Segmentation and classification of renal tumors based on convolutional neural network

Zheng Gong and Liang Kan

Department of Urinary Surgery, Shengjing Hospital of China Medical University, Shenyang, China; Department of Geriatrics, Shengjing Hospital of China Medical University, Shenyang, China

ABSTRACT

Kidney tumors are the second most frequent urology tumors. They are of many types, mostly existing as malignant tumors. In order to improve the accuracy of segmentation and classification of kidney tumors, this paper proposes to build a model of simultaneous segmentation and classification of kidney tumors based on convolutional neural networks to assist medical experts in diagnosis. A two-task neural network 2D SCNet is proposed by combining kidney tumor segmentation and classification. Based on our proposed framework, classification can feed back the global contextual information of the network, and segmentation can make the network focus on local features and regions of interest (ROI). Both tasks jointly promote network feature learning and both increase each other’s prior information. The combination of segmentation and classification of 2D SCNet can achieve an accuracy rate of 99.5% in both benign and malignant classification. The results of the ‘2D SCNet + three-label’ segmentation reached Dice coefficients of 0.946 and 0.846, respectively. Compared with PSPNet, our network kidney and tumor segmentation results are improved by 4.9% and 5.0%, respectively, which shows that the addition of classification module is beneficial to the learning of segmentation network. From the cross-validation results, we can see that 2D SCNet and the two-step segmentation strategy can obtain better results in the segmentation and classification tasks of kidney tumors. The base network of 2D SCNet can extract networks for any feature. This paper compares ResNet50+ PPM and DenseNet as the results of segmentation and classification of the base network. ResNet50+ PPM obtains better results. 2D SCNet can help to segment and examine kidney tumors more efficiently and accurately.

1. Introduction

Radical nephrectomy was once the main method of kidney tumor resection. Because this method will seriously damage the kidney function of the patient, laparoscopic partial nephrectomy (LPN) has gradually become the preferred treatment option. In the past 20 years, in most cases of local renal cell carcinoma, nephron-sparing surgery has become an alternative method for oncological radical nephrectomy (Kutikov & Uzzo, 2009). LPN can protect the patient’s kidneys as much as possible.

In the case of function, the removal of cancerous sites is the gold standard for the treatment of 4 cm tumor patients. Some studies have shown that some patients with kidney tumors within the range of 4.1 cm-7 cm can also perform LPN. In this case, the surgeon’s judgment on the surgery will play a decisive role in the type of surgery (conservative treatment or LPN) to be performed (Ficarra et al., 2009). At this time, the type, shape, location, etc. of the tumor will have an impact on the surgical method and postoperative recovery. The internal tumor of kidney tumors is complex, the texture and position are variable, and the degree of medical imaging enhancement is different.

If only relying on radiology imaging, even experienced urologists cannot accurately characterize certain tumors and need to use pathological reports to diagnose. At present, the research of relevant workers focuses on the location and grading of renal tumors by means of radiography. CT and MRI imaging, as a mature and widely used imaging method, have become the main data form of kidney tumor related research. Predicting the benign and malignant, shape and location of kidney tumors based on CT or MRI imaging is of great clinical significance.

Computer-aided diagnosis is increasingly valued by experts and related researchers. The digitization of images and the computational power of computer exponential growth provide researchers with more opportunities and room for development. The appeal of having a computer that can be executed repeatedly, efficiently and with clear tasks is obvious. The computer will perform a given task consistently and tirelessly. Relatively, complicated and lengthy work is a burden for doctors. They could have spent more energy on treating patients, so humans hope to be able to use the computer system to share part of the doctor’s work.

ARTICLE HISTORY
Received 21 June 2021
Accepted 19 September 2021

KEYWORDS
Convolutional neural network; dual task; kidney tumor; image segmentation; image classification

CONTACT Liang Kan, kanliang31cmu@163.com, Department of Geriatrics, Shengjing Hospital of China Medical University, Shenyang, China

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
The application of machine learning in the medical field mainly includes computer-assisted diagnosis, medical imaging, pathological image analysis, etc. (Hinton et al., 2006) to provide clinicians with fast, more accurate, and repeatable examination results analysis to assist diagnosis. In recent years, with the emergence of graphics processors (GPUs), the development of mathematical models, the availability of big data, and the availability of low-cost sensors, a subset of machine learning is deep learning. Deep learning technology has gradually become the main technology of computer vision (Mahony et al., 2019).

