Journal of Clinical and Medical Images

Research Article

ISSN: 2640-9615 | Volume 5

Traumatic Bleeding Detection Based on Fusion of 3D Shape and Local Texture Features

Yang L[1,2*], Nakaguchi T[3], Jiang H[2], Yang T[1], Kimura H[1], Arai K[4], Nakada T[4] and Xuan P[5]

[1]Graduate School of Science and Technology, Chiba University, 263 8522, NishiChiba, Japan

[2]School of Information Engineering and Research Center of Digital Medical Image Technique, Zhengzhou University, 450001, Zhengzhou, Henan, China

[3]Center for Frontier Medical Engineering, Chiba University, 263 8522, NishiChiba, Japan

[4]Department of Emergency and Critical Care Medicine, Chiba University Graduate School of Medicine, 1 Chome81 Inohana, Chuo Ward, Chiba, 2608677 Japan

[5]School of Computer Science and Technology, Heilongjiang University, 150080, Harbin, Heilongjiang, China

[*]Corresponding author:

Lei Yang,
Graduate School of Science and Technology, Chiba University, 263 8522, NishiChiba, Japan, E-mail: leiyang9201@gmail.com

Received: 10 Aug 2021
Accepted: 28 Aug 2021
Published: 02 Sep 2021

Keywords:

CT Images; 3D Convolutional Networks; Feature Fusion; 3D point cloud; Computer aided diagnosis

Citation:

Yang L et.al. Traumatic Bleeding Detection Based on Fusion of 3D Shape and Local Texture Features. J Clin Med Img. 2021; V5(15): 1-12

1. Abstract

Early detection of traumatic bleedings is very important for subsequent diagnosis and improving the survival rate of patients. In this study, an automatic detection approach for traumatic bleeding from contrast enhanced Computer Tomography (CT) images of whole body is proposed by combining the traditional image processing and deep learning method. Firstly, aiming at the problem of small target bleeding detection, a 3D reconstruction pre-processing method is designed to convert CT slice images into 3D volume data with smaller resolution. Secondly, a bleeding region segmentation model is constructed based on the 3D U-Net framework to obtain three-dimensional candidate features. The proposed model uses the spatial pyramid to extract multi-scale features, introduces two attention modules to enhance the difference between bleeding area and other tissues both in position and channel level, and adds a 3D discriminator to correctly distinguish the target from the background, helping to maximize the extraction of relevant 3D candidate bleeding regions from CT images. Thirdly, the extracted 3D bleeding regions are fused with two-dimensional candidate bleeding features extracted by a multi-threshold method to reduce false positive prediction. Finally, the fused candidate features are formed into a 3D point cloud, which represents the shape information of the bleeding region. At the same time, the texture features are calculated from the Gray Level Co-occurrence Matrix (GLCM) that describes the statistical characteristics of normalized CT values. Then a PointNet based double encoder classification network is established to further decrease false positive predictions by performing a binary classification based on the 3D point cloud and the texture features. We evaluated the final results according to the standard of emergency medical experts. The results show that the total sensitivity is 89.19% with the small data set and relatively simple network structure, and there are 0.9367 false positive predictions per image on average.

2. Introduction

Worldwide, there are many challenges in human health and emergency medicine in regard to traumatic hemorrhage [1]. Traumatic bleeding is usually caused by a serious impact such as a car accident or falling objects hitting the human body. Traumatic bleeding may occur in any part of the human body, and the amount of bleeding may be small in the early stage, but increases rapidly in a short period of time, seriously endangering human life. Uncontrolled bleeding greatly increases the risk of death [2]. Therefore, early detection and control of traumatic bleedings are very important for subsequent diagnosis and improving the survival rate of patients. It is a common method to detect internal bleeding in early stage of the whole body by enhanced Computer Tomography (CT) imaging

in patients with trauma [3]. According to the whole-body enhanced CT image, the doctor can identify the bleeding location, diagnose the severity, and subsequently make the treatment plan.

The whole-body CT examination of each case can obtain three-dimensional volume data composed of hundreds or even thousands of two-dimensional CT slice images, which can provide three-dimensional shape and local texture information of different parts, and support doctors to make an accurate diagnosis from multiple angles and dimensions.

Although enhanced CT images are of great value for early detection and accurate diagnosis of internal bleeding, it is very difficult to make correct diagnoses by relying on the doctor's eye to observe these two-dimensional CT slices. Even a trained doctor can take a long time to locate all the bleeding areas by looking at the whole-body CT image. In the emergency state, due to fatigue and subjective factors, it is easy to miss the diagnosis of a small target bleeding point. Therefore, there is an urgent need to use Computer-Aided Diagnosis (CAD) and artificial intelligence technology to improve the diagnosis efficiency and level of doctors. In order to help emergency doctors rapidly identify the bleeding area in the whole body and carry out emergency rescue for patients, this study investigates the automatic detection method of traumatic bleeding from the enhanced CT images of the whole body.

3. Related Work

The CAD system can assist with disease diagnosis and reduce the time needed in the diagnosis process, so that emergency doctors have more time to treat and rescue patients. CAD systems can be divided into two types: Computer-Aided Detection (CADe) and Computer- Aided Diagnosis (CADx). The goal of CADe is to locate the regions of interest in the image to find specific anomalies. The CADx system provides medical assistance for clinicians to distinguish the type, severity, stage, progression, and deterioration of disease [4]. Numerous related studies and methods have been proposed and can be classified into three major categories: the traditional image processing methods for anomalies detection [5], deep learning-based approaches [6, 7], and related applications [8].

The traditional image processing algorithms often used for anomalies detection includes the threshold methods, segmentation, machine learning, and recognition algorithms. For example, Bhadauria et al. proposed a method for detecting cerebral hemorrhage using the fuzzy C-means means and Level Set algorithm, and their proposed method had an average sensitivity of 99.58% [9]. Kumar et al. proposed an entropy-based automatic unsupervised approach for brain intracranial hemorrhage segmentation [10]. Kadam and Dhole applied GLCM (gray level co-occurrence matrix) and KSCM (Kernel Support Vector Machine) for brain tumor detection [11]. In these studies, bleeding detection based on specific body regions has been achieved. However, the above traditional image processing algorithms all suffer from low reliability for three reasons: (1) the image intensity is distributed unevenly, (2) the shallow features are extracted with expert knowledge, and (3) lack of a unified detection framework. Different algorithms need to be designed to detect internal bleeding in different body parts. In addition, due to the diversity of bleeding areas in the images of whole-body CT scanning, the bleeding areas may have arbitrary size, direction, shape, and texture features, and it is very difficult to detect various bleeding areas in different body parts at the same time by using the traditional image processing algorithms.

To address these challenges, deep learning-based methods are increasingly studied in recent years. They are able to extract features automatically and have been achieving good performance in image-related tasks [12–15].

