80380_Auto_EditedC.docx

Retrospective Study

Convolutional neural network-based segmentation network applied to image recognition of angiodysplasias lesion under capsule endoscopy

Ye C *et al.* Distortion recognition of small intestinal vessels

Ye Chu, Fang Huang, Min Gao, Duo-Wu Zou, Jie Zhong, Wei Wu, Qi Wang, Xiao-Nan Shen, Ting-Ting Gong, Yuan-Yi Li, Li-Fu Wang

Abstract

BACKGROUND

Small intestinal vascular malformations (angiodysplasias) are common causes of small intestinal bleeding[1,2]. While capsule endoscopy has become the primary diagnostic method for angiodysplasia, manual reading of the entire gastrointestinal tract is time-consuming and requires a heavy workload, which affects the accuracy of diagnosis.

AIM

To evaluate whether artificial intelligence can assist the diagnosis and increase the detection rate of angiodysplasias in the small intestine, achieve automatic disease detection, and shorten the capsule endoscopy (CE) reading time.

METHODS

A convolutional neural network semantic segmentation model with a feature fusion method, which automatically recognizes the category of vascular dysplasia under CE and draws the lesion contour, thus improving the efficiency and accuracy of identifying small

intestinal vascular malformation lesions, was proposed. Resnet-50 was used as the skeleton network to design the fusion mechanism, fuse the shallow and depth features, and classify the images at the pixel level to achieve the segmentation and recognition of vascular dysplasia. The training set and test set were constructed and compared with PSPNet, Deeplab3+, and UperNet.

RESULTS
The test set constructed in the study achieved satisfactory results, where pixel accuracy was 99%, mean intersection over union was 0.69, negative predictive value was 98.74%, and positive predictive value was 94.27%. The model parameter was 46.38 M, the float calculation was 467.2 G, and the time length to segment and recognize a picture was 0.6 s.

CONCLUSION
Constructing a segmentation network based on deep learning to segment and recognize angiodysplasias lesions is an effective and feasible method for diagnosing angiodysplasias lesions.

Key Words: Artificial intelligence; Image segmentation; Capsule endoscopy; Angiodysplasias

Chu Y, Huang F, Gao M, Zou DW, Zhong J, Wu W, Wang Q, Shen XN, Gong TT, Li YY, Wang LS. Convolutional neural network-based segmentation network applied to image recognition of angiodysplasias lesion under capsule endoscopy. *World J Gastroenterol* 2022; In press

Core Tip: Small intestinal vascular malformation (vascular dysplasia) is a common cause of small intestinal bleeding. Herein, we proposed a semantic recognition segmentation network to recognize small intestinal vascular malformation lesions. This method can

assist doctors in identifying lesions, improving the detection rate of intestinal vascular dysplasia, realizing automatic disease detection, and shortening the capsule endoscopy reading time.

INTRODUCTION

Angiodysplasias are degenerative lesions that manifest as abnormalities of arteries, veins, or capillaries of the original normal blood vessels. Occasionally, the term angiodysplasias include various synonymous disease concepts, such as angioectasia (AE), Dieulafoy's lesion (DL), and arteriovenous malformation. According to the Yano-Yamamoto classification, small bowel vascular lesions are classified into four types under endoscopy[3]. AE includes small erythemas and can be defined as type 1a: punctuate (< 1 mm), or type 1b: patchy (a few mm). They are characterized by thin, dilated, and tortuous veins lacking smooth muscle layers, which explain their weakness and tendency to bleed. Typically, DLs consist of small mucosal defects and can be classified as type 2a: punctuate lesions with pulsatile bleeding or type 2b: pulsatile red protrusions without surrounding venous dilatation[4]. Some arteriovenous malformations and pulsatile red protrusions with dilated peripheral veins are defined as type 3. Congenital intestinal arteriovenous malformations manifest as polypoid or cluster type[5,6] and are classified as type 4. Nevertheless, the Yano-Yamamoto classification cannot fully reflect the histopathological findings.