From the 1970s to the 1990s, some researchers have used simple image processing algorithms to obtain information and deal with some problems such as detection and segmentation (Dominik Müller, 2019). However, these algorithms are mostly based on the processing of medical images, which is greatly limited by the equipment and imaging environment, and does not have the value and conditions for a wide range of clinical applications.

Convolutional neural network (CNN) is one of the classic models of deep learning. It is used in image recognition and image classification for great success, and provides new research directions for medical imaging and pathological histology. A typical CNN network construction consists of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer (Brattain et al., 2018): the input layer preprocesses the original image data to generate an output vector and input it to the convolutional layer; the convolutional layer performs feature learning to detect local features in all positions of the input image. The main function of the pooling layer is to downsample, remove unimportant samples, simplify the output of the convolution layer, reduce the number of network parameters, and reduce the complexity of the model; the fully connected layer will accumulate and pooled multiple features are fully connected, and these features are used to achieve the final classification. Different from traditional ML classification methods (such as support vector machine, random forest and other classifiers for classification), CNN-based image classification does not require manual extraction of image data features, which can be automatically recognized by machine to identify features, similar to biological vision.

The features recognized by the system, on the one hand, avoid the time-consuming and labor-intensive definition, extraction, screening, and combination of features, on the other hand, they can mine the hidden information in the data more fully, and so higher classification accuracy can be obtained (Brattain et al., 2018; Litjes et al., 2017; Predictive Inference, 1995). CNN based on deep learning framework model is used for segmentation and classification of kidney tumor images (Doi, 2006). CNN has good performance in image detection and classification, and also has gratifying achievements in the medical field. Its short-term, accurate, and repeatable image analysis can assist radiologists and pathologists to complete the diagnosis, and has high clinical value.

Renal cell carcinoma accounts for approximately 3% of adult solid cancers and has become the 14th most common cancer in the world, with more than 400,000 new cases of kidney cancer each year. Despite mature modern imaging techniques and early detection mechanisms, one-third of patients still have metastatic disease, and more than half have local disease (Kirkali et al., 2001). The most common subtypes of renal cell carcinoma are clear cell carcinoma, renal smooth muscle lipoma, and renal chromophobe cell carcinoma, accounting for about 70%–80%, 14%–17%, and 4%–8% of renal cell carcinoma, respectively (Smith-Bindman et al., 2008). Renal clear cell carcinoma is the main subtype.

2. Materials and methods

2.1. Convolutional neural network architecture

The principle diagram of the convolutional neural network model architecture in this paper is shown in Figure 1. Input the unprocessed kidney CT image sequence. In order to reduce the impact of surrounding tissue on segmentation and improve the accuracy of segmentation, the first step of the algorithm is to locate the location of the kidney tumor and calculate the tumor region of interest (ROI).

The convolution layer uses a preset convolution kernel to perform convolution operations on the original image pixel matrix. Applied to filter signals and finding features in signals, the lower convolutional layer extracts lower-order features, and the later convolutional layer extracts higher-order features. The size of the convolution kernel in this study is 5 × 5. The method of convolutional layer operation is shown in Figure 2(a). For a 6 × 6 pixel original image, if a 3 × 3 convolution kernel is used, the convolution kernel will perform a convolution operation on the part inside the red frame of the original image, and then the
convolution kernel moves to the right, and the blue part inside the box is convolved, and the result of the two convolution operations is then convolved in the corresponding color box in the output. Next, the convolution kernel continues to move to the right of the line to the right, and then repeats the above moving convolution process starting from the part in the yellow box on the left of the next line. We follow the above method to complete all convolutions from top to bottom, from left to right, that is, complete...
convolution output. The convolution of the convolution kernel and any \( 3 \times 3 \) pixel matrix in the original image follows the discrete two-dimensional convolution as shown by Equation (1):

\[
B(i, j) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} K(m, n) \times A(i - m, j - n)
\]  