The deep convolutional network-based method, which is now widely used in medical image processing [16], is also used for bleeding detection. The excellent feature learning ability of deep convolutional networks can obtain good accuracy and robustness from the high density information included in medical image data [17]. Arjun et al. applied a deep convolutional network for intracranial hemorrhage detection and obtained a sensitivity of 81% [18]. Candefjord et al. investigated the ability of broadband microwave technology in detecting subdural hematoma, they applied a human cranium phantom and numerical simulations of SDH and identified 100% and 82-96% sensitivity, respectively [19].

However, the deep learning method that can be applied to the detection of systemic hemorrhage has not been reported in the literature. The problem to be solved in this study is how to obtain the most accurate and reliable bleeding area detection results based on whole-body enhanced CT images in light of the many challenges. (1) The first challenge is that the deep learning model is driven by big data. Large amounts of data are required for deep learning methods to avoid over fitting. We have not found the relevant public database because it is time-consuming and laborious to collect and annotate the data. (2) The second challenge is that the deep learning network does not perform well with highly unbalanced data. The deep learning model trained by unbalanced data will reduce the detection sensitivity of the model. (3) The third challenge is that common deep learning models for target detection, such as Faster R-CNN [20] and YOLO Series [21–23], are not suitable for bleeding detection, because bleeding diagnosis needs to locate and measure the number of bleeding targets and the amount of bleeding. Moreover, in every case, the number of bleedings differ, and most of the bleeding areas are small, so it is more challenging to detect multiple small targets from large-scale images covering the whole body. (4) The fourth challenge is to improve the detection accuracy by solving the inadequacy of automatic feature extraction of the deep learning method: the convolutional neural network does not consider the relationship between different image blocks and has lost the overall boundary features of the detection target.

In this study, whole-body CT enhanced scan data based on trau-

matic bleeding cases were collected, and all bleeding areas in each case were labeled by two emergency specialists. An automatic detection approach for traumatic bleeding from contrast enhanced CT images of whole body is proposed by combining the traditional image processing and deep learning method. CT values of all DICOM images are normalized to the range of 8-bit depth image by window processing. Processed images are cropped to a smaller resolution to help detect small targets and balance the ratio of segmentation target and background. Pre-processed images are sent to the adversarial segmentation network with attention modules and spatial pyramid modules, which fuse the features in different channels, positions and scales. The reconstructed segmentation prediction is fused with results of multi CT value thresholds extraction by a logical and operator to obtain global features and decrease false positives. The post-processed result is sent to a classifier, which combines the shape feature and texture feature by taking point cloud data converted by segmentation results and texture features calculated from GLCM. The output of the classifier reduces the false positive prediction by controlling the probability score to obtain the final result.

4. Materials and Methods

The dataset is provided by the Affiliated Hospital of Chiba University, including 33 contrast-enhanced CT cases with different amounts of traumatic bleedings.

4.1. Dataset

Main The detailed information of the dataset is shown in (Table 1). The dataset consists of 33 contrast-enhanced whole-body CT cases. Each case contains 665 to 2113 slices of DICOM images and 1 to 60 spatial isolated traumatic bleedings. The original resolution of all DICOM images is 512×512 pixels and slice thickness 1.0 mm; the pixel dimension ranges from 0.546×0.546 to 0.976×0.976 mm. Bleeding labels are provided in the form of raw files.

Figure 1 shows examples of bleeding in the head, abdomen, and thighs. As shown in Figure 1, the contrast agent can make the bleeding site have higher CT value in enhanced CT scan and make the CT value of the bleeding site become larger and brighter. However, due to the uneven distribution of contrast agents in the human body, it is difficult to obtain high detection sensitivity by only relying on the convolution layer of deep learning to extract the local features from the input images. Furthermore, there are many tissues in the human body, such as blood vessels and bones, whose CT values are similar to the bleeding areas, which can easily lead to detection of a large number of false positives. Especially, as shown in (Figure 1 (b)), when the bleeding area is very close to or directly adheres to the blood vessels, bones and other tissues, the existing methods often misjudge it as a whole bleeding area, thus affecting the detection accuracy.

From the actual needs of emergency doctors, automatic detection of bleeding should first avoid missed diagnoses, otherwise it will endanger the lives of patients, and the number of false positive prediction should be as small as possible to save diagnosis time. Therefore, it is necessary to develop a high sensitivity and high specificity automatic bleeding detection method.

4.2. Proposed Detection Method

If the bleeding area is detected in the two-dimensional CT slice image, it is easy to cause false detection or missed detection because the bleeding area, blood vessels, and bones all show imaging characteristics of white highlighted areas. Therefore, the accurate detection of bleeding areas is more suitable for direct three-dimensional space.

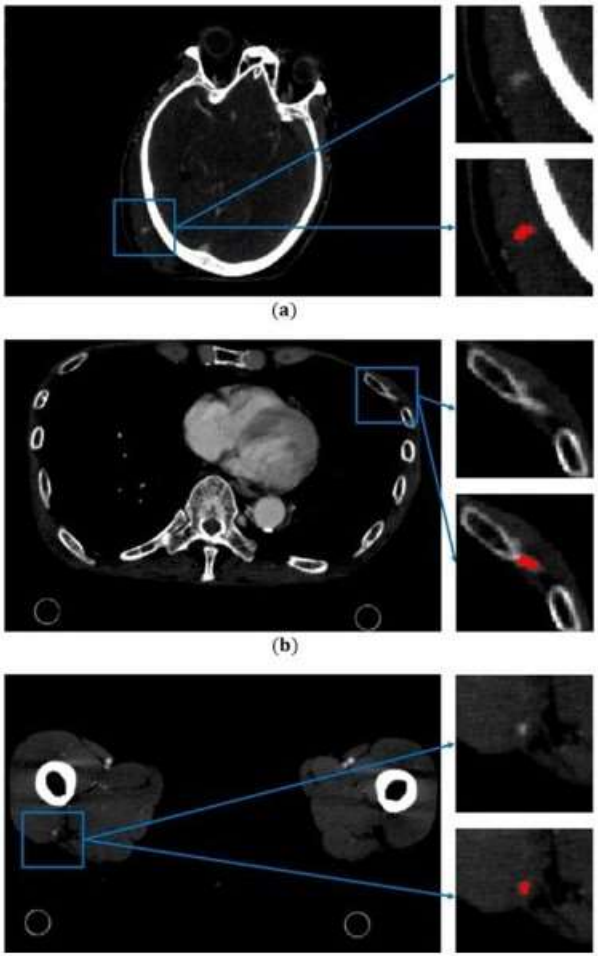

Figure 1: Example of bleedings in the dataset, red parts are bleeding labels.

Table 1: Details of the dataset

Element	Value
Number of cases	33
Image size	512 ´ 512 pixel
Pixel dimension	0.546 × 0.546 - 0.976 × 0.976 mm
Slice thickness	1.0 mm

The overview of the entire process is shown in (Figure2), including segmentation and classification phases. Normalized images are processed to generate 3D volumes as the input of the generator, which is a 3D version of the U-shape network [24] with attention modules and spatial pyramid blocks. A 3D discriminator helps update parameters of the generator network and form an adversarial training process. The features from the gray and texture levels from every spatial isolated region in the post-processed prediction is extracted, then converted to point cloud data. A double encoder classifier uses point cloud data and texture features as input to perform binary classification. The final result takes the probability scores of true positive predictions as a threshold to prevent sensitivity from decreasing.

