Detecting multiple lesions of lung cancer-caused metastasis with bone scans using a self-defined object detection model based on SSD framework

Qiang Lin\textsuperscript{1,2,3}, Xiaoyan Chen\textsuperscript{2,3}, Liangxia Liu\textsuperscript{1,2}, Yongchun Cao\textsuperscript{1,2,3}, Zhengxing Man\textsuperscript{1,2,3}, Xianwu Zeng\textsuperscript{4} and Xiaodi Huang\textsuperscript{5}

\textsuperscript{1} School of Mathematics and Computer Science, Northwest Minzu University, People’s Republic of China
\textsuperscript{2} Key Laboratory of China’s Ethnic Languages and Information Technology of Ministry of Education, Northwest Minzu University, People’s Republic of China
\textsuperscript{3} Key Laboratory of Streaming Data Computing Technologies and Application, Northwest Minzu University, People’s Republic of China
\textsuperscript{4} Department of Nuclear Medicine, Gansu Provincial Tumour Hospital, People’s Republic of China
\textsuperscript{5} School of Computing and Mathematics, Charles Sturt University, Albury, NSW, Australia

E-mail: qiang.lin2010@hotmail.com and 1021619285@qq.com

Keywords: bone scan, \textsuperscript{99m}Tc MDP-SPECT, metastasis, lesion detection, convolutional neural network

Abstract

Objective. To facilitate manual diagnosis of lung cancer-caused metastasis, in this work, we propose a deep learning-based method to automatically identify and locate the hotspots in a bone scan image which denote the lesions metastasized from lung cancer. Approach. An end-to-end metastasis lesion detection model is proposed by following the classical object detection framework single shot multibox object detector (SSD). The proposed model casts lesion detection problem into automatically learning the hierarchical representations of lesion features, locating the spatial position of lesion areas, and boxing the detected lesions. Main results. Experimental evaluation conducted on clinical data of retrospective bone scans shows the comparable performance with a mean score of 0.7911 for average precision. A comparative analysis between our network and others including SSD shows the feasibility of the proposed detection network on automatically detecting multiple lesions of metastasis lesions caused by lung cancer. Significance. The proposed method has the potential to be used as an auxiliary tool for improving the accuracy and efficiency of metastasis diagnosis routinely conducted by nuclear medicine physicians.

1. Introduction

Object detection has been a hot topic in medical image analysis especially in the automated localization and identifying of regions of interest (e.g. organs, tissues, and lesions) in images (Litjens et al 2017). Physicians need pay attention only to the detected regions of interest while ignoring large-size background area of an image during manual diagnosis of diseases, having the huge potential to improve the accuracy and efficiency of diagnosis.

Bone scan (bone scintigraphy) is of the widely-accepted clinical tools for screening bone metastasis originated from a variety of various solid tumors including lung cancer. With \textsuperscript{99m}Tc MDP (99m Technetium methylene diphosphonate), SPECT (single photon emission computed tomography) imaging displays a bone metastasis lesion as an area with high uptake of radiopharmaceutical (Bombardieri et al 2003). It has been proved that \textsuperscript{99m}Tc-MDP SPECT is more affordable and available than PET (positron emission tomography) due to its low-cost equipment and radiopharmaceutical (Lin et al 2020a).

The \textsuperscript{99m}Tc-MDP SPECT imaging is characterized by low specificity and inferior resolution (Nathan et al 2013), significantly impeding human manual analysis of bone scan images for bone metastasis diagnosis. First, it is a challenging task to accurately distinguish a real metastasis lesion from benign processes. This is because, for example, osteoarthritis and bone injury often manifest as high-uptake areas in a SPECT bone scan image, which
would bring misinterpretation to human diagnosis. Second, there has no clear boundary between a metastasis lesion and the normal bone as compared to anatomical imaging modalities like computed tomography (CT) and magnetic resonance imaging (MRI). Segmenting a low-resolution $^{99m}$Tc-MDP SPECT image to accurately measure metastasis lesions is impracticable in routine nuclear medicine practice.

Using object detection algorithms to automatically identify and locate lesions plays a vital role in the field of automated medical image analysis. Not only the disease type that a lesion belongs to but also the location that such a lesion is present in can be determined in an automated way. There has seen an increasing interest in automated detection of lesions within anatomical medical images since the prevalence of convolutional neural network (CNN), which has the ability of automatically learning hierarchal representations of images and dividing high-level features in an end-to-end fashion. With x-ray images, a four-layer CNN model has been developed to detect lung nodules (Lo and Lou 1995). A set of deep CNNs were investigated to detect lymph nodes, sclerotic metastases, and colon polyps in CT scans (Roth et al. 2014, 2016, Yao et al. 2015). Teramoto et al. (Teramoto et al. 2016), studied to propose a multi-stream CNN model for detecting lung nodules with hybrid PET/CT images. A 3D CNN model was proposed to detect brain microhemorrhages with MRI images (Dou et al. 2015). Weak supervised deep learning was also studied to develop CNN-based detection of lung nodules in x-ray mammograms (Hwang and Kim 2016). CheXNeXt (Rajpurkar et al. 2018) is a CNN-based model used for simultaneously detecting 14 different types of lesions including pneumonia, pleural effusion, pulmonary masses, and pulmonary nodules.