(1)

\( A \) is the convolution matrix, \( K \) is the convolution kernel, and \( B \) is the convolution result. Simply put, in the following figure, the number in the red box of the convolution output is equal to the sum of the product of the two numbers in the same position in the convolution kernel and the original image box, that is:

\[
B_{00} = A_{00}K_{00} + A_{01}K_{01} + A_{02}K_{02} + A_{10}K_{10} + A_{11}K_{11} + A_{12}K_{12} + A_{20}K_{20} + A_{21}K_{21} + A_{22}K_{22}
\]  

(2)

The pooling layer immediately follows the convolutional layer and is divided into the largest pooling layer and the average pooling layer. In this study, the maximum pooling layer is selected, and the step size is 2. It uses the maximum value of each \( 2 \times 2 \) pixels in the original image as the output of the corresponding position after pooling, as shown in Figure 2 (B). That is to say, every time through a pooling layer, the number of rows and columns of the matrix will become \( 1/2 \) of the original, this layer can greatly reduce the spatial resolution of the input image data. At the same time, various parameters in the maximum pooling layer can reduce the possibility of overfitting. The output of each pooling layer will use the ReLu activation function to add nonlinear elements.

The flat layer is used to one-dimensionalize the multi-dimensional input, and is often used in the transition from the convolutional layer to the fully connected layer. In this study, after the data passes through the flat layer, the multi-dimensional matrix is transformed into a one-dimensional matrix. The characteristic of the fully connected layer is that each neuron in this layer has all the neurons in the previous layer as input. The number of neurons in the last fully connected layer should be consistent with the number of classified classes, and the output value is the judged category. The fully connected layer in this study is responsible for reducing the number of neural network nodes and ultimately predicting the data category. For each fully connected layer, there is a discard parameter set to 0.5 to prevent the model from overfitting. The output layer is also processed by the Sigmoid activation function, and the final judgment result is obtained.

Since the number of effective images in some fields such as medicine is often limited, an overly complex network structure is likely to cause overfitting, which means that the network structure is very effective for this batch of inputs, but once the new input is replaced, the model. The performance will be greatly reduced, that is, the model lacks the ability to generalize. In order to avoid this situation, this study adds a downsampling layer to the convolutional neural network. During training, for the input of the previous layer, a certain proportion of network parameters are randomly discarded, and different subnetworks are used for iterative optimization each time. If the total number of network parameters is large enough, the problem of overfitting a certain network can be avoided. At the same time, as the number of parameters decreases, the calculation speed will also increase.

### 2.2. 2D SCNet

The feature extraction layer is shared by the segmentation and classification tasks in the dual-task network used in this paper. As mentioned above, U-Net and PSPNet have good feature learning capabilities in the segmented network. However, the U-Net network structure has a high degree of symmetry. If the U-Net encoder is stripped as the feature sharing layer of the dual-task network, the symmetry of U-Net will be destroyed, so 2D SCNet improves its network structure based on PSPNet.

The structure of 2D SCNet includes three parts: feature sharing network, segmentation network and classification network. Figure 3 shows the main structure of 2D SCNet. The feature sharing network is composed of ResNet50 + Pyramid Pooling Module (PPM), the classification network is composed of 4 residual blocks, and the segmentation network adopts a two-step segmentation strategy.