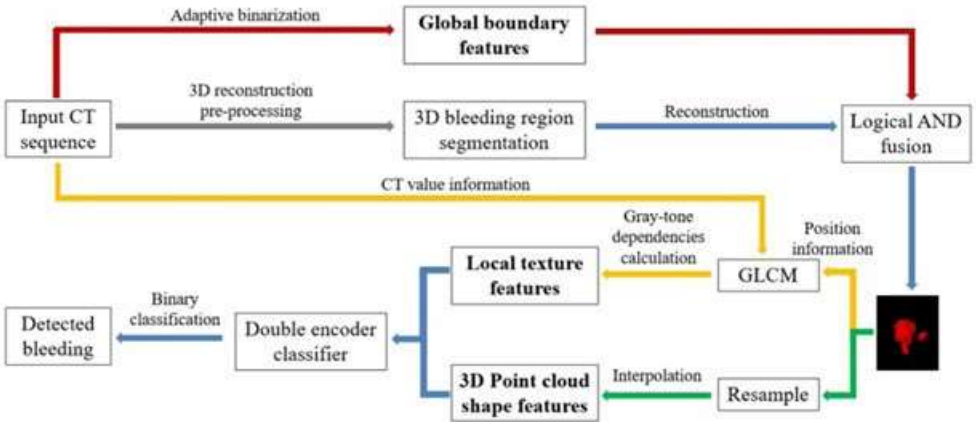

Figure 2: Overview of proposed method, including 3D bleeding region segmentation by deep learning networks and the adaptive threshold method, and the classification approach using fusion features of boundary features from point cloud and texture features from GLCM calculation.

4.2.1. 3D Reconstruction Pre-Processing

In order to solve the problem of insufficient samples with small object detection, an operation of local 3D reconstruction pre-processing is applied to balance the ratio of target to the background. Firstly, CT images are normalized to grayscale images with 8-bit depth by window processing. The window is set to 300 window width and 200 window center, covering most CT value ranges of bleeding regions, as shown in the histogram of bleeding CT values (Figure 3). Since the slice thickness of all cases is 1.0 mm, a linear interpolation to axial images is applied so that the spatial resolution of all CT cases is unified to 1.0 × 1.0 × 1.0 mm.

The proportion of bleeding regions in the entire image is very small. The size of the axial CT image without uniform pixel dimension is 512 × 512 pixels, while the volume of bleeding regions in our data set ranges from 31 to 63656 voxels. Axial CT images with voxel dimension unified are cropped into images with resolution 96 × 96 pixels. Every 32 consecutive images are reconstructed to a 3D volume as the input of the segmentation network, and only volumes with bleeding are selected as training data, as shown in (Figure 4). The large number of whole-body CT images and small size of bleeding regions make the labeling of traumatic bleeding a difficult task, which also results in the small scale of the data set in this study. Therefore, a variety of data augmentation methods are used to increase the number of training samples and perform few-shot learning, including rotation, mirroring, Gaussian noise, and shifting of the dividing grids.

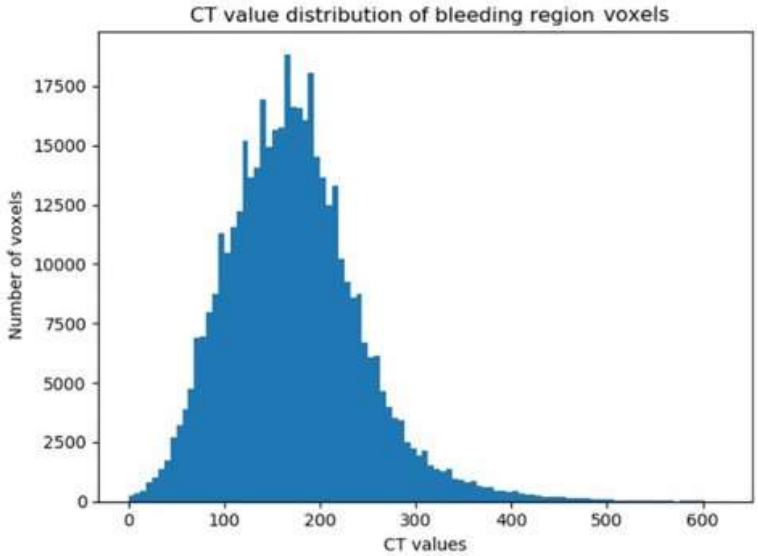

Figure 3: Histogram of bleeding CT values, made according to bleeding labels of all 33 cases. The windowsetting in this study is made to cover the main range of bleeding regions, from CT value 50 to 350.

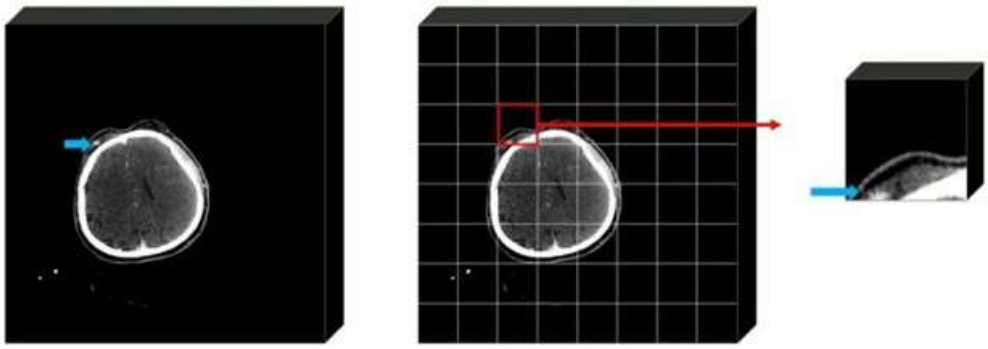

Figure 4: Generation of Segmentation network input volume. Blue arrows show the position of bleeding.

4.2.2. 3D Bleeding Region Segmentation

Compared with traditional detection approaches which generate bounding boxes as output, prediction results of the segmentation network are more conducive to describing multiple smaller targets and being further processed. In this study, the segmentation process is designed as the first stage of the proposed detection method to obtain all bleeding candidate regions. The segmentation network is composed of a generator based on a 3D U-shaped network and a discriminator to form an adversarial network structure for adversarial training as baseline. The generator is divided into two symmetrical parts of the up-sampling path and the down-sampling path. Each part is divided into 4 stages, composed of several convolutional layers, and the corresponding up-sampling and down-sampling stages are connected.

The network structure is selected as the generator to give preliminary segmentation predictions because it has been proven to achieve accurate results for small data sets. The shallow convolutional network depth can also prevent the feature map from being too small to lose information when processing small resolution images. The skip connection structure can combine the output of each convolutional down-sampling stage with the results of the up-sampling stage, so that the features extracted by different convolutional layers can be combined, which is beneficial to improve the performance of the prediction.

