD) 123.5, (CROSS) 513.2, and (THUNDER) 128.9 s for a single VOI. However, the research regarding RDF graph and storing data should be done to link the histogram data and biological characteristics.

Vieira, et al. enhances the system with automatic detection of the tumor. This solution was proposed because the previously used tool WCE was working for the diagnosis of the tumor, but it gave much information, which was difficult to diagnose by the practitioner [18]. The proposed solution was carried out by converting the image into CIELab color space where the lightning component was removed which gave a better vision of the image. Then EM Algorithm was used to differentiate ROI and background so that the features can be extracted separately and computed with histogram-based measures to detect the tumor. This solution is outperformed by 5% according to the texture features selected from Wavelets and Curvlets with support vector machine. As a result, there is no significant loss in diversity. However, the algorithm for the segmentation should be adequately worked for accuracy, and the best classification scheme should be selected especially for deep learning. W. K. Moon, et al. enhanced the CNN architecture to create a system that helps to diagnose tumors using the image fusion method and image content representations and CNN architecture on ultrasound images [11]. The author used VGG – 16, VGG-like, ResNet, and DenseNet methods to extract the best features and improve the time consumption, memory usage and balance the error between the weak and strong classifier. It provides the performance of CAD system with the accuracy, sensitivity, specificity, AUC with more images and SNUH dataset is 91.10%, 85.14%, 95.77% and 0.969 and with few images with BUSI dataset is 90.77%, 96.67%, 89.00%, and 0.948 showing improved performance in more images. However, there should be a consideration regarding the automatic detection and segmentation method for the extraction of the contour of the tumor, its location, and effects of the surrounding tissues, which may be helpful in the diagnosis.

Chen, et al. enhanced the ultrasound image segmentation algorithm to develop a computer-aided diagnosis system to differentiate neoplastic and non-neoplastic gallbladder polyp [3]. They use the shape prior image segmentation method, PCA and AdaBoost so that a rough contour can be drawn and the accuracy of the dimension can be determined. The proposed segmented system achieved 95% accuracy for outlining the gallbladder region of interest. The 86.52%, and 89.40% respectively compared to 69.05%, 67.86% and 70.17% with convolutional neural network. As a result, the accuracy is really high, but it requires high-quality images and the incorporation of clinical data to improve performance. However, machine learning tools and deep learning networks can be looked into consideration H. Bi, et al. enhanced accurate and efficient methods of ultrasound image segmentation for Active shape models with Rayleigh mixture model clustering to overcome problems with low signal noise inhomogeneity, nonstandard shape, and size of the prostate. In this solution, the images are generated from the original image fetching the boundaries and shape model [1]. Then the image is processed into ASM based model for the segmentation of the image. Extracting the region will solve the inhomogeneity of the ultrasound image and RMM for clustering of the image. The time consumed for the segmentation is less than 8 s. There is a limitation of
inhomogeneity getting grey level if used directly with ASM. Huang, et al. enhanced segmentation of the breast tumor using classification and patch merging to overcome the challenge of the poor quality of the image [7]. The image is first cropped into the tumor to figure out the area of effect. Then the filter is applied to remove the noise and improve the homogeneity of the ultrasound image. An equalizer is applied so that the contrast of the image can be maintained which is the main drawback of the ultrasound image. Using different measures of the grey histogram, grey level co-occurrence matrix, and local binary pattern, the semantics are extracted. The semantics are labeled and classified using the KNN classification method to eliminate the disturbance of misclassification.

Similarly, Fan, et al. presented a study on camouflaged object detection (COD), which aims to identify objects that are “seamlessly” embedded in their surroundings and develops an efficient end-to-end SINet framework, termed Search Identification Network (SINet), which generates more visually favorable results [4]. Le, et al. explored the camouflaged object segmentation problem by proposing a framework that can boot segmentation using classification, it applied to different fully convolutional networks [8]. The classification and segmentation stream are effectively combined to show baseline performance. Yan et al. proposed a novel bio-inspired network, named the MirrorNet, that leverages both instance segmentation and mirror stream for camouflaged object segmentation [20]. The network possesses two segmentation streams: the main and mirror stream corresponding with the original image and its flipped image, respectively. The output from the mirror stream is then fused into the main stream’s result for the final camouflage map to boost segmentation accuracy. The method achieves 89% inaccuracy, Zhang et al. proposed an adaptive context-selection-based encoder-decoder framework that is composed of (LCA) Local-Context Attention module, (ASM) Adaptive Selection Module, and (GCM) Global Context Module [21]. Local-Context Attention modules deliver local-context features from encoder layers to decoder layers, enhancing the attention to the challenging region that is determined by predicting the map of the previous layer. Global-Context Module aims to further explore the global-context features and send them to the decoder layers. Adaptive Selection Module is used for adaptive-selection and aggregation of context features through channel-wise attention. This approach was evaluated on Kvasir-SEG and EndoScene datasets and results in a good performance.