The original ResNet50 network input size is \( 224 \times 224 \), and the output of the network after all residual blocks is \( 1/32 \) of the original image. Our image size is \( 150 \times 150 \). If the original ResNet50 feature size reduction strategy is used, the size of the output feature map is too small, and the size cannot be reduced after the feature is sent to PPM. In addition, for the segmentation task, it is unreasonable for the encoder to output the feature map with a too small size and then upsample it to the original image size. Therefore, in order to increase the size of the output feature map of ResNet50, and considering the data size of \( 150 \times 150 \), this chapter adjusts the initial convolutional layer and pooling layer of ResNet50 as follows: (1) Convolution kernel \( 3 \times 3 \) size, a convolutional layer with a step size of 1, padding of 1, and a channel number of 64; (2) A pooling layer with a core size of \( 3 \times 3 \), a step size of 2, and a padding of 1. There are 3 feature size reductions in the residual block of the original ResNet50 (the step size of the convolutional layer is 2), this paper adjusts the convolution step size of all the residual blocks in ResNet50 to 1, so that the output characteristics of the residual block the size is unchanged.
The size of the feature map (7 × 7) fed into the PPM in the original PSPNet network is limited, so its PPM learns the receptive field features of different sizes by setting convolution kernels of different sizes instead of steps. However, according to Equation 2, the step size has a multiplicative effect on the receptive field, and the size of the convolution kernel only participates in the multiplication calculation once. Therefore, when the feature size is sufficient, it is more meaningful for the PPM to choose a different step size. ResNet50 output feature size (39 × 39) is large, which can be used to reduce the size of the feature with different steps to achieve the purpose of increasing the difference in the size of the receptive field. SCNet’s PPM uses asynchronous convolution. As shown in Figure 4, the PPM’s convolution kernel step size is set to 2, 4, and 6, respectively, and the three convolution outputs are about 1/2 of the input feature size, respectively. 1/4 and 1/6 are the same as the original PSPNet, the module output and input have a direct connection similar to ‘shortcut.’

For the classification network, the classifier was not directly added after the feature sharing network, and the shared features were sent to the classifier after 3 residual blocks. This design is so that when back propagating, the loss of classification can pass through several convolutional layers. Then propagate to the shared feature layer, so that the result of the classification indirectly changes the parameter value of the feature shared layer, avoiding the direct result of the classification during network training influences.

![Figure 3. 2D SCNet network structure.](image)

![Figure 4. Pyramid pooling structure.](image)
2.3. Data set composition

CT angiography is an imaging technique that can clearly display the details of blood vessels throughout the body. After the contrast agent is injected into the blood vessel, because the density of the contrast agent is generally higher than that of the surrounding tissue, the attenuation of X-rays through the contrast agent is large, so the blood vessel containing the contrast agent flows through the blood vessel imaging gray value (CT value) is high. After injection of contrast medium, the patient, the arteries and vessels are filled with imaging agents. At this time, the renal blood vessels and cortical parts of the image will be highlighted, and the other positions will be displayed normally. The obtained scan image is a CT angiography image (CTA). For kidneys with tumors, CTA can clearly reflect the contour of the renal cortex and the size of the tumor. It can outline the branch involvement of the aortic dissection and the compression of the true cavity against the false cavity. After a certain period of time, when the contrast agent fills the urinary tract, the scanned images of the blood vessels and cortex are CT urography images (CTU). Note that CTU can clearly reflect renal parenchyma and renal pelvis. The CTA and CTU images together can clearly show the whole image of the kidney part.

A total of 168 patient data, each patient contains three-dimensional CTA and CTU two-stage data, a total of four types of renal tumor subtypes, they contain the number of patients are: 38 cases of lipid-free smooth muscle lipoma, 64 cases of renal cell carcinoma, 23 Papillary renal cell carcinoma and 43 cases of chromophobe cell carcinoma.

2.4. Judgment criteria for comparative experiments

In the experiment, the index of Dice coefficient is used, and the expression of Dice coefficient is shown in Equation 3.

\[ \text{Dice}(G, P) = \frac{2|G \cap P|}{|G| + |P|} \]  

\( \text{Dice}(G, P) \) represents the Dice coefficient of the i-th object in the manual segmentation result G and the dense prediction result P by experts. \(|G|\) represents the number of elements, \(G \cap P\) represents the common element of the i-th set of objects in the expert manual segmentation result and the i-th set of objects in the dense prediction result, \(|G| + |P|\) represents the i sum of pixels in the object-like area.

The classification criteria are accuracy rate and recall rate. True positive (TP), false negative (FN), false positive (FP), and true negative (TN) indicate the number of positive predictions and the number of positive prediction errors, respectively. The number of negative prediction errors and the number of negative prediction errors. The calculation of accuracy rate and recall rate are shown in Equations 4 and 5, respectively.