Capsule endoscopy (CE) is a painless and well-tolerated approach that can achieve complete visualization of the small intestine[7]. It captures images for > 8 h[8]. Previous studies have demonstrated the probability of CE diagnosis of angiodysplasias was 30%-70%, and > 50% of obscure gastrointestinal bleeding patients have angiodysplasias[9-11]. The detection rate of CE was reported to be higher than other diagnostic methods, such as small bowel computed tomography, mesenteric angiography, and enteroscopy. Therefore, using CE as a first-line inspection tool for the diagnosis of angiodysplasias is recommended[12]. Nonetheless, CE has some limitations, and only 69% of angiodysplasias can be diagnosed by gastroenterologists[13]. Less relevant lesions, such as erosions or tiny red spots, are regarded as negative results; however, distinguishing highly relevant lesions from less relevant lesions could be challenging. In addition, the diagnostic efficiency of CE decreases when the presence of bile pigments, food residues, or bubbles affects the observation of the intestinal mucosa. The doctor's manual reading of the entire

gastrointestinal tract is time-consuming, and the heavy workload affects the accuracy of the diagnosis. Therefore, making diagnosis of angiodysplasias solely based on CE is challenging.

The detection rate of angiodysplasias in the small intestine can be increased by using artificial intelligence (AI) to assess the effect of automatic diagnosis, which has been successfully applied for the recognition and diagnosis of gastrointestinal endoscopic images[14]. AI assists in the recognition and diagnosis of CE images, eliminates errors in manual reading, reduces the workload of doctors, and improves diagnosis efficiency. The clinical application of AI-based deep learning technology in wireless CE has been a research focus, which has gained increasing interest in the past two years[15-32]. Several studies[15,23-26] have used deep learning to identify ulcers from CE data. Pogorelov *et al*[27] used the color texture features to detect small intestinal bleeding in CE data. Blanes-Vidal *et al*[28] constructed a classification network to identify intestinal polyp lesions in CE data. Kundu *et al*[29] and Hajabdollahi *et al*[30] identified small bowel bleeding in CE data using a classification neural network. Obscure gastrointestinal bleeding is the main indication for small intestinal CE, and the potential risk of bleeding from vascular malformations is high[14]. Therefore, we focused on AI-assisted recognition technology for angiodysplasias in the present study. Hitherto, there are few semantic segmentation networks based on deep learning to segment and recognize angiodysplasias lesions in CE, which prompted us to introduce a segmentation model in the study. Compared with the classification model and target detection model in deep learning, the segmentation model based on deep learning can more accurately locate the focus of small intestinal vascular malformation, better assist doctors in diagnosing small intestinal vascular malformation, and improve the accuracy and efficiency of doctors' diagnosis.

Currently, significant progress has been made in semantic segmentation in the field of deep learning. To the best of our knowledge, this is the first paper that proposed using a semantic segmentation network to solve the pixel-level small intestinal vascular malformation focus recognition and location. Resnet-50 was used as the skeleton network, and the fusion mechanism based on shallow features and deep features was

introduced so that the segmentation model could accurately locate the location and category of lesions. Shallow features can perceive the texture details of lesions, while deep features can perceive the semantic information between lesions. By combining these two features to segment the image, the phenomenon where the lesion area is divided into uncorrelated small areas is reduced, the pixel accuracy is improved, and the missed detection rate of the lesion is reduced. This paper introduced the proposed network structure in detail and compared three common segmentation models, *i.e.*, pspnet[31], deeplabv3+[32], and supernet[33]. The obtained results confirmed that the model proposed in this paper had high-performance indicators.

