US 20200410671A1

(19) **United States**

(12) **Patent Application Publication**

PENG et al.

(10) **Pub. No.: US 2020/0410671 A1**

(43) **Pub. Date: Dec. 31, 2020**

(54) **CT LYMPH NODE DETECTION SYSTEM BASED ON SPATIAL-TEMPORAL RECURRENT ATTENTION MECHANISM**

(71) Applicant: **SHANDONG UNIVERSITY OF SCIENCE AND TECHNOLOGY**, Qingdao (CN)

(72) Inventors: **Haixin PENG**, Qingdao (CN); **Yingran MA**, Qingdao (CN); **Yuanhong WANG**, Qingdao (CN); **Yanjun PENG**, Qingdao (CN); **Xinming LU**, Qingdao (CN)

(21) Appl. No.: **16/957,305**

(22) PCT Filed: **Oct. 21, 2019**

(86) PCT No.: **PCT/CN2019/112219**

§ 371 (c)(1),
(2) Date: **Jun. 23, 2020**

(30) **Foreign Application Priority Data**

Jun. 28, 2019 (CN) 201910570878.1

Publication Classification

(51) **Int. Cl.**

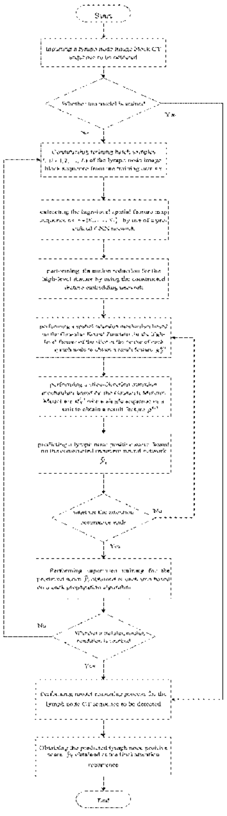
G06T 7/00	(2006.01)
G06T 7/70	(2006.01)
G06N 3/08	(2006.01)
G06N 7/00	(2006.01)
G06N 20/10	(2006.01)

(52) **U.S. Cl.**

CPC ***G06T 7/0012*** (2013.01); ***G06T 7/70*** (2017.01); ***G06N 3/084*** (2013.01); ***G06N 7/005*** (2013.01); ***G06N 20/10*** (2019.01); *G06T 2207/20021* (2013.01); *G06T 2207/30096* (2013.01); *G06T 2207/20084* (2013.01); *G06T 2207/20076* (2013.01); *G06T 2207/20081* (2013.01); *G06T 2207/10081* (2013.01)

(57) **ABSTRACT**

The present disclosure discloses a CT lymph node detection system based on a spatial-temporal recurrent attention mechanism and specifically relates to the field of medical image analysis technologies. Based on a deep convolutional neural network and a recurrent attention mechanism, the present disclosure can construct an attention feature map adaptive to a lesion size in a slice direction and a spatial direction of a lymph node CT sequence. Firstly, a high-level spatial feature corresponding to the lymph node CT image is extracted by use of a pre-trained convolutional network; secondly, a recurrent attention mechanism based on a Gaussian Kernel Function is constructed with a slice at the center of the lymph node as a reference in a spatial domain; based on this, a temporal (slice direction) attention mechanism based on a Gaussian Mixture Model is performed; in addition, a predicted attention position is constrained based on the prior information of position distribution of the lymph node in the CT slice sequence; finally, in combination with the high-level features extracted by the two attention methods, the recurrent neural network performs classification to obtain a lymph node detection result.