The discriminator part is composed of 6 convolutional layers. The network input is divided into two categories: (i) CT image and binarized bleeding ground truth used as positive samples after an element-wise multiply, and (ii) generator prediction element-wise multiplied with corresponding CT image used as negative samples.

The loss function of the segmentation network is shown in (1) and (2). The generator loss consists of two parts, one is the DICE loss [25] obtained based on segmentation prediction and ground truth, and the other is the binary cross entropy loss obtained by element-wise multiplication of the segmentation prediction with the corresponding CT image as input.

The discriminator loss is obtained by adding the positive score and negative score [26], obtained by the same sigmoid cross entropy loss function accepting different inputs: CT image and binarized bleeding ground truth are used as positive samples after an element-wise multiply, and CT image and generator prediction element-wise multiplied with corresponding CT image are used as negative samples, as shown in (1) and (2):

$$LOSS_G(I,G) = LOSS_{DICE}(G,P) + \lambda \mathrm{LOSS}_{SCE}(I \otimes \mathrm{P}, \mathrm{TRUE}), \quad (1)$$

$$LOSS_D(P,G) = LOSS_{SCE}(I \otimes \mathrm{P}, \mathrm{False}) + LOSS_{SCE}(I \otimes \mathrm{G}, \mathrm{True}), \quad (2)$$

where I is the input volume of the generator, G is the corresponding ground truth, P is the prediction of the generator, $Loss_{DICE}$ is the DICE loss function and $Loss_{SCE}$ is the sigmoid cross-entropy loss with logits. $Loss_G$ is the loss of generator, $Loss_D$is the loss of discriminator. $\otimes$ is element-wise multiplication. λ is the combination weight of sigmoid cross-entropy loss summed with segmentation loss. In this study λ is set to 0.01, same with the weight set in [26].

The segmentation result based on the baseline network obtains a certain degree of sensitivity, but there are over-segmentations caused by the connection of tissues with similar CT values and lost bleedings due to too small size. Based on the baseline network structure, several custom improvements for this study are applied to improve the segmentation performance. The structure of the modified generator and discriminator is shown as (Figure 5a and Figure 5b), respectively.

Position Attention Modules (PAM) [27] are added to the connection of the up-sampling and down-sampling paths, and a Channel Attention Module (CAM) [27] is added to the bottom of the down-sampling path, where the number of channels is largest. The detail of the two modules are shown in (Figure 6a and Figure 6b). The PAM module sends the input A into three convolutional layers respectively to obtain three feature maps. The feature map B is multiplied by the reshaped C after reshaping and transposing, and a soft-max operation is performed to obtain the position feature map S:

$$s_{ji} = \frac{exp(B_i \cdot C_j)}{\sum_{i=1}^{N} exp(B_i \cdot C_j)}, \quad (3)$$

where s_{ji} is the influence of the feature of the i-th position on the j-th position. If the two positions happen to belong to the same type of pixels, then s_{ji} will produce a large value, which will highlight the relation between similar features. Two PAM modules are located at the concatenation of the second and fourth stages, be-

cause in all our location experiments of PAM modules, this network structure has obtained the best accuracy. CAM is similar to the PAM module but focuses on the connection between different channels rather than between locations.

The severity of bleeding caused by different degrees of trauma is also different, which causes the size of the bleeding regions to be vary. Therefore, Spatial Pyramid (SP) blocks [28] composed of several convolutional layers are placed in the part where each up-sampling path and down-sampling path are connected. There are a total of 4 SP modules in the segmentation network, and the positions are shown as SP1 to SP4 in (Figure 5). The detailed structure of these blocks is shown in (Figure 7). Convolutional kernels of different sizes allow them to extract features of different scales from feature maps of each down-sampling stage, then superimpose them and feed back to the corresponding up-sampling stage, so that the network is able to detect bleeding regions of different sizes.

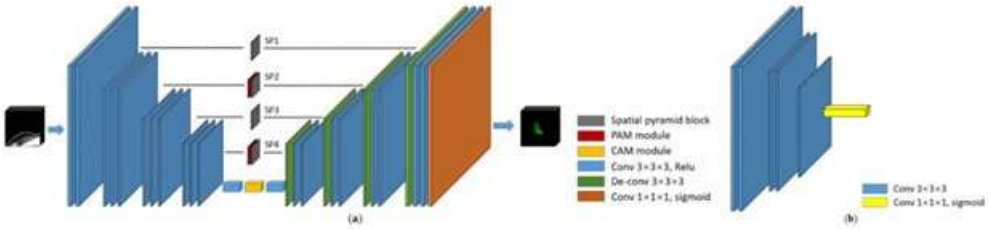

Figure 5: Proposed segmentation network. (a) Structure of the generator part, including the basic 3D U-Netand additional attention modules and spatial pyramid blocks; (b) the discriminator consists of six convolutional layers.

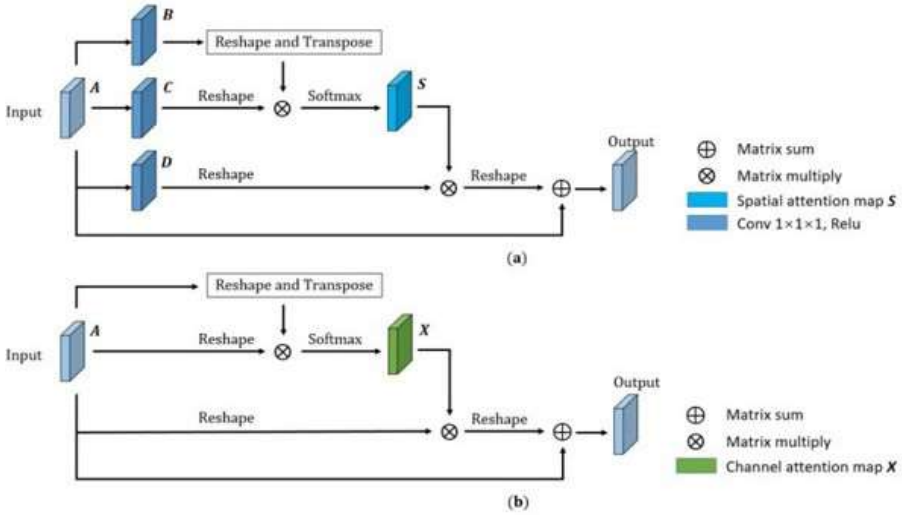

Figure 6: Attention modules. (a) The position module, A is input, B, C and D is outputs of convolutionallayers; (b) the channel attention module