In the domain of $^{99m}$Tc-MDP SPECT image analysis, existing research efforts lie in developing CNN-based image classification methods, targeting at answering whether a metastasis lesion is present (i.e. two-class classification (Dang 2016, Papandrianos et al. 2020a, 2020b, 2020c, 2020d, Pi et al. 2020, Zhao et al. 2020, Cheng et al. 2021a, 2021b, Lin et al. 2021a)) or how many lesions of different diseases are present in a bone scan image (i.e. multi-class classification (Lin et al. 2021b, 2021c, Guo et al. 2022, Li et al. 2022)). A CNN-based supervised segmentation model was proposed to automatically delineate metastasis lesions in regional SPECT bone scan images (Lin et al. 2020b), achieving a mean score of 0.6103 for intersection over union (IoU). The objective of this work is to propose a self-defined lesion detection model by following the single shot multibox object detector (SSD) framework, having different objective from image classification (Dang 2016, Pi et al. 2020, Papandrianos et al. 2020a, 2020b, 2020c, 2020d, Zhao et al. 2020, Cheng et al. 2021a, 2021b, Lin et al. 2021a, 2021b, 2021c, Guo et al. 2022, Li et al. 2022) and segmentation (Lin et al. 2020b) tasks.

CNN-based automated detection of metastasis lesions is still untillled in the $^{99m}$Tc-MDP SPECT image analysis field. To facilitate manual diagnosis of lung cancer-caused metastasis, in this work, we propose a CNN-based lesion detection method by following the classical object detection framework SSD (Liu et al. 2016). The proposed method can identify and locate a bone metastasis lesion in a SPECT bone scan image, enabling to improve diagnosis accuracy and efficiency.

The main contributions of this work can be summarized as: First, to the best of our knowledge, we are the first to try to automatically detect bone metastasis lesions with $^{99m}$Tc-MDP SPECT bone scans. Second, by following the classical SSD framework, a CNN-based end-to-end model is developed to transform the lesion detection problem into learning the hierarchal representations of lesion features, locating the spatial position of lesion areas, and boxing the detected lesions. Lastly, a set of clinical data of retrospective SPECT bone scans are used to evaluate the proposed method, showing the comparable detection performance with a mean score of 0.7911 for the composite metric average precision (AP).

The rest of this paper is organized as follows. We present in section 2 the data used and the proposed lesion detection method. We report in section 3 the experimental evaluation conducted on clinical $^{99m}$Tc-MDP SPECT scan bone images. We provide in section 4 a brief discussion about the pros and cons of the proposed method. In section 5, we conclude this work and point out the future research directions.

2. Materials and methods

The SPECT bone scans used and the proposed lesion detection method are detailed in this section.

2.1. Bone scan image and preprocessing

In this retrospective study, the $^{99m}$Tc-MDP SPECT bone scan images used were collected from the Department of Nuclear Medicine, Gansu Provincial Tumor Hospital. During the SPECT imaging, a single-head gamma camera (GE SPECT Millennium MPR) was used to acquire the anterior- and posterior-view whole-body images from patients who were clinically diagnosed with lung cancer, where $^{99m}$Tc MDP (20–25 mCi) was intravenously injected into the body of a patient.

A total of 527 patients with lung cancer were involved in the dataset, resulting in 1054 whole-body SPECT bone scan images. To focus only on the thoracic region that was widely identified as one of the most common
area of bone metastasis (Nathan et al. 2013), we extracted the regional thorax sub-image from every whole-body image to construct a dataset consisting of 306 regional sub-images. Those images containing bone metastasis in other areas were excluded. In other word, this work aims to develop a bone metastasis lesion detection model conducting on regional SPECT bone scan images. An extracted regional sub-image has the size of $256 \times 256$, by filling in edge with background if necessary.

Three experienced nuclear medicine physicians from our group manually delineated the boundary of each lesion using a LabelMe (http://labelme.csail.mit.edu/Release3.0/) based annotation system. The labeled lesions act as ground truth in the experiments, which will be fed into the detection model for training purpose.

2.2. The proposed method

Figure 1 outlines the proposed SSD-based lesion detection method. An inputted $256 \times 256$ image is first convoluted using a $7 \times 7$ filter ($7 \times 7 \text{Conv}, c_{\text{out}} = 64$) to produce feature maps, which is followed by a down-sampling using a $3 \times 3$ pooling layer ($3 \times 3 \text{MaxPool}, S = 1$), where $c_{\text{out}}$ and $S$ is the channel number and stride length, respectively. The feature extraction sub-network works to extract shallow-to-deep image features, aiming to yield smaller feature maps. Lesion localization & boxing stage is used to locate lesion areas in images (feature maps) and box each area with a rectangle.

2.2.1. Feature extraction

To facilitate the detection of varied-size metastasis lesions in low-resolution SPECT images, in this work, we propose a feature extraction sub-network consisting of cascaded convolution blocks with residual connections (see figure 2).