Fan, et al. proposed a parallel reverse attention network (PraNet) for an accurate polyp segmentation in the colonoscopy images [5]. The features had been aggregated in high-level layers using a parallel partial decoder (PPD). A global map had been generated as the initial guidance area based on the combined feature. Also, the boundary cues had been mined by using the (RA) reverse attention module, which can establish the relationship between boundary cues and areas. Similarly, Tian et al. addressed the inappropriate sensitivity to outliers by also learning from inliers. He proposed a few shot anomaly-detection method based on an encoder trained to maximize the mutual information between normal images, feature embeddings, and, followed by a few shot score inference networks, trained with a large set of inliers and a substantially smaller set of outliers [17]. The proposed method was evaluated on the clinical problem of detecting frames containing polyps from the colonoscopy video sequences.

### 2.2 State-of-the-art

Zhang, et al. enhanced a Deep Neural Network to overcome the limitation of classifying images and robust the variations like shape, size, and texture of lesion with multi-scale kernel and skip connections [13]. The proposed method consists of two components to determine the existence
of malignant tissue in an image and recognize solid nodules. They offered an Inception Module to let the network learn about the features with different types, shapes, and sizes and a Class Activation Mapping algorithm to generate a heat map of the same size as the feature map, which can be used for extraction of the features to differentiate the malignant and non-malignant tissues. Class activation mapping algorithm helps to improve classification accuracy and sensitivity for both networks. The accuracy of the system is on the basis of two features Mt-net with 94.98% and Sn-net with 90.13%. As a result, the ultrasound image is used to extract features where a region of interest could be used for more accuracy. This part presents the good features in the blue dotted highlighted box and limitations in the red dotted highlighted box in Fig. 1.

[13] enhanced Class Activation Mapping is the algorithm in which two networks Mt-Net and Sn-Net work collaboratively where the ultra-sonographic image and visualization of Sn-Net is taken as an input for the aggregation of the high features. The classification result gives the output of malignant and non-malignant tumors. The Sn-Net takes the input of ultra-sonographic image and visualization of Mt-net and works in a similar way to get the output of solid nodules [13]. The above-mentioned algorithm works of Mt-Net and Sn-Net, which seems to be partially dependent upon each other, so the Cross Training Algorithm is proposed to train those networks alternatively at a time. Initially, no visualization input is available for Mt-Net, so one of them is picked randomly. This training methodology is proposed to minimize the cost by turning off the facility of choosing the visualization input from networks [13].

Pre processing Simple image is generated into square size so that the size of the image for the processing remains the same and it maintains the ratio of the aspect. To generate the image into square, the first scaling is done [13]. This size is further cropped to 299 × 299 pixels, the desired size that the convolutional neural requires [13]. Limitation: The main limitation of this state-of-the-art is the use of ultrasound images as input. Here, so scaling and cropping of an image are done, but it does not show up the region of interest, which is the main infected area that needs to be diagnosed.

Fig. 1 Block diagram of state-of-the-art (The blue dotted box shows the good features and red dotted box shows the limitations) [13]
**Feature extraction**  
*Class Activation Mapping* is the algorithm in which two networks, Mt-Net and Sn-Net, are worked collaboratively where the ultra-sonographic image and visualization of Sn-Net are taken as an input for the aggregation of the high features [13]. The classification result gives the output of malignant and non-malignant tumors. The Sn-Net takes the input of ultra-sonographic image and visualization of Mt-net and works in a similar way to get the output of solid nodules [13].