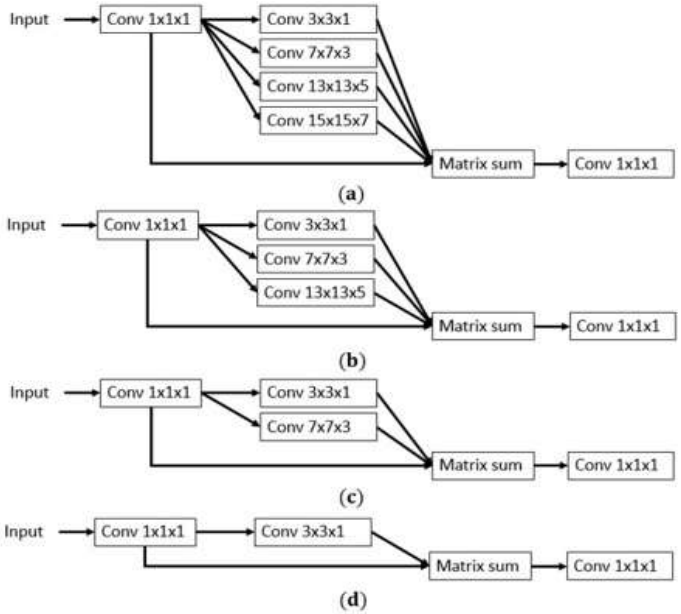

Figure 7: Spatial pyramid blocks, corresponding to four SP blocks in FIGURE 4. (a) SP1, (b) SP2, (c) SP3and (d) SP4.

4.3. Feature Fusion

4.3.1. Global Feature Extraction

The segmentation results had a large number of false positive predictions when extracting as many areas as possible that could be bleeding regions. The reason is that there are a large number of tissues in the human body that are very similar in local features to bleeding areas. While cropping the image to a smaller size in order to detect small targets

and balance the foreground and background ratios, global features of these tissues are lost, resulting in a large number of false positives. At the same time, some tissues are cropped into several parts by dividing grids in pre-processing and sent to the segmentation network as different volumes, accompanied by the loss of global features. In reconstructed segmentation prediction, different parts of the same tissue may be judged as bleeding and non-bleeding respectively, resulting in the generation of irregular boundaries and incomplete shapes, which will distort the bleeding and non-bleeding shapes and have a negative effect on the subsequent classification tasks. One example is shown in (Figure 8).

In response to the above two problems, a feature fusion based on CT thresholds extraction is applied to segmentation result as post-processing. The specific operation is to set thresholds of CT values and extract all regions above the thresholds from the entire DICOM cases, then perform the logical and fusion with the segmentation result and remove all the regions that do not overlap. According to the actual situation of existing bleeding regions in our data set and comments from medical experts, the volume of the bleeding area is usually between 30 and 75,000 voxels, so all regions whose volume is not within this range in the threshold extraction result will be removed.

Considering that the bleeding of different parts will vary due to the difference of the contrast agent distribution, the dual threshold is adopted to cover as many bleeding regions as possible. In this study, based on experiments on existing data sets, the two thresholds were set to 101 and 155. These two values can acquire as many bleeding regions as possible while removing most false positive predictions. The reduction in the number of false positive predictions after post-processing is shown in (Figure 9). Through the fusion of segmentation results that are generated according to local features and the threshold extraction results that include global features, the number of false positive predictions can be largely reduced, and the remaining segmentation results obtain more complete boundary and shape information, as shown in (Figure 8c).

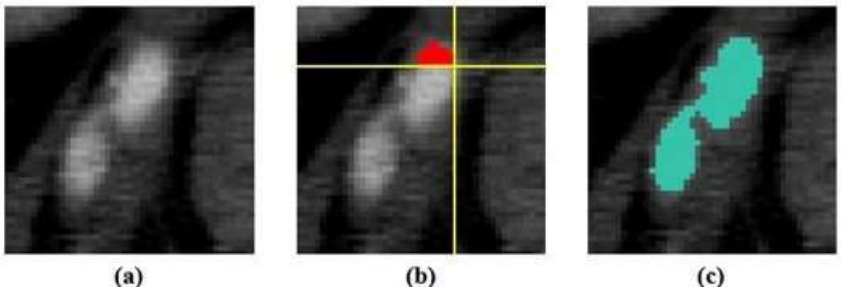

Figure 8: Example of segmentation results affected by dividing grids. (a) The original CT image; red region in (b) is the segmentation result and yellow lines are dividing grids, the straight line boundary overlaps to the dividing grids in the segmentation result can be seen; (c) in the post-processed segmentation result, regions thatare connected to the prediction and have similar CT values are fused.

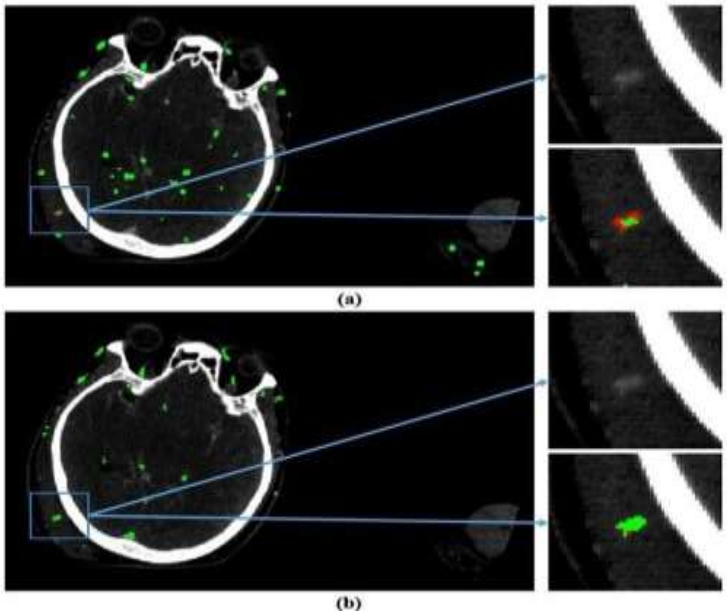

Figure 9: Segmentation result. (a) Before post-processing and (b) after. The number of false positives largely decreases. Green regions are predictions, and the red part in the blue box is the bleeding label. There are fewerfalse positives in (b) than (a), and the true positive prediction in (b) overlaps with the ground truth more than (a).

4.3.2. Point Cloud Generation

After post-processing, the segmentation results need to be converted to extracted features for the classifier. The reason for using point cloud data as input is that the boundary features can be clearly described by the combination of a certain number of points. Unlike other 3D convolutional networks in which the input size is limited by GPU memory, the point cloud data does not need to retain the original size after normalization, and the actual input of the network are coordinates of the point cloud, which saves more computing resources. In this study, every spatially independent volume in the segmentation result after threshing is re-sampled as a point cloud by linear interpolation, composed of 1024 points, and normalized to the range of (-1, 1). A rigid transformation is applied to augment the dataset by rotating point clouds by 45°, 90°, 135°, 180°, 225°, 270° and 315°, respectively. The bleeding region before and after being converted to point cloud is shown as (Figure 10).