Four groups of convolution blocks are included in the defined feature extraction sub-network, with each block consisting of a $3 \times 3$ convolution layer ($3 \times 3 \text{Conv}, c_{\text{out}}$) and a $1 \times 1$ convolution layer ($1 \times 1 \text{Conv}, c_{\text{out}}$), where $c_{\text{out}}$ is the channel number. The number of blocks in these groups is indicated by $\{3, 3, 5, 3\}$. The size of feature maps evolves from larger to smaller while the extracted image features changing from shallower to deeper.

As depicted in figure 3, there is a residual connection (i.e. Intra-res) between two adjacent convolutional layers within a block and a residual connection (i.e. Inter-res) between two convolutional layers of different blocks.
The extracted higher-level features will be fed into the lesion localization & boxing stage to identify lesion areas and label these areas with boxes.

2.2.2. Lesion localization and boxing

Figure 4(a) details the structure of feature extraction sub-network, which outputs a group of varied-size feature maps of \( \{32 \times 32, 16 \times 16, 13 \times 13, 11 \times 11, 9 \times 9, 7 \times 7, 5 \times 5, 3 \times 3, 1 \times 1\} \). With these feature maps, a two-stage operation consisting of candidate box (CB) generation and valid candidate box (VCB) selection is conducted to locate and box each lesion area in an image (see figure 4(b)).

2.2.2.1. CB generation

In the CB generation stage, we need to first establish a mapping between the feature maps and the manual labels in the original image to facilitate locating the lesions areas. An input image is divided into grids according to the size of feature maps by regarding the geometric center of each grid as a midpoint, where candidate boxes with six types of sizes are used.

Let \( SC_k \) denote the width of a square CB and \( \text{Input}_{-}\text{width} \) be the width of an original image, then \( S_k = SC_k / \text{Input}_{-}\text{width} \) can be calculated according to equation (1).

\[
S_k = S_{\text{min}} + \frac{S_{\text{max}} - S_{\text{min}}}{m - 1} (k - 1), \quad k \in [1, m],
\]  

where \( m \) denotes the number of the feature maps; and \( S_{\text{max}} \) and \( S_{\text{min}} \) is the maximum and minimum of \( S_k \), respectively.

A value of 0.9/0.2 for \( S_{\text{max}}/S_{\text{min}} \) works well in the experiments. The varied-size candidate boxes can thus be obtained by adjusting the value of \( S_k \). Particularly, the square CB pertaining to the first feature map (i.e. \( 32 \times 32 \)) has \( S_0 = S_{\text{min}}/2 \), meaning that the width of the square CB of the first feature map is \( SC_0 = S_0 \times \text{Input}_{-}\text{width} = S_{\text{min}}/2 \times \text{Input}_{-}\text{width} = 0.1 \times 256 \approx 25 \). For any rectangle CB, the height \( h \) and the width \( w \) can be calculated according to equation (2).

\[
\begin{align*}
    h &= SC_k / \sqrt{ar} \\
    w &= SC_k \times \sqrt{ar}
\end{align*}
\]  

where \( ar = \text{width} : \text{height} \).

The visual observation of bone metastasis lesions in SPECT bone scan images suggests several empirical values of \( \{1′, 1, 2, 3, 1/2, 1/3\} \) for \( ar \), where \( 1′ \) denotes \( SC_k = (SC_k \times SC_{k+1})^{1/2} \) (Liu et al 2016). Since feature maps in the shallower network layers are often used to detect smaller objects while feature maps in the deeper layers corresponding to larger objects, we arrange \( ar \in \{1′, 1, 2, 1/2\} \) for the feature maps of \( \{32 \times 32, 3 \times 3, 1 \times 1\} \). The parameter settings are outlined in table 1.

2.2.2.2. VCB selection

As shown in table 1, a total of 8342 candidate boxes are generated in the CB generation stage, which can be further divided into positive and negative samples. Specifically, a CB is called positive sample if it partially or fully covers a real lesion; it is a negative sample otherwise. VCB selection conducts to reduce the number of the positive and negative samples to speed up the model training, which is implemented by matching positive samples and mining hard negative samples.
Matching positive sample: a positive CB will be selected as a VCB if it has \( \text{IoU} > \theta \) (strong positive sample), where IoU measures the overlap between this CB and its ground truth (i.e. manual label). The one with largest IoU will be also selected as a VCB if there is no strong positive sample for a feature map.

Mining hard negative sample: differing from the natural images, the proportion of the lesion areas (foreground) is far less than the background for bone scan images. This means the most of the generated 8342

Table 1. Parameter settings in CB generation.

| Feature map | \( SC_k \) | \( ar \)          | CB in single feature map | Total number of CB |
|-------------|------------|------------------|--------------------------|--------------------|
| 32 × 32     | 25         | \{1',1,2,1/2\}   | 4                        | 32 × 32 × 4        |
| 16 × 16     | 51         | \{1',1,2,3,1/2,1/3\} | 6                        | 16 × 16 × 6        |
| 13 × 13     | 73         | \{1',1,2,3,1/2,1/3\} | 6                        | 13 × 13 × 6        |
| 11 × 11     | 96         | \{1',1,2,3,1/2,1/3\} | 6                        | 11 × 11 × 6        |
| 9 × 9       | 118        | \{1',1,2,3,1/2,1/3\} | 6                        | 9 × 9 × 6          |
| 7 × 7       | 140        | \{1',1,2,3,1/2,1/3\} | 6                        | 7 × 7 × 6          |
| 5 × 5       | 163        | \{1',1,2,3,1/2,1/3\} | 6                        | 5 × 5 × 6          |
| 3 × 3       | 185        | \{1',1,2,1/2\}    | 4                        | 3 × 3 × 4          |
| 1 × 1       | 208        | \{1',1,2,1/2\}    | 4                        | 1 × 1 × 4          |