**Inception module / cross training algorithm** The above-mentioned algorithm works of Mt-Net and Sn-Net which seem to be partially dependent upon each other so Cross Training Algorithm is proposed to train those networks alternatively at a time [13]. In the beginning, no visualization input is available for Mt-Net, so one of them is picked randomly. This training methodology is proposed to minimize the cost by turning off the facility of choosing the visualization input from networks [13].

The operation of each convolutional layer is defined as in Eq. (1). This is used for the extraction of the features [13].

\[
a^{l+1}_n = f \left( \sum_{m=1}^{M} w^{l}_{nm} \cdot a^l_m + b^{l+1}_n \right)
\]

Where,
- \(a^{l+1}_n\) nth feature map in \((l+1)\)th layer
- \(a^l_m\) mth feature map in th layer.
- \(M\) Number of feature map in layer l.
- \(w^{l}_{nm}\) 2d convolution kernel of the size \(I \times J\) from the mth feature map in the lth layer \((a^l_m)\) to the nth feature map in the posterior layer \((a^{l+1}_n)\)
- \(b^{l+1}_n\) Result of the previous layer containing more than one feature.
- \(f\) nonlinear activation function
- \(l\) layer

According to [13] the ReLU function for nonlinear activation function is defined as in Eq. (2) [13]:

\[
f(x) = \begin{cases} 
0 & x \leq 0, \\
\alpha x & x > 0
\end{cases}
\]

Where,
- \(\alpha\) input space.

Calculation of visualization is done by Class Activation Mapping. It helps to differentiate malignant and non-malignant tissues and extracts solid nodules [13]. The Class Activation Mapping is defined as in Eq. (3) [13]:

\[
CAM_c = \sum_{n=1}^{N} W_{cn}^{l+1} \cdot a^{l-1}_n
\]

Where,
- \(c\) Classification result
Convolutional kernel in layer after global pooling.

nth feature map before global pooling.

The output of the last layer after the processing from the class activation mapping and the high level of feature aggregation is defined as in Eq. (4) [13]:

\[ E_{al} = F_{img} (M_{m}^{MsGKFCM} + F_{MsVFC}) + \lambda \times F_{cam}(X_{cam}) \]  

Where,

- \( a_{l} \) softmax output of the last layer
- \( X_{img} \) breast ultrasound image
- \( X_{cam} \) class activation map
- \( F_{img} \) and \( F_{cam} \) convolutional layers that the two inputs go through
- \( \lambda \) scale factor for the features from \( X_{cam} \)

### 3 Proposed system

After researching for an accurate system that works in distinguishing the high risk and low-risk tumors, a lot of limitations and advantages are figured out. Accuracy, Processing time, and image segmentation are some of the issues to be considered. Class Activation Mapping is the algorithm proposed by [13] in which two networks Mt-Net and Sn-Net work collaboratively where the ultrasonographic image and visualization of Sn-Net are taken as an input for the aggregation of the high features. The classification result gives the output of malignant and non-malignant tumors. The Sn-Net takes the input of ultrasoundographic image and visualization of Mt-net and works in a similar way to get the output of solid nodules. In the proposed research, this state-of-the-art helps to figure out the malignant and non-malignant tumors. This solution not only gives the result of malignant and non-malignant tissue, but it also gives the solid nodule output with the features of different types, shape and the size of the lesion. The limitation of this algorithm is the use of input images which can be enhanced so that the accuracy of the system can be achieved more.

Along with that, Panigrahi et al. proposed Multi-scale Gaussian Kernel Induced Fuzzy C-Means (MsGKFCM) and Multi-scale Vector Field Convolution (MsVFC) to generate a lesion segmentation technique in breast Ultrasound with high accuracy [12]. In the state-of-the-art solution, the input for the processing of feature extraction is the ultrasound image with multiple and various factors to be considered, so segmentation of the image with a proper Region of Interest is required [13]. The blockram of the proposed solution is shown in Fig. 2.

Panigrahi et al. proposed MsGKFCM, a Clustering Method to initialize contour and get accurate lesion margin in an ultrasound image. This Region of Interest is the boundary set out to the area of infection which can be further studied and used for better feature extraction considering the factor of shape, size, and type.