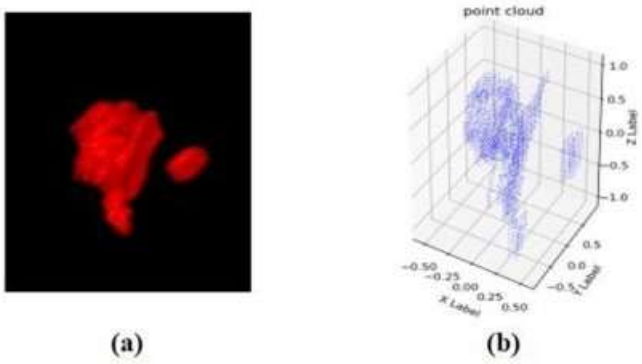

Figure 10: Example of conversion of bleeding regions. (a) The bleeding prediction after 3D reconstruction; (b)the corresponding point cloud. It can be seen that the boundary and shape features are retained.

4.3.3. Texture Features From GLCM

Considering that although the point cloud data can reflect the shape features of the target in detail, other features such as voxel value, density and volume size are also lost due to the change of data type. Therefore, the second encoder takes the feature vector composed of feature values as input to make up for the missing features of the first encoder. The vector as the input of the second encoder contains 16 feature values, 4 of which are obtained from the gray level histogram: Average CT value, Max CT value, Min CT value and number of voxels. The remaining 12 features are calculated from GLCM [29]: Angular Second Moment, Contrast, Correlation, Sum of Squares, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy and Information Measures of Correlation. One example of 2D GLCM is shown in (Figure11). After expanding to the 3D level, each volume can generate GLCMs in 26 directions. Assuming two directions that are 180° opposite are combined into one, the adjacent distance between voxels is 1, and there are 13 GLCMs corresponding to 13 sets of feature values. The final input of the network is the mean of these 13 sets. The range of CT values is very wide, but most of them are air or bones with higher density that are not related to bleeding detection. If the entire CT value range is normalized, the area not related to bleeding will account for the majority and the range of gray value after normalization will largely increase the number of rows and columns of GLCM, making the calculation more cumbersome. Therefore, this study selected the CT value range from -200 to 1200 and quantified it to the range of 0 to 63 to construct a GLCM with 64 rows and columns.

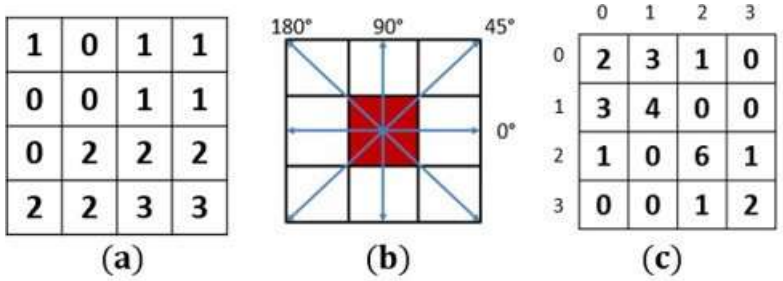

Figure 11: Example of 2D GLCM. Assuming that (a) is a normalized gray scale image, (b) shows the 4directions of each pixel in (a), and (c) is the sum of GLCMs calculated from (a) in the directions of (b).

4.4. Double Classification Network

After post-processing, a certain amount of false positive predictions still exists in the segmentation result. In order to further decrease false positives and reduce time as much as possible in an emergency diagnosis, a classification network is built to classify the segmentation result. All spatially independent 3D volumes in the reconstructed segmentation result are sent to the classifier as input and judged as bleeding or non-bleeding. By adjusting the threshold of probability scores, the classifier can reduce the number of false detections while ensuring that the sensitivity does not decrease.

Details of the classification network are shown in (Figure 12a). The network consists of two encoders and one decoder. Two encoders receive processed segmentation results in different forms as input, extract features and fuse them to conduct the classification. The structure of the first encoder is based on PointNet [30]. The network is designed for point cloud data classification, segmentation and scene segmentation tasks. PointNet encoder takes a point cloud as input, and consists of convolutional layers, fully connected layers and two transform blocks, as shown in (Figure 12). The purpose of the two blocks is to generate a rotation matrix that is supposed to transform the point cloud to produce a positive effect on the classification task. Detail of two transform blocks are shown in Figure 12(b) and Figure 12(c). After accepting the shape features provided by the point cloud data and texture features calculated from GLCMs, two encoders generate their own feature vectors from their respective inputs. The output of the two encoders is fused into one feature vector of length 16 after passing through several fully-connected layers, and concatenated into a 32 length vector. After another fully-connected layer, the final output of the two possibility scores required for binary classification is generated.

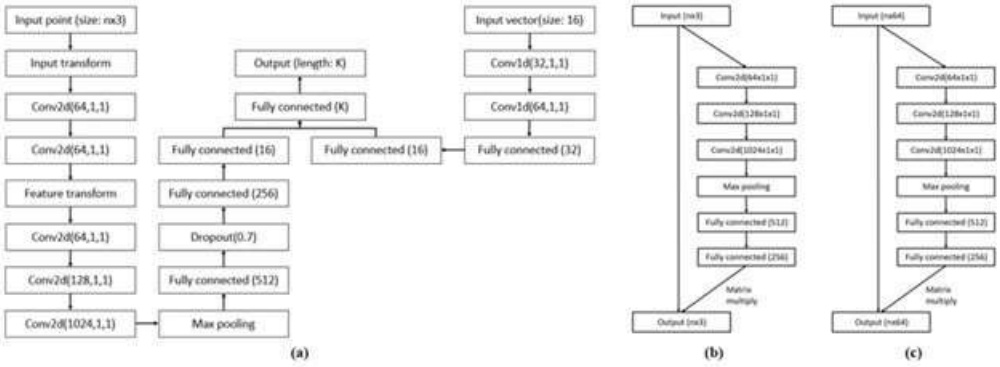

Figure 12: Proposed double encoder classifier. (b) and (c) are the input transform module and feature transformmodule in (a), respectively. *n*is the number of input points. *k* is the number of classification categories. In this study, *n*=1024, *k*=2.

5. Results and Discussion

A series of experiments were conducted to evaluate the performance of the proposed detection scheme. This section presents the results of our experiments and some discussions.

5.1. Experimental Setup

The approach proposed in this study is implemented based on the Keras toolkit under the Windows10 operating system. An NVIDIA GeForce RTX 2080Ti GPU with 11GB memory is used as the hardware to train both the segmentation network and classification network. From the 33 CT cases containing traumatic bleedings provided by the Affiliated Hospital of Chiba University, 30 cases are selected as training data, containing 312 bleedings, and the remaining cases are selected as test data, containing 37 bleedings in total.