- Matching positive sample: a positive CB will be selected as a VCB if it has IoU > \( \theta \) (strong positive sample), where IoU measures the overlap between this CB and its ground truth (i.e. manual label). The one with largest IoU will be also selected as a VCB if there is no strong positive sample for a feature map.
- Mining hard negative sample: differing from the natural images, the proportion of the lesion areas (foreground) is far less than the background for bone scan images. This means the most of the generated 8342
candidate boxes are negative samples. To keep balance between the positive and negative samples, we need to select valid negative samples by mining the hard negative samples. Specifically, a negative sample is a hard negative sample if it has largest negative loss, \( L_{\text{Neg}} \), which is defined in equation (3).

\[
L_{\text{Neg}} = -\frac{1}{N} \sum_{i \in \text{Neg}} \log(\hat{c}_i^0) \mid \hat{c}_i^0 = \frac{\exp(c_i^0)}{\sum_0 \exp(c_i^0)},
\]

where \( N \) is the number of negative samples, and \( \hat{c}_i^0 \) is the probability of the background class.

With the selected VCBs, the proposed lesion detection model can be trained. During the model test stage, we use the non-maximize suppression algorithm (Rosenfeld and Thurston 1971) to reduce the overlaps of predictions in the form of boxes. Suppose \( n \) boxes \( \{pb_1, pb_2, \ldots, pb_n\} \) with each having a class score \( s_i \) (\( 1 \leq i \leq n \)), the non-maximize suppression algorithm works as follows:

1. Let \( A = \{pb_1, pb_2, \ldots, pb_n\} \) and \( B = \emptyset \).
2. Moving \( pb_i \) from \( A \) to \( B \) if it has the current largest score \( s_i \).
3. \( \forall \ pb_j \in A \ (j \neq i) \), if IoU \( (pb_i, pb_j) \geq \theta \), removing \( pb_j \) from \( A \), where \( \theta \) is a predefined threshold.
4. Goto 2.

The algorithm above stops if no element in the set \( A \) meets the selection requirement. The boxes in the set \( B \) are the resultant outputs of the proposed detection model.

### 2.3. Experimental setup

The experimental evaluation metric used in this work is AP, which is defined as the area under the P–R curve, where \( P = \text{Precision} \) and \( R = \text{Recall} \) are as follows.

\[
\text{Precision} = \frac{TP}{TP + FP}, \tag{4}
\]

\[
\text{Recall} = \text{Sensitivity} = \frac{TP}{TP + FN}, \tag{5}
\]

where the annotations are defined as follow.

\( TP = \text{True Positive}: \) The number of the predicted boxes with IoU \( > \theta \), where IoU measures the overlap between a predicted positive box and its ground truth; \( FP = \text{False Positive}: \) The number of the predicted positive boxes with IoU \( < \theta \); \( FN = \text{False Negative}: \) The number of the predicted boxes covering no ground truth (manual labels).

The dataset consisting of 306 thoracic SPECT images are divided into two parts: the training set (\( n = 212, \sim 70\% \)) and the test set (\( n = 94, \sim 30\% \)). The experimental results reported below are ten-fold cross-validation scores of evaluation metrics.

The parameter settings of the CNN-based lesion detection model are outlined in table 2.

### 3. Results

This section reports the experimental evaluation conducted on a set of clinical \(^{99m}\text{Tc-MDP} \) SPECT bone scan images. The experiments are run in Tensorflow 2.0 on an Intel Core i7-9700 PC with 32GB RAM running Windows 10 operating system.
3.1. Performance results
Table 3 reports the performance obtained by the proposed lesion detection model on the test samples (~30% of the images in the dataset used) by providing the scores of AP, Precision, and Recall (sensitivity).

High score of Precision reveals that the proposed detection model has the ability to successfully identify true positives while suppressing false positives. However, the model obtains low score for Recall, which is contributed by the high false negatives. This is due to the inconsistent appearance of metastasis lesions in SPECT bone scan images among patients with various bony metabolic activities. Some positive pixels were incorrectly detected as the negative, hence the high false negatives. Fortunately, the composite metric AP that measures the area under the Precision–Recall (P–R) curve obtains the relatively high score, which is depicted in figure 5.

The P–R curve depicted in figure 5 shows that a score of no less than 0.7 can be obtained for Recall when Precision is not more than 0.8. This combined with the mean score of 0.7911 for AP proves the feasibility, to some content, of the proposed model on detecting metastasis lesions in low-resolution SPECT bone scan images.