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

\[ R = \frac{TP}{TP + FN} \]

3. Results

3.1. Pretreatment

The CTU and CTA images are registered using the toolkit elasitx. Each patient’s data is drawn by the hospital’s radiology experts to outline the kidneys and tumors, and the experts manually segment the results. In the result of expert manual segmentation, gray represents the kidney and red represents the tumor. Figure 5 shows the two-dimensional images of CTA and CTU of kidney tumor and the result of expert manual segmentation.

The comparison of the output results of the segmentation network is shown in Figure 6. From the two-dimensional situation in Figure 6, compared with the other two networks, the edge of the 2D SCNet segmentation result is smoother, and the tumor area is more complete, which is better for smaller tumors.

Table 1 shows the data distribution of training set and test set. There are 168 data in total, and the ratio of training set and test set is 2:1. Due to the lack of data, no verification set is set in this paper.
For the training of convolutional neural networks, the data volume of 168 three-dimensional data is small, and the benign tumors in the data set account for less than 23%, and renal clear cell carcinoma accounts for nearly 40%. Therefore, a small number of this paper and an uneven distribution of categories are needed. The data set is expanded. This paper uses different enhancement ratios for the proportion of benign and malignant tumors to make the data set balanced. We implemented some common methods to enhance the data set to expand the data set, such as rotation, flipping, and cropping.

The first one adds different labels normal to these special two-dimensional images, and finally there are three types of images in the two-dimensional data set. The labels are benign, malignant and normal. The second method is to exclude sections without tumor tissue, and the final labels are benign and malignant. Subsequent experiments will use three labels and two labels to refer to these two schemes.

### 3.2. Testing the validity of the model

In order to test the effectiveness of our model, we conduct comparison experiments based on the following points: (1) Compare with different single classification models to verify the accuracy of 2D SCNet classification; (2) Compare with different single segmentation models, Compare the misclassification and omission of different models of kidney and tumor regions; (3) Based on the 2D SCNet segmentation and classification results of different base networks. We observe the accuracy and misclassification of different base networks. The base network refers to the network structure used by the shared feature layer.

---

**Figure 6.** Comparison of two-dimensional images of kidney tumors, expert manual segmentation results and the results of various models depicting the (a) CT images; (b) expert manual segmentation results; (c) 2D SCNet + three labels + two-step segmentation; (d) PSPNet segmentation; (e) U-Net segmentation.

**Table 1.** Number distribution of tumors of different subtypes.

| Tumor type                        | Training set | Test set | Total | Proportion | Two-dimensional image | Expanded two-dimensional image |
|-----------------------------------|--------------|----------|-------|------------|-----------------------|-------------------------------|
| Lean fatty renal smooth muscle lipoma | 27           | 11       | 38    | 22.62%     | 2700                  | 18,500                        |
| Renal clear cell carcinoma        | 46           | 18       | 64    | 38.10%     | -                     | -                             |
| Papillary cell carcinoma          | 14           | 9        | 23    | 13.69%     | 6500                  | 21,000                        |
| Chromophobe cell carcinoma        | 28           | 15       | 43    | 25.39%     | -                     | -                             |
| Total                             | 115          | 53       | 168   | -          | 8400                  | 37,500                        |

---

Z. GONG AND L. KAN
3.2.1. Classification of benign and malignant kidney tumors based on different models

Due to the small number of patients in each subcategory, we used a 10-fold cross-validation experiment for the classification network and compared 2D SCNet, SVM, ResNet50, and DenseNet. As shown in Table 2, although other algorithms can get better classification results, the combination of segmentation and classification of 2D SCNet can achieve 99.5% accuracy in both benign and malignant classification, both based on the ‘two labels’. Higher than SVM 18.1% and 12.8% respectively. Based on the ‘three labels,’ it is 9.2% and 7.1% higher than Resnet50 and 18.4% and 9.6% higher than DenseNet. The combination of segmentation and classification tasks, the manual segmentation results of the segmentation task increase the priori information of the classification task and help the feature learning of the ROI region. From the experimental results, it can be seen that the dual-task network can obtain better classification than the single classification task. Better network effect.