MATERIALS AND METHODS

ResNet was introduced in 2015 and won first place in the classification task of the ImageNet competition on account of being "simple and practical". Afterward, many methods, which were based on ResNet50 or ResNet101, have been widely used in detection, segmentation, recognition, and other fields. This method makes a reference (X) for the input of each layer, learning to form residual functions rather than learning some functions without reference (X). This residual function is easier to optimize and can greatly deepen the number of network layers. Moreover, the extracted image features have strong robustness. ResNet50 is faster than ResNet100. Therefore, ResNet50 is selected as the skeleton network of the semantic segmentation network in this paper. Based on the fusion of shallow and deep features, Resnet-50 was used as the skeleton network to construct an improved CNN segmentation network model that automatically recognizes the type of angiodysplasias under CE and draws the outline of the lesion in the study. The present study aimed to assist doctors in diagnosing angiodysplasias lesions with CE.

The model proposed in this study was composed of three sub-units, *i.e.*, down-sampling, up-sampling, and classifier. CE small intestine data were used as input in the module, and the final output was image lesion category information and lesion boundary information.

Research data set

In order to train and evaluate the segmentation model, 378 patients with angiodysplasias who underwent OMOM CE (China Chongqing Kingsoft Technology Co., Ltd) at the Ruijin Hospital between January 2014 and December 2020 were recruited in this study. The sampling frequency of OMOM capsules of 2fps, the working time of > 12 h, and the apex field of view of 150° were used to diagnose the patients. A total of 12403 pictures were identified with an image resolution of 256 × 240. The patient data were anonymized, any personal identification information was omitted, and examination information (such as examination date and patient name) was deleted from the original image. All patients provided written informed consent, and the ethics committee approved the study [the certification number was (2017) provisional ethics review No. 138]. The annotated data were marked by an experienced endoscopy group that included three experts from Ruijin Hospital Affiliated to Shanghai Jiao Tong University. The average age of the experts was 35 years, and their average CE reading experience was 5 years, with an average of 150 CE cases each year. The five types of lesions of vascular malformation were annotated, and 12403 image data and 12403 annotated mask image data were generated. The data sample map is shown in Figure 1.

This project used the image data of 178 cases as the training set and the remaining 200 cases as the test set. The training set was divided into training and verification data at a ratio of 7:3 during the training process. The test set contained 1500 images without lesions and 1500 images with lesions. The training set and test set image data are summarized in Table 1.

Data preprocessing

The training data were preprocessed to meet the requirements of the deep learning model. The preprocessing steps of the model constructed in the study were as follows: (1) resizing the image to 256 × 240 × 3; (2) using enhancement methods (rotation, flip, and tilt) on the resized image; and (3) normalizing all images. In order to train a deep learning

model, the dataset was split. The dataset image was randomly divided into two parts: 70% for training and 30% for verification.

Segmentation network details

The network structure proposed in this study is shown in Figure 2. The construction of the network model was inspired by the UperNet model. ResNet-50 was used as the skeleton network. The fusion mechanism of shallow features and deep features were introduced. Subsequently, the feature with the same size as the original image was obtained through the down-sampling operation. Finally, the classifier was connected to realize the pixel-level segmentation task of the image.

Based on the new semantic segmentation recognition network framework, a single end-to-end network could be trained to capture and analyze the semantic information of the CE small intestine data. In order to fuse the shallow features and deep feature information, the last feature mapping set output by each stage in ResNet was expressed as C1, C2, C3, and C4, and the two-by-two fusion of features were utilized as down-sampling operation input, where the down-sampling rates were 4, 8, 16, and 32, respectively. The texture features of the lesion were captured at the highest layer, and the pixel-level segmentation of the lesion was completed based on the lowest layer features.

The last down-sampling operation generated a feature map with the same resolution as the original image, with a size of 256 × 240. After the feature was operated by Flatten, a classifier composed of a fully connected layer was connected to complete the segmentation and recognition tasks of the capsule data.