3.2. Ablation study
How the feature extraction and lesion localization & boxing operations impact the model’s detection performance measured by AP is studied in this subsection.

3.2.1. Impact of candidate boxes on AP
As mentioned in subsection 2.2.2, the feature extraction sub-network outputs a group of varied-size feature maps, which relate to the size and number of the candidate boxes. Table 4 lists several different types of candidate boxes, where the scheme S3# represents the one used in previous subsection.

On the different types of candidate boxes, figure 6 provides the scores of AP obtained by the proposed detection model. Using the scheme S3#, the proposed detection model performs best since the nine candidate boxes in this scheme can better cover the lesion areas in feature maps than others. It can also be seen that using the 16 × 16 CB is more suitable than the 19 × 19 one as shown by the relatively inferior performance of scheme S4#.

3.2.2. Impact of ar values on AP
Another factor relating to the detection performance is ar, which is defined as the ratio of width and height. Several groups of empirical values for ar are provided in table 5, with ar_3 indicating the one used previously.

| Metric                | Value |
|-----------------------|-------|
| AP                    | 0.7911|
| Precision             | 0.9130|
| Recall (sensitivity)  | 0.1333|

Figure 5. The P–R curve of the proposed model on the test set.

Table 3. Scores of evaluation metrics obtained by the proposed model on the test set.
The experimental results presented in figure 7 show that using the six various ratios in ar_3 to scale the candidate boxes in scheme S3 as shown in table 4 can appropriately box the metastasis lesions in input images.

Table 4. The types of the candidate boxes.

| Scheme | Type of candidate boxes |
|--------|-------------------------|
| S1#    | \(32 \times 32; 16 \times 16; 11 \times 11; 7 \times 7; 5 \times 5; 3 \times 3; 1 \times 1\) |
| S2#    | \(32 \times 32; 19 \times 19; 11 \times 11; 7 \times 7; 5 \times 5; 3 \times 3; 1 \times 1\) |
| S3#    | \(32 \times 32; 16 \times 16; 13 \times 13; 11 \times 11; 9 \times 9; 7 \times 7; 5 \times 5; 3 \times 3; 1 \times 1\) |
| S4#    | \(32 \times 32; 19 \times 19; 13 \times 13; 11 \times 11; 9 \times 9; 7 \times 7; 5 \times 5; 3 \times 3; 1 \times 1\) |
| S5#    | \(32 \times 32; 19 \times 19; 16 \times 16; 13 \times 13; 11 \times 11; 9 \times 9; 7 \times 7; 5 \times 5; 3 \times 3; 1 \times 1\) |

Table 5. Different groups of values for ar.

| ar     | Value                                      |
|--------|--------------------------------------------|
| ar_1   | \(\{1^*, 1\}\)                             |
| ar_2   | \(\{1^*, 1, 2, 1/2\}\)                     |
| ar_3   | \(\{1^*, 1, 2, 3, 1/2, 1/3\}\)             |
| ar_4   | \(\{1^*, 1, 2, 3, 4, 1/2, 1/3, 1/4\}\)     |
| ar_5   | \(\{1^*, 1, 2, 3, 4, 5, 1/2, 1/3, 1/4, 1/5\}\) |

The experimental results presented in figure 7 show that using the six various ratios in ar_3 to scale the candidate boxes in scheme S3# as shown in table 4 can appropriately box the metastasis lesions in input images.
Table 6. Structures of different feature extraction sub-networks.

| Layer       | VGG-16 | Resnet-18 | Resnet-27 | Ours | Resnet-31 | Resnet-34 |
|-------------|--------|-----------|-----------|------|-----------|-----------|
| Conv        | [3 × 3, 64] × 2 | [3 × 3, 64] × 2 | [3 × 3, 64] × 2 | [3 × 3, 64] × 2 | [3 × 3, 64] × 2 | [3 × 3, 64] × 2 |
| Pool        | [3 × 3, 128] × 2 | [3 × 3, 128] × 2 | [3 × 3, 128] × 2 | [3 × 3, 128] × 2 | [3 × 3, 128] × 2 | [3 × 3, 128] × 2 |
| Conv1_x     | [3 × 3, 256] × 2 | [3 × 3, 256] × 2 | [3 × 3, 256] × 2 | [3 × 3, 256] × 2 | [3 × 3, 256] × 2 | [3 × 3, 256] × 2 |
| Conv2_x     | [3 × 3, 512] × 2 | [3 × 3, 512] × 2 | [3 × 3, 512] × 2 | [3 × 3, 512] × 2 | [3 × 3, 512] × 2 | [3 × 3, 512] × 2 |
| Conv3_x     | [3 × 3, 1, 64] × 2 | [3 × 3, 1, 64] × 2 | [3 × 3, 1, 64] × 2 | [3 × 3, 1, 64] × 2 | [3 × 3, 1, 64] × 2 | [3 × 3, 1, 64] × 2 |
| Conv4_x     | [3 × 3, 1, 128] × 2 | [3 × 3, 1, 128] × 2 | [3 × 3, 1, 128] × 2 | [3 × 3, 1, 128] × 2 | [3 × 3, 1, 128] × 2 | [3 × 3, 1, 128] × 2 |

Configuration

7 × 7, 64, stride = 1
3 × 3 max pool, stride = 1
3.2.3 Impact of network structure on AP

The structure of the feature extraction sub-network is also examined to investigate whether it has an impact on the detection performance. The classical CNNs including VGG (Simonyan and Zisserman 2014) and Resnet (He et al 2016) (and its variants) are used as feature extraction sub-networks to compare with the proposed one, which are outlined in table 6.