It can be seen from Table 2 that for all models, the malignant classification results are slightly better than benign. This is because despite the data expansion, the actual number of malignant tumors is more than benign, and the network has better generalization ability for malignant tumors. In addition, DenseNet’s classification results compared with ResNet50, the accuracy rate of benign tumors decreased by 9.2%, and the accuracy rate of malignant tumors decreased by 2.5%. Because when the network complexity increases, the limited data set will increase the training difficulty, and it is easy to over fit.

3.2.2. Segmentation of kidney tumors based on different models

For the segmentation task, we compared 2D SCNet, PSPNet and U-Net. As shown in Table 3, under the condition of three labels, the segmentation results of 2D SCNet+three labels reached the Dice coefficients 0.946 and 0.846, respectively. Compared with PSPNet, our network kidney and tumor segmentation results have improved by 4.9% and 5.0%, respectively, which indicates that the addition of classification modules is beneficial to the learning of segmentation networks. 2D SCNet+Three Labels +Two-Step Segmentation Dice coefficients obtained based on the two-step segmentation method are 0.8% higher than the direct use of the dense prediction classifier 2D SCNet+Three Labels and 5.8% higher than PSPNet. The network performance has been improved.

Table 3. Dice coefficients of kidney tumor segmentation based on different models.

| Model                        | Patient-based segmentation Dice coefficient |
|------------------------------|--------------------------------------------|
|                              | Kidney | Tumor |
| U-Net                        | 0.912  | 0.831 |
| PSNet                        | 0.897  | 0.796 |
| 2D SCNet+ three labels       | 0.946  | 0.846 |
| 2D SCNet+ three labels +two-step segmentation | 0.923  | 0.854 |

Table 4. Segmentation of kidney tumors based on different table.

| Model                        | Patient-based classification Dice coefficient |
|------------------------------|-----------------------------------------------|
|                              | Patient-based classification                  | Patient-based segmentation Dice coefficient |
|                              | Benign recall rate | Vicious recall rate | Accuracy | Kidney | Tumor |
| 2D SCNet+ three labels       | 0.995 | 0.995 | 0.995 | 0.946 | 0.846 |
| 2D SCNet+ three labels +two-step segmentation | 0.995 | 0.995 | 0.995 | 0.923 | 0.837 |
| 2D SCNet+ two labels         | 0.995 | 0.995 | 0.995 | 0.913 | 0.856 |
| 2D SCNet+ two labels +two-step segmentation | 0.995 | 0.995 | 0.995 | 0.942 | 0.887 |
3.2.4. **2D SCNet based on different base networks**

This paper also tested the 2D SCNet experiment results based on different base networks-based on ResNet50 + PPM and DenseNet. The DenseNet used in the experiment contains 3 dense blocks, and each block contains 32 layers of convolution. Since our training image is small, the pooling layer in the transition layer is removed from the DenseNet structure, and a 3 × 3 convolution and stepping pooling layer are added at the beginning of the network. The experimental results are shown in Table 5.

It can be seen from Table 5 that the Dice Net-SCNet tumor area and kidney area segmentation Dice coefficient is lower than ResNet50 + PPM-2D SCNet.

### Table 5. 2D SCNet based on different base networks.

| Model                  | Patient-based classification +two-step segmentation | Patient-based segmentation Dice coefficient |
|------------------------|----------------------------------------------------|---------------------------------------------|
| ResNet50+ PPM-2D       |                                                     |                                             |
| SCNet + three labels   | 0.993                                              | 0.987                                       |
| +two-step segmentation  |                                                    | 0.990                                       |
| 2D SCNet++ three labels | 0.993                                              | 0.988                                       |
| +two-step segmentation  |                                                    | 0.986                                       |

4. Discussion

CNN is a multi-layer perceptron model designed by simulating biological neural networks (Bayramoglu et al., 2016; Havaei et al., 2017; LeCun et al., 1989). Its basic unit is a neuron. All the links between the neurons of the multi-layer perceptron lead to more parameters, while the local connections and weights of CNN sharing two major features greatly reduces the number of parameters. Local connection means that each layer of neurons is only connected to the neurons in the upper part of the region, which is the receptive field (receptive field), so that the neurons are locally connected in the spatial dimension (spatial), but they are all connected in depth; fact above, for two-dimensional images, the correlation of local pixels is strong, and the local connection can fully guarantee the response of the filter to the local input features.