**Feature extraction** Class Activation Mapping is the algorithm proposed in which two network Mt-Net and Sn-Net work collaboratively where the ultra-sonographic image and visualization of Sn-Net is taken as an input for the aggregation of the high features [13]. The classification result gives the output of malignant and non-malignant tumors. The Sn-Net
takes the input of ultra-sonographic image and visualization of Mt-net and works similar way to get the output of solid nodules.

3.1 Proposed equation

The equation with the input image, which was a normal ultrasound image without any segmentation of the region of interest, had been modified. With the help of a segmented image, the features can be extracted better than the normal image. The state-of-the-art has no segmentation provided, which was a limitation. -art We have proposed to solve with the multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution. Multi-scale Gaussian kernel fuzzy clustering (MsGKFCM) is minimized and defined as in Eq. (5) [12]:

\[
J_m^{MsGKFCM} = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ji}^m (1 - K(x_i, a_j)) + \alpha \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ji}^m (1 - K(x_i', a_j)) + \beta \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ji} - \mu_{ji}')^m (1 - K(x_i, a_j)) \tag{5}
\]

Where,
\[\sum_{j=1}^{c} \mu_{ji} = 1,\]

\[\mu_{ji}\] Membership function of pixel.

\[\mu_{ji}'\] Membership function of previous pixel.
Table 1  Enhanced automatic segmentation of image using CNN

| Algorithm: Enhanced Automatic Segmentation of Image to increase accuracy |
|---------------------------------------------------------------|
| Input: Ultrasound image                                       |
| Output: Segmented Image                                       |

BEGIN

Step 1: Initialize MsGKFCM clustering process.
Step 2: Set clustering C > 1 and stopping condition S > 0.
Step 3: Estimate \( \sigma^2 \).
Step 4: Compute membership value \( \mu_{ij} \) and centroid \( a_j \)
Step 5: Update \( \mu_{ij} \) and \( a_j \).
Step 6: Check stopping condition
Step 7: Segment image and cluster centre using equation (7).
Step 8: Initialize MsVFC.
Step 9: Compute \( F_{MAFVC} \) using equation (6)
Step 10: Check stopping condition
Step 11: Boundary of segmented image.
Step 12: Implement CNN relu function for feature extraction.
Step 13: Divide into 1 x 1, 3 x 3 and 5 x 5 convolution layer size.
Step 14: Concat 3 layers into 1 x 1 convolution size.
Step 15: Implement class activation map with ultrasound image and visualization result of Sn-Net using equation (3).
Step 16: Forward computation of the Sn-Net, to be used as the input for the Mt-Net
Step 17: Repeat Step 1-8 to compute the Mt-Net, to be used as the input for Sn-Net

END

Fig. 3 Flowchart of proposed system
Table 2  Accuracy and processing time of state-of-the-art and proposed system using USS1

| Dataset | Original Image | state-of-the-art | Proposed System |
|---------|---------------|------------------|------------------|
| USS1    |               | Frames | Accuracy | Processing Time in ms | Frames | Accuracy | Processing Time in ms |
|         |               | 84     | 32       | 90                 | 26     |
|         |               | 85     | 28       | 88                 | 25     |
|         |               | 82     | 27       | 85                 | 24     |
|         |               | 83     | 22       | 88                 | 22     |
|         |               | 91     | 25       | 94                 | 21     |
|         |               | 85     | 27       | 87                 | 25     |
|         |               | 79     | 32       | 85                 | 27     |
|         |               | 82     | 26       | 87                 | 24     |
|         |               | 90     | 28       | 94                 | 23     |
|         |               | 86     | 28       | 91                 | 24     |
Table 3  Accuracy and processing time of state-of-the-art and proposed system using SH hospital

| Dataset          | Original Image | Frames | Accuracy | Processing Time in ms | Frames | Accuracy | Processing Time in ms |
|------------------|----------------|--------|----------|-----------------------|--------|----------|-----------------------|
| SH Hospital      |                | 87     | 27       | 82                    | 28     | 86       | 25                    |
|                  |                | 78     | 34       | 88                    | 29     | 85       | 26                    |
|                  |                | 61     | 28       | 84                    | 24     | 83       | 23                    |
|                  |                | 40     | 29       | 84                    | 25     | 83       | 24                    |
|                  |                | 52     | 28       | 86                    | 26     | 82       | 25                    |
|                  |                | 51     | 29       | 88                    | 25     | 81       | 24                    |
|                  |                | 65     | 26       | 89                    | 25     | 84       | 24                    |
|                  |                | 64     | 27       | 88                    | 22     | 87       | 23                    |
Table 4  Accuracy and processing time of state-of-the-art and proposed system using SNUH