The experimental results presented in figure 8 reveal that the proposed 29-layer sub-network using residual connections (i.e. the inter- and intra-connections) can extract more representative features of low-resolution SPECT bone scan images than others, hence the best lesion detection performance.

3.3 Comparative analysis

In our previous work (Lin et al 2020b), CNN-based methods were proposed to segment $^{99m}$Tc-MDP SPECT bone scan images by introducing residual connection and attention mechanism into the classical models U-Net (Ronneberger et al 2015) and Mask R-CNN (He et al 2020), respectively. By adjusting the structures of the segmentation networks to fit the lesion detection task, we examine their performance on lesion detection task using same test set of SPECT bone scan images.

As shown in table 7, the proposed detection method outperforms all segmentation models with the much higher score for AP. This further demonstrates the superiority of the proposed lesion detection model.

4 Discussion

A brief analysis of the strengths and weaknesses of the proposed bone metastasis lesion detection model is presented in this section.

Despite the inferior spatial resolution of SPECT imaging, our model achieves promising lesion detection performance with the mean score of 0.7911 for AP. The proposed model outperforms the others in which different structures of feature extraction sub-network were used, demonstrating the feasibility of the cascaded residual convolution blocks in learning hierarchal representations of low-resolution images. The model detection performance depends also on the amount and shape of valid candidate boxes, where $ar \in \{1', 1, 2, 3, 1/2, 1/3\}$ corresponds to the best detection performance.

The experimental results in table 3 show that the proposed model achieves high detection precision but low recall. According to equation (5), the false negatives and false positives account for decreasing Recall and

![Figure 8. Scores of AP obtained by the proposed detection model when different feature extraction sub-networks are used.](image)

| Model                                      | AP   |
|--------------------------------------------|------|
| U-Net (Ronneberger et al 2015)             | 0.5106 |
| U-Net-Res (Lin et al 2020b)                | 0.5374 |
| Mask R-CNN (He et al 2020)                 | 0.4378 |
| Mask R-CNN-Att (Lin et al 2020b)           | 0.4872 |
| Ours                                       | 0.7911 |
Precision metric, respectively. Figure 9 explains this with three typical cases of detecting multiple lesions of bone metastasis.

For the case in figure 9(a), six lesions (i.e. $l_1$, $l_2$, and $l_5$–$l_8$) were correctly detected, which were assigned with a value of 1 for TP and a value of 0 for FP and FN. Two areas $l_3$ and $l_4$ with slightly high uptake were incorrectly detected as lesions, denoting the false-positive predictions that contribute to decreasing the precision. For this case, the value of Precision is calculated as $\text{Precision} = \sum_{i=1}^{8} [TP_i/(TP_i + FP_i)] = 6/(6 + 2) = 3/4$ and the value of Recall is calculated as $\text{Recall} = \sum_{i=1}^{8} [TP_i/(TP_i + FN_i)] = 6/(6 + 0) = 1$.

For the case in figure 9(b), four lesions (i.e. $l_2$–$l_5$) were correctly detected. A lesion $l_1$ was incorrectly identified as the normal, denoting the false-negative prediction that contributes to decreasing the recall. For this case, the value of Precision is calculated as $\text{Precision} = \sum_{i=1}^{5} [TP_i/(TP_i + FP_i)] = 4/(4 + 0) = 1$ and the value of Recall is calculated as $\text{Recall} = \sum_{i=1}^{5} [TP_i/(TP_i + FN_i)] = 4/(4 + 1) = 4/5$.

Figure 9. Illustration of the detected multiple lesions in three images by the proposed model with the prediction and ground truth being labeled with red box and green box, respectively. (a) The case contains false-positive predictions in a posterior-view image; (b) The case contains false-negative prediction in an anterior-view image; and (c) The case contains false-positive predictions in the organs (i.e. kidneys) in a posterior-view image.
Similarly, we can calculate the values of Precision and Recall for the case in figure 9(c) as Precision = \[ \sum_{i=1}^{5} \frac{TP_i}{(TP_i + FP_i)} = \frac{2}{(2 + 2)} = \frac{1}{2} \] and Recall = \[ \sum_{i=1}^{5} \frac{TP_i}{(TP_i + FN_i)} = \frac{2}{(2 + 1)} = \frac{2}{3}, \] respectively.

We can conclude the reasons of false-negative and false-positive predictions as follows:

- Unique characteristics of SPECT imaging: \(^{99m}\text{Tc} \text{MDP-SPECT} \) is a type of functional medical imaging techniques. Not only the lesion areas but also the normal bones would display hotspots of high uptake of radiopharmaceutical. Furthermore, hotspots can also be seen in the kidneys and bladder for some patients due to the accumulation of radiopharmaceutical (\(^{99m}\text{Tc} \text{-MDP} \)) during the process of excretion. This is why the normal skeletal areas indicated by \( l_1 \) and \( l_4 \) in figure 9(a) and the kidneys indicated by \( l_3 \) and \( l_4 \) in figure 9(c) were incorrectly detected as lesions by the model.