The upper layers of convolutional neural networks generally pay more attention to high-level semantic information such as classes and objects contained in the image, while the lower layers tend to learn low-level semantic information such as position and texture (Kawahara et al., 2017; Linguraru et al., 2011; Neocognitron; 1980; Way et al., 2010). Although the deeper the neural network layers, the larger the effective receptive field size, but the experiment in the paper (Zhou et al., 2014) proved that the actual receptive field size of the network is smaller than the theoretical calculation. Shin et al. (Bo et al., 2021; Chen et al., 2014; He et al., 2016; I-C et al., 2016; Long et al., 2021; Ronneberger et al., 2015) use dilated convolution to increase the receptive field without pooling loss information so that each convolution contains a larger range of information. Yu et al. (Yu & Multi, 2015) proposed a global average pooling method to replace the fully connected layer. This structure can add global semantics to the final classifier, which solves the problem of insufficient actual receptive field. Liu et al. applied the idea of global average pooling to semantic segmentation (Lin et al., 2013). Liu et al. (Liu et al., 2015) added a pyramid pooling module on the basis of global average pooling, which combined regional features of different sizes to further improve the receptive field (Zhao et al., 2017).

Scholars such as Iglésias et al. initially focused on segmentation of kidney-free areas or cysts without tumors. For example, the use of regional growth or multi-template-based registration can accurately segment kidney contours (Iglésias & Sabuncu, 2015). Some work segmented the cyst area. For example, Badura et al. extracted the kidney area through the mark control watershed algorithm, used shape-related three-dimensional features and convolutional neural networks as classifiers to roughly detect cysts, and used anisotropic diffusion filtering and mixing in the fine stage. Level set method (Badura et al., 2016). Bae et al. (Bae et al., 2013) used k-means clustering to obtain cyst areas that are brighter than the renal parenchyma, and also used filters such as anisotropic diffusion filtering to obtain cyst contours. The shape of the kidney area is uniform, and the gray distribution of the cyst area is similar and evenly distributed. On the contrary, the kidney shape of the tumor is irregular, the tumor position is irregular, and the gray distribution is very different. Although the kidney or cyst area can obtain good segmentation results, the complex composition of kidney tumors makes the tumor segmentation very difficult, and traditional kidney or tumor segmentation methods are powerless.

For the segmentation study of kidney area and tumor area, as early as 2009, some scholars used three phases of contrast-enhanced CT data (before contrast agent administration, arterial phase and portal vein phase) for segmentation (Linguraru et al., 2011). First, the three-phase image is registered (aligned), and then fast marching and active contour model are used for segmentation. The article mentioned that because the lesions may be heterogeneous, the segmentation algorithm may stop at the edge of the internal lesion, which also illustrates the difficulty faced by the segmentation problem-the variable shape and texture distribution.

Lee et al. (2019) pays attention to the segmentation of kidney tumors with a diameter of less than 4 cm. The method proposed in this paper first detects the ROI.
region of the kidney through grayscale and threshold, extracts candidate tissues, and then uses block-based texture and context feature classification to reduce false positives (misdetection) in the kidney area. Ravinder et al. proposed a spatial intuitive fuzzy C-means clustering method (SIFCM) to divide the lesion area (Kaur et al., 2019).

In addition to using some clustering or feature-based segmentation methods, some work uses convolutional neural networks for segmentation. Yang et al. used a three-dimensional fully convolutional neural network to segment the kidney region, and used hollow convolution and pyramid pooling modules to increase the receptive field. Haghighi et al. used a cascaded three-dimensional convolutional neural network to first coarsely locate the kidney region and then finely segment the kidney region (Haghighi et al., 2017).