In order to assess the fusion of features of different scales, bilinear interpolation was used to adjust them according to the size, after which a non-evolutionary layer was applied to fuse the features of different levels and reduce the channel size. All non-classifier convolutional layers underwent batch normalization Relu operations after output. The learning rate of the current iteration was equivalent to the initial learning rate multiplied by $(1\text{-iter}/\text{max-iter\_size})^{power}$, and the initial learning rate and power were set to 0.02 and 0.9, respectively.

RESULTS

Evaluation index

The performance of the segmentation model of angiodysplasias lesions in CE was evaluated based on the following indicators: Positive predictive value (PPV), negative predictive value (NPV), mean intersection over union (mIOU), and pixel accuracy (PA). PPV and NPV were calculated using formulae 1 and 2, respectively.

Where true positive (TP) and true negative (TN) are the true number of positive samples and the true number of negative samples, respectively; FP and FN are false positives and false negatives, respectively. IOU and mIOU calculation formulae are shown as formulae 3 and 4, respectively.

$$PPV = \frac{TP}{TP + FP} \qquad (1)$$

$$NPV = \frac{TN}{TN + FN} \qquad (2)$$

$$IOU = \frac{Target\ area \cap Prediction\ area}{Target\ area \cup Prediction\ area} \qquad (3)$$

$$MIOU = \frac{1}{k + 1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij} + \sum_{j=0}^{k}P_{ji} - P_{ii}} \qquad (4)$$

Supposedly, there were K+1 categories (including an empty category or background) in semantic segmentation, which indicated that class i is predicted as i, and class j is predicted as j. The PA is calculated by formula 5:

$$PA = \frac{\sum_{i=0}^{k}P_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}P_{ij}} \qquad (5)$$

Experimental design

Python 3 is a good deep-learning programming language that supports multiple deep-learning frameworks. The model was implemented using Python3 and Torch framework. The training server has a graphics processing unit. All images were first passed to the image data generation class in Pytorch, and the preprocessing operations were

performed, including enhancement, resize, and normalization operations. Then, the generated images were sent to the model to start the training. The layers in the backbone network ResNet-50 used pre-trained weights on ImageNet. An optimizer (SGD) was used to train the model, after which a weight decay of 0.0001 and a momentum of 0.9 were applied. Each model ran approximately 25000 Epochs; each Epoch iterated eight times, and the batch size was 8. In the model training process, the loss change, PA index, and MIOU changes were detected (Figure 3).

Comparison results of multiple models

On a test set consisting of 3000 image data, the following test indicators were compared on the four models: PPV, NPV, mean IOU, PA, parameter quantity, float calculation quantity, and duration. The results are shown in Table 2.

Based on the method of fusion of shallow and deep features, the CNN segmentation network model was improved and optimized, and the segmentation and recognition of five types of angiodysplasias lesions, *i.e.,* blue branch, blue cluster, red branch, red cluster, and red spider nevus, were realized. This method fully uses the shallow and deep features extracted from the skeleton network to perceive the global information and lesion texture information of the small intestine capsule image data as a whole. Thus, it significantly improves the PPV and NPV of the segmentation model in the angiodysplasias lesion image. In order to obtain the highest PPV, the NPV has to be the highest. The unified perception of the global and local information of the small intestine capsule data was completed through a CNN network, which reduced the number of network model parameters, the number of float calculations, and the inference time of the deep learning model. Furthermore, a comparative experiment was designed and compared to the current advanced segmentation network models: PSPNet, DeeplabV3+, and UperNet. Our model showed that the NPV reached the highest 98.74% when the PPV was the highest.

The comparison of the segmentation and recognition effects of the four models on the vascular aberration lesions of the CE small intestine data is shown in Figure 4. The model proposed in the study was similar to that of the expert's annotation results.

Compared with relevant literature, Leenhardt *et al*[19] applied technology for segmentation, achieving the highest level of lesion detection, with an NPV value of 96%. However, the algorithm presented in this paper had some advantages in the test set. Also, our NPV value was 98%.