- Nature of bone metastasis: asymmetry of hotspots is a core cue of clinical bone metastasis diagnosis in nuclear medicine practice. An irregular, asymmetric or eccentric radiotracer uptake in bone scans may be towards malignant involvement (Nathan et al 2013). A recent work (Saito et al 2021) studies to develop CNN-based image classification method for automatically identifying the presence or absence of bone metastasis, which takes the left-right asymmetry of hotspots into consideration. Although there are only five images in our dataset contain remarkable higher uptake in kidneys like \( l_3 \) and \( l_4 \) in figure 9(c), how to alleviate or even eliminate their negative effects needs to be further studied. Technological solutions such as the data normalization conducted within varied-length window would be helpful for reducing the false positives caused by symmetric hotspots in both the bone and organs.

- Scarcity of samples: it is a common challenge to build a large enough image dataset to train a model in the medical image analysis field. The performance of deep learning-based analysis, however, often positively relates to the size of the datasets. Limited samples provide no sufficient instances for an object detection model to learn rich representations of metastasis lesions with varied location, shape, and intensity. The situation becomes even much worse when human experts manually label low-resolution SPECT bone scan images to obtain ground truth. This is because manual annotation is very time-consuming, laborious, and subjective.

5. Conclusions

Targeting the automated detection of lung cancer-caused bone metastasis in low-resolution SPECT bone scan images, we proposed a SSD-based automated lesion detection model. The structure of the proposed model has been presented by detailing the processes of extracting hierarchical representations of SPECT bone scan images, locating the spatial location of a lesion, and boxing the identified lesion. Experimental evaluation conducted on clinical data of \(^{99m}\text{Tc} \text{-MDP-SPECT} \) images has demonstrated the feasibility of the proposed model on automated detection of metastasis lesions with low-resolution images. A comparative analysis has also been conducted to show the superiority of the proposed detection network against the classical SSD model and others with different feature extraction sub-networks.

We plan to extend our work in several directions in the future. First, we intend to collect more data of clinical \(^{99m}\text{Tc} \text{-MDP-SPECT} \) bone scan images to test the proposed model, to further improve and optimize the detection network. Second, emerging techniques including image super-resolution and varied-length-window data normalization will be adopted to improve the quality of the raw imaging data as much as possible. Lastly, we attempt to integrate field knowledge (e.g. structural symmetry of the body) into data patterns to develop the knowledge- and data-driven models for high-performance detection of bone metastasis lesions with \(^{99m}\text{Tc} \text{-MDP} \) SPECT imaging data.

Declarations

Funding

This work was supported by the Key R&D Plan of Gansu Province (21YF5GA063), the Youth PhD Foundation of Education Department of Gansu Province (2021QB-063), the Natural Science Foundation of Gansu Province (20J3RA511), the Fundamental Research Fund for the Central Universities (31920220020, 31920220054, 31920210013), the National Natural Science Foundation of China (61562075), the Gansu Provincial First-class Discipline Program of Northwest Minzu University (11080305), and the Program for Innovative Research Team of SEAC (2018)98.
Conflict of interest

Authors declare that they have no conflict of interest.

Availability of data and material

Anyone can get the validation subset by emailing the corresponding author by stating that the data is used for research purposes only. The whole dataset will be publicly available in the future.

Code availability

The software implementing the convolutional neural network can be made available for distribution through a collaboration agreement by contacting the corresponding author.

Author contributions

Conceptualization, Qiang Lin and Xiaoyan Chen; methodology, Qiang Lin, Yongchun Cao, and Xiaoyan Chen; software, Xiaoyan Chen, Liangxia Liu; validation, Qiang Lin, Yongchun Cao, Xiaoyan Chen, and Liangxia Liu; formal analysis, Qiang Lin, Yongchun Cao, and Zhengxing Man; investigation, Qiang Lin and Xiaoyan Chen; resources, Xianwu Zeng; data curation, Xianwu Zeng, Qiang Lin, and Yongchun Cao; writing—original draft preparation, Qiang Lin and Xiaoyan Chen; writing—review and editing, Qiang Lin, Xiaodi Huang; visualization, Xiaoyan Chen; supervision, Qiang Lin and Xiaoyan Chen; project administration, Qiang Lin; funding acquisition, Qiang Lin. All authors have read and agreed to the published version of the manuscript.

Ethics approval

The study was approved by the Ethics Committee of Gansu Provincial Tumor Hospital (Approval No.: A202106100014).

Consent to participate

All procedures performed in this study involving human participants were in accordance with the ethical standards of the Gansu Provincial Tumor Hospital research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The used bone SPECT images were de-identified before the authors received the data. The fully anonymised image data was received by the authors on June 01, 2021. A requirement for informed consent was waived for this study because of the anonymous nature of the data.

Consent for publication

Not applicable.