5. Conclusion

For segmentation and classification of kidney tumors, this paper mainly introduces and network design based on two-dimensional convolutional neural network. For our segmentation and classification tasks, the diversity of kidney sizes requires the network to be able to combine features of different sizes and learn global context information. The characteristics of different sizes mean that the size of the receptive field is a key factor. Although the network structure of U-Net and PSPNet satisfies this point, the receptive field is not equal to the global context information. Our network combines segmentation and classification tasks, and utilizes the correlation between global context information and existing categories to reduce false segmentation. In addition, this paper also adopts a two-step segmentation strategy to further improve the experimental results. From the cross-validation results, we can see that 2D SCNet and the two-step segmentation strategy can obtain better results in the segmentation and classification tasks of kidney tumors.

Disclosure statement

No potential conflict of interest was reported by the author(s)

Ethical compliance

There is no ethics approval required for this paper.

References

Pawel Badura, Wojciech Wieclawek, Bartłomiej Pycinski. (2016). Automatic 3D segmentation of renal cysts in CT [M]. Information Technologies in Medicine. Springer International Publishing. pp 149–163. DOI:10.1007/978-3-319-39796-2_13

Bae, K., Park, B., Sun, H., Wang, J., Tao, C., Chapman, A. B., Torres, V. E., Grantham, J. J., Mrug, M., Bennett, W. M., Flessner, M. F., Landsittel, D. P., & Bae, K. T. (2013). Segmentation of individual renal cysts from MR images in patients with autosomal dominant polycystic kidney disease [J]. Clinical Journal of the American Society of Nephrology, 8(7), 1089–1097. https://doi.org/10.2215/CJN.10561012

Neslihan Bayramoğlu, Juho Kannala, Janne Heikkilä. (2016). Deep learning for magnification independent breast cancer histopathology image classification[C]. 2016 23rd International conference on pattern recognition (ICPR).IEEE, 2440–2445. DOI:10.1109/ICPR.2016.7900002

Bo Hu, Changjiang Zhou, Hongbing Wang, Lairong Yin. (2021). Prediction and validation of dynamic characteristics of a valve train system with flexible components and gyroscopic effect. Mechanism and Machine Theory, 157, 104222. https://doi.org/10.1016/j.mechmachtheory.2020.104222

Laura J. Brattain, Brian A. Telfer, Manish Dhyani, Joseph R. Grajo & Anthony E. Samir Abdominal Radiology. (2018). Machine learning for medical ultrasound: Status, methods, and future opportunities [J]. Abdom Radiol, 43(4), 786–799. https://doi.org/10.1007/s00261-018-1517-0

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille. (2014). Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. arXiv Preprint arXiv, 412.7062, 1. https://arxiv.org/abs/1412.7062

Doi, K. (2006). Diagnostic imaging over the last 50 years: Research and development in medical imaging science and technology [J]. Physics in Medicine and Biology, 51 (13), R5. https://doi.org/10.1088/0031-9155/51/13/R02

Dominik Müller & Frank Kramer. (2009). Preoperative aspects and dimensions used for an anatomical (Padua) classification of renal tumours in patients who are candidates for nephron-sparing surgery[J]. European Urology, 56(5), 786–793. https://doi.org/10.1016/j.eururo.2009.07.040

Marzieh Haghighi, Simon K. Warfield, Sila Kurugol. (2018). Automatic renal segmentation in DCE-MRI using convolutional neural networks [J]. IEEE 9(4), 1945–8452. DOI:10.1109/ISBI.2018.8363865

Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin, Hugo Larochelle. (2017). Brain tumor segmentation with deep neural networks [J]. Medical Image Analysis, 35, 18–31. https://doi.org/10.1016/j.media.2016.05.004

Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. (2016). Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 770–778. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Hinton, G. E., Osindero, S., & The, Y. W. (2006). A fast learning algorithm for deep belief nets [J]. Neural Computation, 18(7), 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527

I-c, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNNarchitectures, dataset characteristics and transfer learning[J]. IEEE Transactions on Medical Imaging, 35(5), 1285–1298. https://doi.org/10.1109/TMI.2016.2528162

Iglesias, J. E., & Sabuncu, M. R. (2015). Multi-atlas segmentation of biomedical images: A survey [J]. Medical Image Analysis, 24(1), 205–219. https://doi.org/10.1016/j.media.2015.06.012