| Dataset | Original Image | Frame | Accuracy | Processing Time in ms | Frame | Accuracy | Processing Time in ms |
|---------|----------------|-------|----------|-----------------------|-------|----------|-----------------------|
| SNUH    |                | 86    | 92       | 25                    | 92    | 90       | 24                    |
|         |                | 84    | 92       | 25                    | 90    | 89       | 24                    |
|         |                | 86    | 90       | 25                    | 86    | 90       | 24                    |
|         |                | 88    | 90       | 24                    | 88    | 90       | 23                    |
|         |                | 85    | 90       | 24                    | 85    | 90       | 25                    |
|         |                | 83    | 90       | 25                    | 83    | 91       | 26                    |
|         |                | 87    | 90       | 25                    | 87    | 91       | 25                    |
|         |                | 89    | 90       | 25                    | 89    | 92       | 25                    |
|         |                | 89    | 90       | 25                    | 89    | 92       | 27                    |
|         |                | 89    | 90       | 25                    | 89    | 91       | 26                    |
|         |                | 89    | 90       | 25                    | 89    | 91       | 25                    |
Table 5  Accuracy and processing time of state-of-the-art and proposed system using BUSI

| Dataset | Original Image | state-of-the-art | Proposed System |
|---------|----------------|------------------|-----------------|
|         | Frames         | Accuracy         | Processing Time in ms | Frames         | Accuracy         | Processing Time in ms |
| BUSI    | 84             | 32               | 88              | 26             |
|         | 81             | 28               | 86              | 25             |
|         | 83             | 27               | 87              | 24             |
|         | 84             | 22               | 88              | 22             |
|         | 81             | 25               | 84              | 21             |
|         | 83             | 32               | 87              | 27             |
|         | 85             | 26               | 86              | 24             |
|         | 84             | 28               | 89              | 23             |
|         | 89             | 28               | 92              | 24             |
\[ K(x_i, a_j) = \exp\left(-\frac{||x_i - a_j||}{\sigma^2}\right). \]

*K* Kernel  
*\(\mu\)* 0.2 constant  
*\(\alpha\)* scale factor  
*\(\beta\)* scale factor  
*x<sub>i</sub>* feature vector of *i* and *a<sub>j</sub>* pixel  
*a<sub>j</sub>* pixel  
*M* weighted component

Multi-scale vector field convolution (MsVFC) over a range of scale is defined as in Eq. (6) [12]:

Fig. 4 Accuracy and processing time of dataset USS1

Fig. 5 Accuracy and processing time of dataset SH hospital
\[ F_{MaVFC} = f(x, y)^m [m_{norm}(x, y : t), n(x, y)] \]  

Where,

- \( t \) range of scales
- \( x \) and \( y \) edge map
- \( m \) magnitude function

Scale factors \( \alpha \) and \( \beta \) are removed from JMsGKFCM because it was affecting the clustering result [12].

Modified Multi-scale Gaussian kernel fuzzy clustering (MsGKFCM) defined as in Eq. (7):

\[
M^J_{msGKFCM} = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ji}^m \left( 1-K(x_i, a_j) \right) + \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ji}^m \left( 1-K(x_i, a_j) \right) 
+ \sum_{i=1}^{n} \sum_{j=1}^{c} \left( \mu_{ji} - \mu_{ij}^m \right) \left( 1-K(x_i, a_j) \right) 
\]  

Fig. 6 Accuracy and processing time of dataset SNUH

Fig. 7 Accuracy and processing time of dataset BUSI
Where,
\[ \sum_{j=1}^{c} \mu_{ji} = 1, \]

\( \mu_{ji} \) Membership function of pixel.