DISCUSSION

The classification network and the target detection network are the mainstream network structure that combines the deep learning model and the CE diagnosis method. In the present study, we introduced the segmentation network in deep learning, segmented and identified the angiodysplasias lesions, and completed the pixel-level segmentation task of the angiodysplasias lesions. The semantic segmentation network model had clinical practicality application as assessed using the training and test sets in comparative experiments.

The segmentation networks have been obviously developed in the field of deep learning. PSPNet uses the prior knowledge of the global feature layer to understand the semantics of various scenes, combined with the deep supervision loss to develop an effective optimization strategy on ResNet and embed difficult-to-analyze scene information features into the functional connectivity networks prediction framework to establish a pyramid. The pooling module aggregates the contextual information in different regions and improves the ability to obtain global information. This system was used for scene analysis and semantic segmentation and was 83% accurate on the COCO data set. The DeepLabV3+ model was based on an encoder-decoder structure, which improved the accuracy and saved the inference time; an accuracy rate of 89% was obtained in the COCO dataset. UperNet used unified perception analysis to build a network with a hierarchical structure to ensure that multiple levels were resolved at visual concepts, learn the differentiated data in various image datasets, achieve joint

reasoning, and explore the rich visual knowledge in the images. Finally, 79.98%-pixel accuracy was obtained on the ADE20K data set. UperNet used a unified perception analysis module from scenes, objects, parts, materials, and textures to simultaneously analyze the multilevel visual concepts of images, such that many objects could be segmented and recognized, and the rate of missed objects could be reduced. The CE small intestine image data has a simple scene and fewer semantic levels. The use of large segmentation network models would cause over-fitting in training and high computational complexity. This study was inspired by UperNet and optimized basic CNN segmentation network, which led to the creation of a network model suitable for the segmentation and recognition of angiodysplasias with CE.

On the other hand, a case-based dataset encompassing typical vascular malformation images, atypical angiodysplasias images, and normal images was constructed, including pictures with poor intestinal cleanliness. According to the color and morphology of the angiodysplasias lesions in the cases, the five types of angiodysplasias lesions were summarized as blue branched, blue cluster, red branched, red cluster, and red spider nevus. The dataset constructed in this study verified the clinical applicability of the semantic segmentation model. Thus, the dataset was essential in diagnosing CE small bowel vascular malformation based on the deep learning model.

CONCLUSION

The deep learning model constructed in this study showed high PPV and NPV for the segmentation and recognition of angiodysplasias lesions. In the future, it could be used to assist capsule endoscopists in the real-time diagnosis of angiodysplasias lesions. Deep learning does not require prior knowledge, as it can directly learn the most predictive features from image data, as well as segment and recognize the image. The larger the amount of data, the higher the advantages of deep learning and the higher the recognition accuracy. AI facilitates grassroots' CE to obtain the same diagnosis effect as senior experts. However, the current uneven distribution of medical resources and the technical level of grassroots CE are the driving forces for the development of AI. In conclusion, the

segmentation model based on deep learning can assist doctors in identifying the lesions of small intestinal vascular malformations.

ARTICLE HIGHLIGHTS

Research background

Small intestinal vascular malformations (angiodysplasias) commonly cause small intestinal bleeding. Therefore, capsule endoscopy has become the primary diagnostic method for angiodysplasias. Nevertheless, manual reading of the entire gastrointestinal tract is a time-consuming heavy workload, which affects the accuracy of diagnosis.

Research motivation

The doctor's manual reading of the entire gastrointestinal tract is time-consuming, and the heavy workload affects the accuracy of the diagnosis. Also, significant progress has been made in semantic segmentation in the field of deep learning.

Research objectives

This study aimed to assist in the diagnosis and increase the detection rate of angiodysplasias in the small intestine, achieve automatic disease detection, and shorten the capsule endoscopy (CE) reading time.