5.2. Evaluation Metrics

In order to evaluate the detection result of the proposed method, an evaluation method is proposed after consultation with medical doctors. The evaluation method extracts all spatially independent volumes from the reconstructed detection result and compares them with bleeding labels one by one to determine whether they are true positives or false positives. The method of comparison is to calculate the coordinates of the center of gravity of each bleeding region, and determine whether there is a prediction volume whose distance from its' gravity center to the gravity center of the bleeding label is less than 7.5 mm in all predictions. For a bleeding region, if there is a prediction that meets the above requirements, it is determined that the bleeding is detected; if the prediction of the gravity center does not meet the requirements, it is determined that the bleeding is not detected. For a prediction volume, if there is a bleeding label with a distance from its' gravity center to the gravity center of predictions less than 7.5 mm, then the prediction is judged as a true positive prediction; otherwise, it is judged as a false positive.

$$S = \frac{TP}{TP+FN}, \qquad (4)$$

For the final result, the overall sensitivity, as shown in (4) and average false positive predictions per case and per slice are counted for evaluation, where S is sensitivity, TP is the number of true positive predictions and FN is the number of false negative predictions.

5.3. Segmentation Results

The segmentation results are shown in (Table 2), including the result of the 3D U-Net and proposed segmentation network, as well as the post-processed result of the proposed segmentation network. The original 3D U-Net acquired a certain sensitivity at 83.78%, 1.9542 average false positives per slice. Compared with the above result, the proposed segmentation network in this study acquired a higher sensitivity at 97.29%, but it produces 6.3277 average false positives per slice. We infer that this is attributed to the small features extracted by the small kernel convolutional layer in the spatial pyramid. Therefore, it is necessary to remove some false positives by post-processing before the classification phase. By fusing deep learning and CT value thresholding approaches, about 78.53% of false positives in the segmentation results are removed. However, during post-processing, the sensitivity decreases from 97.29% to 89.19%, and detection of three bleeding regions were lost. The additional false negative bleedings are shown in (Figure 13); detection of two of them are lost because of connection with other tissues that have similar CT values, resulting in the shift of gravity center. Even when using higher CT values as thresholds, it is difficult to separate the bleeding regions from the connected tissues. The remaining bleeding region was not detected because the low CT values cannot be extracted by the current thresholds.

However, using a low threshold to detect the above bleeding region will cause a shift of gravity center of other bleeding regions; moreover, sensitivity is not improved and more false positives will be produced.

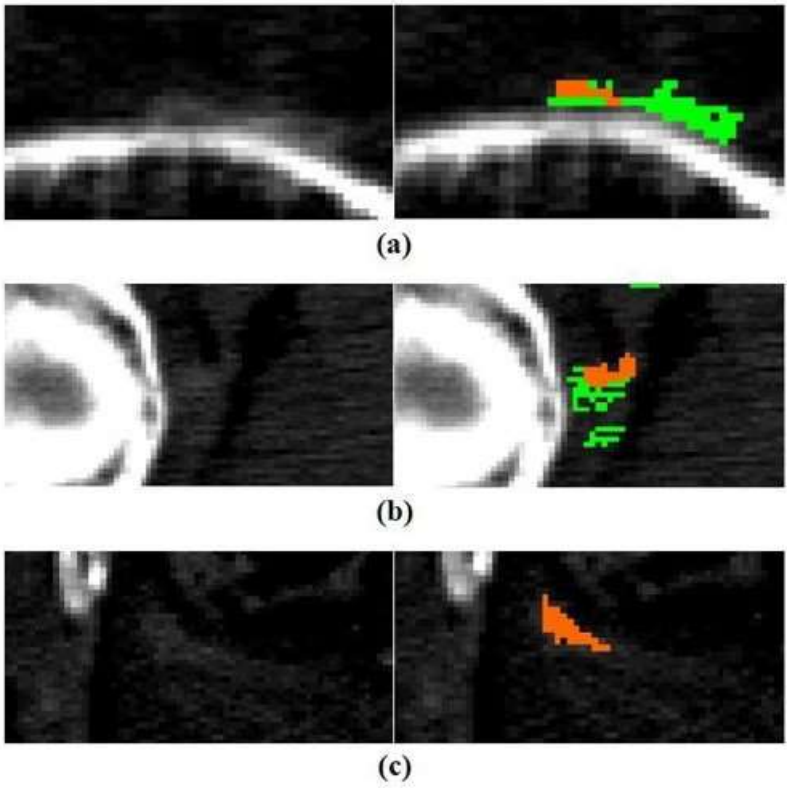

Figure 13: Three additional false negatives after post-processing. Left side are CT images, right side show prediction (in green) and ground truth (in orange). (a) and (b) are caused by gravity centers shifting, and (c) is caused by low CT values of the bleeding region.

Table 2: Segmentation results

Approach	Sensitivity	Average FP per slice
3D U-Net	83.78%	1.9542
Proposed segmentation network	97.29%	6.3277
Proposed segmentation network and post-processing	89.19%	1.3587

5.4. Classification Results

The classification result is shown in (Table 3), including the results of original PointNet and the proposed double encoder classifier. According to the results, the proposed method produces an average of 0.9367 false positives per slice; about 31% of the false positives in the segmentation result are removed.

After using the average value, the proposed method removes from the original PointNet 0.2503 false positive per slice, thus proving that the additional gray level and texture features make up for the missing features in the point cloud data. The fusion of shape features and texture features help the classifier learn the input samples better and achieve a higher performance.

Table 3: Classification results

Approach	Sensitivity	Average FP per slice
PointNet	89.19%	1.1871
Proposed classification network	89.19%	0.9367

5.5. Ablation Study

In order to verify the effectiveness of the modifications that we made, ablation experiments are performed, and results are shown in (Table 4). The baseline is the adversarial network based on 3D U-Net and discriminator. Compared with the result of 3D U-Net in Table 1, the adversarial network decreases the average false positives from 1.9542 to 1.6447 per slice, but sensitivity also decreases to 81.08%. Attention modules increases the sensitivity, and one bleeding region attached to the rib that is lost in the 3D U-Net result only is also detected, as shown in (Figure 14). The additional true positive shows that attention modules improve the ability to distinguish the bleeding from surrounding tissues. After applying spatial pyramid blocks, the sensitivity increases to 97.29% with an average of 6.3277 false positives per slice. The fusion of local features in different scales increases the sensitivity to almost 100%, but the loss of global features causes a large number of false positives. The application of thresholds extraction provides global features and decreases 78.53% of false positives. By fusing the boundary features and texture features, the classifier further decreases the number of false positives.

Table 4: Classification results

Approach	Sensitivity	Average FP per slice
Baseline	81.08%	1.6447
Baseline + attention modules	83.78%	2.2349
Baseline + attention modules + spatial pyramid blocks	97.29%	6.3277
Baseline + attention modules + spatial pyramid blocks + thresholding	89.19%	1.3587
Baseline + attention modules + spatial pyramid blocks + thresholding + classifier	89.19%	0.9367

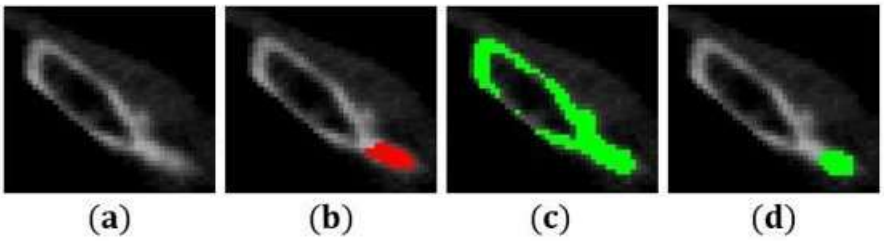

Figure 14: Bleeding detected after applying attention modules. (a) is CT image, (b) ground truth, (c, d) predictions before and after applying attention modules. (c) is evaluated as a false negative because of the gravity center deviation, whereas (d) is evaluated as true positive.