\( \mu'_{ji} \) Membership function of previous pixel.

\[ K(x_i, a_j) = \exp(-\|x_i - a_j\| / \sigma^2). \]

\( K \) Kernel

\( \mu \) 0.2 constant

\( x_i \) feature vector of i and \( a_j \) pixel

\( a_j \) pixel

\( M \) weighted component

Ultrasound image, which was the input for the output layer of the state-of-the-art, is replaced by the summation of Modified Multi-scale Gaussian kernel fuzzy clustering and Multi-scale vector field convolution (FMsVFC). The enhanced output layer is defined as in Eq. (8):

\[ E\alpha' = F_{img}(MJ_{m}^{MsGKFCM} + F_{MsVFC}) + \lambda * F_{cam}(X_{cam}) \] (8)

### 3.2 Area of improvement

The equation with the input image, which was a normal ultrasound image without any segmentation and region of interest, is modified. With the help of a segmented image, the features can be extracted better than with the normal image because the segmentation somehow has already extracted the area of investigation. **Why Automatic Segmentation of an image?** With the use of automatic segmentation region of interest, the actual infection area is extracted and can be used as the segmented region, which is to be further calculated by

---

Table 6  Comparison table

| Name of the solution | Proposed solution | State-of-the-art solution |
|----------------------|-------------------|---------------------------|
| MsGKFCM method       |                   | Deep Neural Network       |
| Accuracy             | Average accuracy 88.4% | Average accuracy 84.275% |
| Processing Time      | 24.575%           | 28.525%                   |
| Proposed equation    | \( J_{m}^{MsGKFCM} = \sum_{i=1}^{n} \sum_{j=1}^{c} \mu_{ij}^{m}(1-K(x_i, a_j)) + \alpha \sum_{i=1}^{n} \sum_{j=1}^{c} \mu'_{ij} \) + \beta \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij} - \mu'_{ij}) \) | \( d' = F_{img}(X_{img}) + \lambda * F_{cam}(X_{cam}) \) |
| Contribution 1       | Automatic enhanced segmentation helps to avoid manual tracing. | Automatic segmentation isn’t implemented. |
| Contribution 2       | Increase accuracy by 5% and decrease performance time by 5% compared to the state-of-the-art. | |

23861 Multimedia Tools and Applications (2022) 81:23845–23865
defining the type of part and differentiating it from the background for the diagnosis of the tumor automatically. The state-of-the-art does not provide the proper Region of Interest which is the limitation of the solution. The importance of the proposed method is to improve the accuracy of the segmentation and define the area of interest. The proposed method helped avoid the manual tracing of the contour and saved the computational time with enhanced accuracy. The algorithm of enhanced automatic segmentation of the image is shown in Table 1, and the flowchart of the proposed system is shown in Fig. 3. We have presented automatic segmentation of the lesion for the generation of the tumor’s contour and diagnosis automatically with the full features of an image provided.

4 Results and discussion

Python v 3.7 is used for testing CNN model and building the result set. Framework Keras, Sklearn, and Tensorflow are used to support the coding. Four datasets, namely (USS1, SH Hospital, SNUH, BUSI), are used for the training and testing of the model. The datasets are collected from the online resource based on different ages, groups, and types of the sample. The size of each image was set to 128 × 128. The frames for training and testing are set to 10. Relu cross-validation function is used to divide the dataset into training and testing sets.

During the segmentation of the image, speckle noises are removed from the ultrasound image using SRAD filter, and MsGKFCM technique is used to segment the ultrasound image. During the feature extraction, the features to define tumors are extracted using CNN. In the classification stage, the features extracted are taken from the previous convolutional layer. The proposed model detects the tumor and tumor areas.

We use the processing time and accuracy as a performance measure. The equation to compute the accuracy is:

\[ \text{Accuracy} = \frac{TN + TP}{TP + FP + FN + FN} \]

where TN, TP, FP, FN, are the number of true negatives, true positives, false positives, and, false negatives respectively.

Four (Tables 2, 3, 4 and 5) are prepared with the four different datasets. The training and testing set are divided into 10 groups each. The results are created during the Classification stage, where the accuracy and the processing time are calculated. The given Fig. 4 below shows that the proposed system has increased accuracy by 5% and decreased processing time by 5% compared to the state-of-the-art.

Figure 4 is the plotted figure of the accuracy and processing time of the dataset USS1. The x-axis is defined as the number of result sets (ultrasound images) of the dataset, and the y-axis is defined as the percentage level of accuracy and processing time for both state-of-the-art and proposed systems.