Research methods

A convolutional neural network semantic segmentation model with feature fusion automatically recognizes the category of vascular dysplasia under CE and draws the lesion contour, thus improving the efficiency and accuracy of identifying small intestinal vascular malformation lesions, was proposed.

Research results

The test set constructed in the study achieved satisfactory results: pixel accuracy was 99%, mean intersection over union was 0.69, negative predictive value was 98.74%, and

positive predictive value was 94.27%. The model parameter was 46.38 M, the float calculation was 467.2 G, and the time needed to segment and recognize a picture was 0.6 s.

Research conclusions

Constructing a segmentation network based on deep learning to segment and recognize angiodysplasias lesions is an effective and feasible method for diagnosing angiodysplasias lesions.

Research perspectives

The model detects the small intestinal malformation lesions in the capsule endoscopy image data and draws the lesion area through segmentation.

80380_Auto_EditedC.docx

11%

SIMILARITY INDEX

PRIMARY SOURCES

1   www.ncbi.nlm.nih.gov
Internet   130 words — 3%

2   mdpi-res.com
Internet   42 words — 1%

3   "Computer Vision – ECCV 2018", Springer Science and Business Media LLC, 2018
Crossref   31 words — 1%

4   Qinghua Han, Xuan Liu, Jie Xu. "Detection and Location of Steel Structure Surface Cracks Based on Unmanned Aerial Vehicle Images", Journal of Building Engineering, 2022
Crossref   22 words — 1%

5   Li Deng, Qian Chen, Yuanhua He, Xiubao Sui, Quanyi Liu, Lin Hu. "Fire detection with infrared images using cascaded neural network", Journal of Algorithms & Computational Technology, 2019
Crossref   21 words — 1%

6   link.springer.com
Internet   21 words — 1%

7   Xing Bai, Jun Zhou. "Parallel global convolutional network for semantic image segmentation", IET Image Processing, 2020
Crossref   18 words — < 1%

8   www.hindawi.com
Internet
18 words — < 1%

9   Mingchao Shao, Chenwei Nie, Minghan Cheng, Xun Yu, Yi Bai, Bo Ming, Hongli Song, Xiuliang Jin. "Quantifying effect of tassels on near-ground maize canopy RGB images using deep learning segmentation algorithm", Precision Agriculture, 2021
Crossref
13 words — < 1%

10   jasn.asnjournals.org
Internet
13 words — < 1%

11   N. Scales. "Evaluation of the logarithmic-sensitivity index as a neural network stopping criterion for rare outcomes", 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine 2003 ITAB-03, 2003
Crossref
12 words — < 1%

12   f6publishing.blob.core.windows.net
Internet
11 words — < 1%

13   ljkzedo.ba
Internet
9 words — < 1%

14   Huihui Huo, Yaochen Li, Chuan Wu, Xiao Wu, Zhiqiang Tian, Yuehu Liu. "Semantic Segmentation and Scene Reconstruction for Traffic Simulation Using CNN", 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), 2019
Crossref
8 words — < 1%

15   Xin Hu, Hua Ouyang, Yang Yin. "Image Recognition based on Convolution Neural Network", 2020 IEEE
8 words — < 1%

9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2020
Crossref

16   Xing Bai, Jun Zhou. "Efficient Semantic Segmentation Using Multi-Path Decoder", Applied Sciences, 2020
Crossref

7 words — < 1%

17   "Artificial Neural Networks and Machine Learning – ICANN 2019: Image Processing", Springer Science and Business Media LLC, 2019
Crossref

6 words — < 1%

18   Yuan Wang, Abdugheni Abliz, Hongbing Ma, Li Liu, Alishir Kurban, Umut Halik, Matti Pietikainen, Wenjuan Wang. "Hyperspectral Estimation of Soil Copper Concentration Based on Improved TabNet Model in the Eastern Junggar Coalfield", IEEE Transactions on Geoscience and Remote Sensing, 2022
Crossref

6 words — < 1%