ORCID iDs

Qiang Lin https://orcid.org/0000-0002-3842-2634

References

Bombardieri E et al 2003 Bone scintigraphy: procedure guidelines for tumor imaging Eur. Nucl. Med. Molecular Imaging 30 99–106
Cheng D C et al 2021a Lesion-based bone metastasis detection in chest bone scintigraphy images of prostate cancer patients using pre-train, negative mining, and deep learning Diagnostics 11 518
Cheng D C et al 2021b Bone metastasis detection in the chest and pelvis from a whole-body bone scan using deep learning and a small dataset Electronics 10 1201
Dang J 2016 Classification in none scintigraphy images using convolutional neural networks Lund Univ.
Dou Q et al 2015 Automatic cerebral microbleeds detection from MR images via independent subspace analysis based hierarchical features Int. Conf. of the IEEE Engineering in Medicine & Biology Society (Milan, Italy)
Guo Y et al 2022 Automated detection of lung cancer–caused metastasis by classifying scintigraphic images using convolutional neural network with residual connection and hybrid attention mechanism Insights into Imaging 13 24
He K et al 2016 Deep residual learning for image recognition IEEE Conf. CVPR 2016 (Las Vegas, NV, 27–30 June 2016) pp 770–8
He K et al 2020 Mask R-CNN IEEE Trans. Pattern Anal. Mach. Intell. 2020 42 386–97
Hwang S and Kim H 2016 Self-transfer learning for fully weakly supervised object localization arXiv:1602.01625
Li T et al 2022 Automated detection of skeletal metastasis of lung cancer with bone scan using convolutional neural network Phys. Med. Biol. 67 015004
Lin Q et al 2020a Classifying functional nuclear images with convolutional neural networks: a survey IET Image Proc. 14 3300–13
Lin Q et al 2020b Deep learning based automatic segmentation of metastasis hotspots in thorax bone SPECT images PLoS One 15 e0243253
Lin Q et al 2021a Deep learning based automated diagnosis of bone metastases with SPECT thoracic bone images Sci. Rep. 11 4223
Lin Q et al 2021b Multiclass classification of whole-body scintigraphic images using a self-defined convolutional neural network with attention modules Med. Phys. 48 5782–93
Lin Q et al 2021c dSPIC: a deep SPECT image classification network for automated multi–disease, multi–lesion diagnosis BMC Med. Imaging 21 122
Litjens G et al 2017 A survey on deep learning in medical image analysis Med. Image Anal. 42 60–88
Liu W et al 2016 SSD: single shot multibox detector European Conf. on Computer Vision (Cham) (Springer) 21–37
Lo S B and Lou S A 1995 Artificial convolution neural network techniques and applications for lung nodule detection IEEE Trans. Med. Imaging 14 711–8
Nathan M et al 2013 Bone Scintigraphy: Patterns, Variants, Limitations and Artefacts (Berlin Heidelberg: Springer)
Papandrianos N et al 2020a Bone metastasis classification using whole body images from prostate cancer patients based on convolutional neural networks application PLoS One 15 e0237213
Papandrianos N et al 2020b Efficient none metastasis diagnosis in bone scintigraphy using a fast convolutional neural network architecture Diagnostica 10 532
Papandrianos N et al 2020d A deep-learning approach for diagnosis of metastatic breast cancer in bones from whole-body scans Appl. Sci. 10 997
Papandrianos N, Papageorgiou E and Anastasios A 2020c Development of convolutional neural networks to identify bone metastasis for prostate cancer patients in bone scintigraphy Ann. Nucl. Med. 34 824–32
Pi Y et al 2020 Automated diagnosis of bone metastasis based on multi-view bone scans using attention–augmented deep neural networks Med. Image Anal. 65 101784
Raiipurkar P et al 2018 Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists PLoS Med. 15 1–17
Ronneberger O, Fischer P and Brox T 2015 U-Net: Convolutional networks for biomedical image segmentation 18th Conf. MICCAI 2015 (Munich, Germany, 3–9 October 2015) pp 234–41
Rosenfeld A and Thurston M 1971 Edge and curve detection for visual scene analysis IEEE Trans. Comput. 20 562–9
Roth H R et al 2016 Improving computer–aided detection using convolutional neural networks and random view aggregation IEEE Trans. Med. Imaging 35 1170–81
Roth H R et al 2014 A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations Med. Image Comput. Assist. Intervent. Intervention 17 520–7
Saito A et al 2021 Extraction of metastasis hotspots in a whole–body bone scintigram Int. J. Comput. Assist. Radiol. Surg. 16 2251–60
Simonyan K and Zisserman A 2014 Very deep convolutional networks for large-scale image recognition arXiv:1409.1556
Teramoto A et al 2016 Automated detection of pulmonary nodules in PET/CT images: ensemble false–positive reduction using a convolutional neural network technique Med. Phys. 43 2821–7
Yao J et al 2015 Detection of sclerotic spine metastases via random aggregation of deep convolutional neural network classifications Springer Int. Publ. 3–12
Zhao Z et al 2020 Deep neural network based artificial intelligence assisted diagnosis of bone scintigraphy for cancer bone metastasis Sci. Rep. 10 17046