Figure 5 is the plotted figure of the accuracy and processing time of the dataset SH Hospital. The x-axis is defined as the number of the result sets (ultrasound images) of the dataset, and y-axis is defined as the percentage level of accuracy and processing time for both state-of-the-art and proposed systems.

Figure 6: the plotted figure of the accuracy and processing time of the dataset SNUH. The x-axis is defined as the number of the result sets (ultrasound images) of the dataset and the y-
axis is defined as the percentage level of accuracy and processing time for both state-of-the-art and proposed systems.

Figure 7 is the plotted figure of the accuracy and processing time of the dataset BUSI. The x-axis is defined as the number of the result sets (ultrasound images) of the dataset and y-axis is defined as the percentage level of accuracy and processing time for both state-of-the-art and proposed system.

From the above results, there is a significant difference between the state-of-the-art and the proposed system, as illustrated in Table 6. The accuracy of the proposed system has increased by 5% and the processing time outperformed by 5%. The improvement had been figured between both the state-of-the-art and proposed system by running both algorithms for the same dataset provided.

Automatic segmentation is used to improve the accuracy of the tumor’s diagnosis and remove the noise and shadowing of the image. Accurate lesion segmentation for the computer-aided diagnosis has been proposed in the article, which was the main problem in state-of-the-art. With the use of automatic segmentation Region of Interest, the actual area of infection is extracted and can be used as the segmented region, which is to be further calculated defining the type of region and differentiating it from the background for the diagnosis of the tumor automatically. The methods MsGKFCM and MsVFC technology helped in avoiding the manual tracing of the contour and saved the computational time. Comparison between the state-of-the-art and the proposed system is illustrated in Table 6.

5 Conclusion and future work

It is stated in most of the articles that CNN has outperformed when it comes to the implementation and detection of the object. The accuracy and the processing time is the good factor of using CNN in most of the diagnosis and detection. Even though it is one of the good methods to use, there is a limitation in the automatic segmentation of the image, which is the dependency upon the expert in the input. Manual tracing of the contour is one of the problems that have been figured out in most of the articles. Automatic segmentation is used so that the maximum accuracy can be taken as output and helps in the detection and diagnosis of the tumor. The proposed solution, Enhanced Automatic Segmentation, helped to get more precise and accurate tumor images which resulted in increased accuracy and performance time compared to the state-of-the-art. Automatic segmentation overcomes the dependency on the expert for drawing the Region of Interest (ROI). Therefore, the accuracy has increased by an average of 5%, with the improved processing time outperformed by 5% compared to the state-of-the-art. The resulting image is represented in 2D while it can be looked further for the 3D images.

Abbreviations  
CNN, Convolution Neural Network; Mt-Net, Malignant Network; Sn-Net, Solid Nodule Network; CAM, Class Activation Mapping; MsGKFCM, multi-scale Gaussian kernel fuzzy; MsVFC, multi-scale vector field convolution; ROI, Region of Interest; US, Ultrasound; Ms, Millisecond