6. Conclusion

In this study, an approach for automatically detecting traumatic bleedings based on the 3D deep convolutional network and feature fusion is proposed to detect bleedings from contrast-enhanced CT images. 8-bit depth gray scale images converted by window processing are cropped and stitched into 96 × 96 × 32 pixel volumes as the input of the segmentation network to resolve the difficulty of detecting small targets. The proposed segmentation network applies spatial pyramid modules to extract and fuse features in different scales, and also applies 3D discriminator and two kinds of attention modules to improve the ability to distinguish bleeding from other body tissues. Through the combination of deep convolutional networks and CT value thresholding extraction, most false positives are removed under the premise of ensuring a certain level of sensitivity. Finally, a classifier that learns shape features, gray level and texture features through double encoders removes about 31% of false positives from post-processed segmentation results. The final result obtains a sensitivity of 89.19%, with an average of 0.9367 false positives per slice. Medical doctors of the Affiliate Hospital of Chiba University give a positive evaluation of our results and consider that the experimental results have certain reference value in image-reading training and actual diagnosis.

For future studies, we plan to gather more clinical data to further improve the performance of our deep convolutional networks. For the data processing, we plan to use different thresholds for different body areas to adapt to different bleeding regions, so as to prevent the sensitivity from decreasing in post-processing. We also plan to consult with doctors to improve the current evaluation methods, for example, adding general evaluation criteria such as IOU or DICE values with a certain weight to improve reliability.

References

1. Spahn DR, Bouillon B, Cerny V, Duranteau J, Filipescu D, Hunt BJ, et al. "The European guideline on management of major bleeding and coagulopathy following trauma: fifth edition," Critical Care. 2019; 23: 98.
2. Behrens AM, Sikorski MJ, Kofinas P. "Hemostatic strategies for traumatic and surgical bleeding," Journal of Biomedical Materials Research. 2014; 102: 4182-94.
3. Löw R, Duber C, Schweden F, Lehmann L, Blum J, Thelen M. "Whole body helical CT for rapid primary diagnosis in severely injured patients," Rofo: Fortschritte auf demGebiete der Rontgenstrahlen und der Nuklearmedizin. 1997; 166: 382-8.
4. Doi K. "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," Computerized Medical Imaging and Graphics. 2007; 31: 198-211.
5. McAuliffe MJ, Lalonde FM, McGarry D, Gandler W, Csaky K, trus BL. "Medical Image Processing, Analysis and Visualization in clinical research," Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS, 2001.
6. Lee JG, Jun S, Cho YW, Lee H, Kim GB Seo JB, et al. "Deep Learning in Medical Imaging: General Overview," Korean J Radiol. 2017; 18: 570-84.
7. K. Suzuki. "Overview of deep learning in medical imaging," Radiological Physics and Technology. 2017; 10: 257-73.
8. Muruti G, Abdul Rahim F, and Ibrahim Z. "A Survey on Anomalies Detection Techniques and Measurement Methods,"2018 IEEE Conference on Application, Information and Network Security (AINS), 2018.
9. Bhadauria HS, Singh A, Dewal ML. "An integrated method for hemorrhage segmentation from brain CT imaging," Computers & Electrical Engineering. 2013; 39: 1527-36.
10. Kumar I, Bhatt C, Singh KU. "Entropy based automatic unsupervised brain intracranial hemorrhage segmentation using CT images," Journal of King Saud University - Computer and Information Sciences, 2020.
11. Kadam M, Dhole A. "Brain Tumor Detection using GLCM with the help of KSVM," International Journal of Engineering and Technical Research (IJETR). 2017; 7: 2454-4698.
12. Zhao Z, Zheng P, Xu S, Wu X. "Object Detection with Deep Learning: A Review," IEEE Transactions on Neural Networks &Learning Systems. 2019; 30: 3212-32.
13. Jiao L, et al., "A Survey of Deep Learning-Based Object Detection," IEEE Access. 2019; 7: 128837-68.
14. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, et al., "Deep Learning for Generic Object Detection: A Survey," International Journal of Computer Vision. 2020; 128: 261-318.
15. Wu X, Sahoo D, Hoi SC. "Recent advances in deep learning for object detection," Neuro computing. 2020; 396: 39-64.
16. Zhou SK, Greenspan H, Davatzikos C, Duncan JS. "A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises," Proceedings of the IEEE Early access. 2020.
17. Rubin DL, "Informatics in radiology: measuring and improving quality in radiology: meeting the challenge with informatics," Radiographics. 2011; 31: 1511–27.
18. Majumdar A Brattain L, Telfer B, Farris C, Scalera J. "Detecting Intracranial Hemorrhage with Deep Learning," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE.2018.
19. Candefjord S, Winges J, Malik AA, Yu Y, Rylander T, Mckelvey T, et al., "Microwave technology for detecting traumatic intracranial bleedings: tests on phantom of subdural hematoma and numerical simulations," Medical & Biological Engineering & Computing. 2017; 55: 1177–88.
20. Ren S, He K, Girshick R, and Sun J. "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017; 39: 1137-49.
21. Redmon J, Divvala S, Girshick R, and Farhadi A. "You only look once: Unified, real-time object detection," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016.
22. Redmon J, Farhadi A. "YOLO9000: Better, faster, stronger.", In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA. 2017.
23. Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv. 2018.
24. Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," MICCAI 2016, LNCS 9901. 2016; 424-32.
25. Sudre CH, Li W, Vercauteren T,Ourselin S, Cardoso MJ. "Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations," DLMIA 2017, ML-CDS. 2017; 240-8.
26. Hunge WC, Tsai YH, Liou YT, Lin YY, Yang MH. "Adversarial Learning for Semi-Supervised Semantic Segmentation," British Machine Vision Conference (BMVC), 2018.
27. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, et al. "Dual Attention Network for Scene Segmentation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 3146-54.
28. Jia H, Xia Y, Song Y, Zhang D, Huag H, Zhang Y, et al. "3D APA-Net: 3D Adversarial Pyramid Anisotropic Convolutional Network for Prostate Segmentation in MR Images," IEEE Transactions on Medical Imaging. 2020; 39: 447-57.
29. Haralick R.M, Shanmugam K, Dinstein I, "Textural Features for Image Classification," IEEE Transactions on Systems,Man and Cybenetics.1973; SCM-3: 610-21.
30. Qi CR, Su H, Mo K, Guibas LJ, Guibas LJ. "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," The IEEE Conference on Computer Vision and Pattern Recognition. 2017.