Funding  Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and
indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Bi H, Jiang Y, Tang H, Yang G, Shu H, Dillenseger J-L (2020) Fast and accurate segmentation method of active shape model with Rayleigh mixture model clustering for prostate ultrasound images. Comput Methods Prog Biomed 184:105097
2. Bonmati E, Hu Y, Gibson E, Uribarri L, Keane G, Gurusami K, Davidson B, Pereira SP, Clarkson MJ, Barratt DC (2018) Determination of optimal ultrasound planes for the initialisation of image registration during endoscopic ultrasound-guided procedures. Int J Comput Assist Radiol Surg 13(6):875–883
3. Chen T, Tu S, Wang H, Liu X, Li F, Jin W, Liang X, Zhang X, Wang J (2020) Computer-aided diagnosis of gallbladder polyps based on high resolution ultrasonography. Comput Methods Prog Biomed 185:105–118
4. Fan D, Ji G, Sun I, Cheng M, Shen J, Shao L (2020) Camouflaged Object Detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
5. Fan D, Ji G, Zhou T, Chen G, Fu H, Shen J, Shao L (2020) PraNet: parallel reverse attention network. arXiv:2006.11392v4 [eess.IV]
6. Henry C, Lee S-W (2019) Automatic trimap generation and artifact reduction in alpha matte using unknown region detection. Expert Syst Appl 133:242–259
7. Huang Q, Huang Y, Luo Y, Yuan F, Li X (2020) Segmentation of breast ultrasound image with semantic classification of superpixels. Med Image Anal 61:101–105
8. Le T, Nguyen TV, Nie Z, Tran M, Sugimoto A (2019) Anabranche network for camouflaged object segmentation. Comp Vision Image Underst 184:45–56
9. Li X, Jiang F, Guo Y, Jin Z, Wang Y (2019) Computer-aided diagnosis of gastrointestinal stromal tumors: a radiomics method on endoscopic ultrasound image. Int J Comput Assist Radiol Surg 14:1635–1645
10. Li J, Udupa JK, Tong Y, Wang L, Torigian DA (2020) LinSEM: linearizing segmentation evaluation metrics for medical images. Med Image Anal 60:101–601
11. Moon WK, Lee Y-W, Ke H-H, Lee SH, Huang C-S, Chang R-F (2020) Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Comput Methods Prog Biomed 190:105–361
12. Panigrahi L, Verma K, Singh BK (2019) Ultrasound image segmentation using a novel multi-scale Gaussian kernel fuzzy clustering and multi-scale vector field convolution. Expert Syst Appl 115:486–498
13. Qi X, Zhang L, Chen Y, Pi Y, Chen Y, Lv Q, Yi Z (2019) Automated diagnosis of breast ultrasonography images using deep neural networks. Med Image Anal 52:185–198
14. Robinson K, Li H, Lan L, Schacht D, Giger M (2019) Radiomics robustness assessment and classification evaluation: a two-stage method demonstrated on multivendor FFDM. Med Phys 46(5):2145–2156
15. Rubinstein E, Salhov M, Nidam-Leshem M, White V, Golan S, Bernstine H, Groshar D, Averbuch A (2019) Unsupervised tumor detection in dynamic PET/CT imaging of the prostate. Med Image Anal 55:27–40
16. Shi Z, Traverso A, Soest J, Dekker A, Wee L (2019) Technical note: ontology-guided radiomics analysis workflow (O-RAW). Med Phys 46(12):5677–5684
17. Tian Y, Maicas G, Pu LZ, Singh R, Verjans J, Carneiro G (2020) Few-shot anomaly detection for polyp frames from colonoscopy. arXiv:2006.14811v1 [cs.CV]
18. Vieira PM, Freitas F, Valente J, Vaz IF, Rolanda C, Lima CS (2019) Automatic detection of small bowel tumors in wireless capsule endoscopy images using ensemble learning. Med Phys 47(1):52–63
19. Wang L, Zhang L, Zhu M, Qi X, Yi Z (2020) Automatic diagnosis for thyroid nodules in ultrasound images by deep neural networks. Med Image Anal 61:101–665
20. Yan J, Le T et al (2020) MirrorNet: Bio-inspired adversarial attack for camouflaged object segmentation. arXiv: 2007.12881v1[cs.CV]
21. Zhang R, Li G, Li Z, Cui S, Qian D, Yu Y (2020) Adaptive Context Selection for Polyp Segmentation. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. Lecture notes in computer science, vol 12266. Springer, Cham. https://doi.org/10.1007/978-3-030-59725-2_25
Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sanira Tuladhar\textsuperscript{1} · Abeer Alsadoon\textsuperscript{1,2,3,4} · P. W. C. Prasad\textsuperscript{1,2,3} · Akbas Ezaldeen Ali\textsuperscript{5} · Ahmad Alrubaie\textsuperscript{6}

\textsuperscript{1} School of Computing Mathematics and Engineering, Charles Sturt University (CSU), Wagga Wagga, Australia
\textsuperscript{2} School of Computer Data and Mathematical Sciences, Western Sydney University (WSU), Sydney, Australia
\textsuperscript{3} Kent Institute Australia, Sydney, Australia
\textsuperscript{4} Asia Pacific International College (APIC), Sydney, Australia
\textsuperscript{5} Department of Computer Science, University Of Technology, Baghdad, Iraq
\textsuperscript{6} Faculty of Medicine, University of New South Wales, Sydney, Australia