Start
Inputting a lymph node image block CT sequence to be detected
Whether the model is trained?
Yes
No
Constructing training batch samples
of the lymph node image block sequence from the training data set
extracting the high-level spatial feature map sequence by use of a pre-trained CNN network
performing dimension reduction for the high-level feature by using the constructed feature embedding network
performing a spatial attention mechanism based on the Gaussian Kernel Function for the high-level feature of the slice at the center of each lymph node to obtain a result feature
performing a slice-direction attention mechanism based on the Gaussian Mixture Model for with a single sequence as a unit to obtain a result feature
predicting a lymph node positive score based on the constructed recurrent neural network
whether the attention recurrence ends
No
Yes
Performing supervised training for the predicted score obtained at each step based on a back propagation algorithm
No
Whether a training ending condition is reached
Yes
Performing model reasoning process for the lymph node CT sequence to be detected
Obtaining the predicted lymph node positive score obtained at the final attention recurrence
End

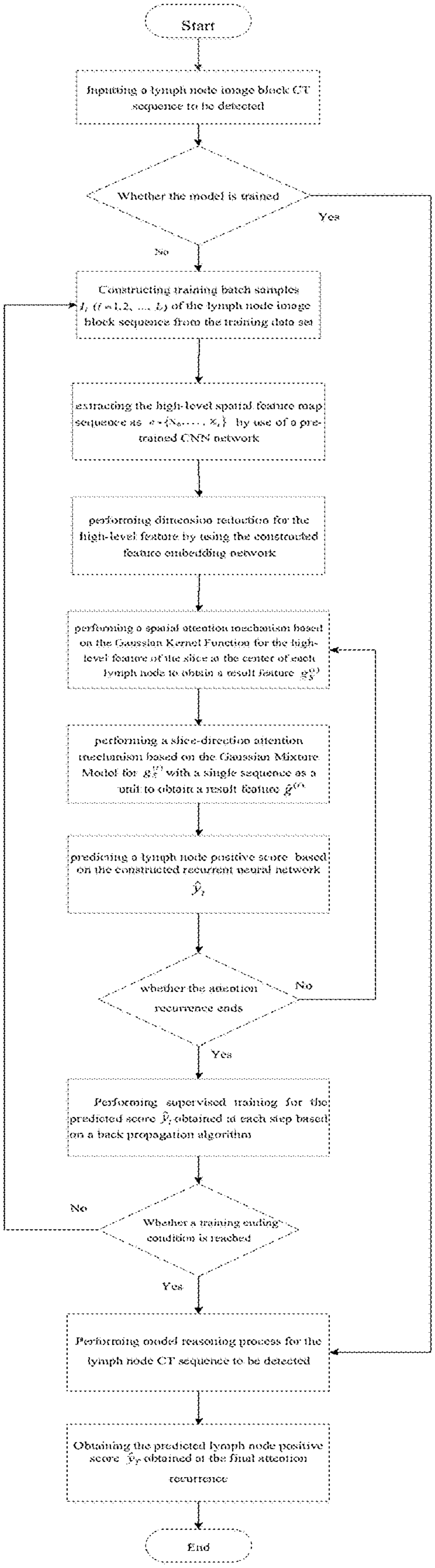

Start
Inputting a lymph node image block CT sequence to be detected
Whether the model is trained
Yes
No
Constructing training batch samples I_i (i=1,2, ..., L) of the lymph node image block sequence from the training data set
extracting the high-level spatial feature map sequence as ϵ={X_0, ..., X_t} by use of a pre-trained CNN network
performing dimension reduction for the high-level feature by using the constructed feature embedding network
performing a spatial attention mechanism based on the Gaussian Kernel Function for the high-level feature of the slice at the center of each lymph node to obtain a result feature $g_S^{(t)}$
performing a slice-direction attention mechanism based on the Gaussian Mixture Model for $g_S^{(t)}$ with a single sequence as a unit to obtain a result feature $\hat{g}^{(t)}$
predicting a lymph node positive score based on the constructed recurrent neural network $\hat{y}_t$
whether the attention recurrence ends
No
Yes
Performing supervised training for the predicted score $\hat{y}_t$ obtained at each step based on a back propagation algorithm
No
Whether a training ending condition is reached
Yes
Performing model reasoning process for the lymph node CT sequence to be detected
Obtaining the predicted lymph node positive score $\hat{y}_T$ obtained at the final attention recurrence
End

FIG. 1

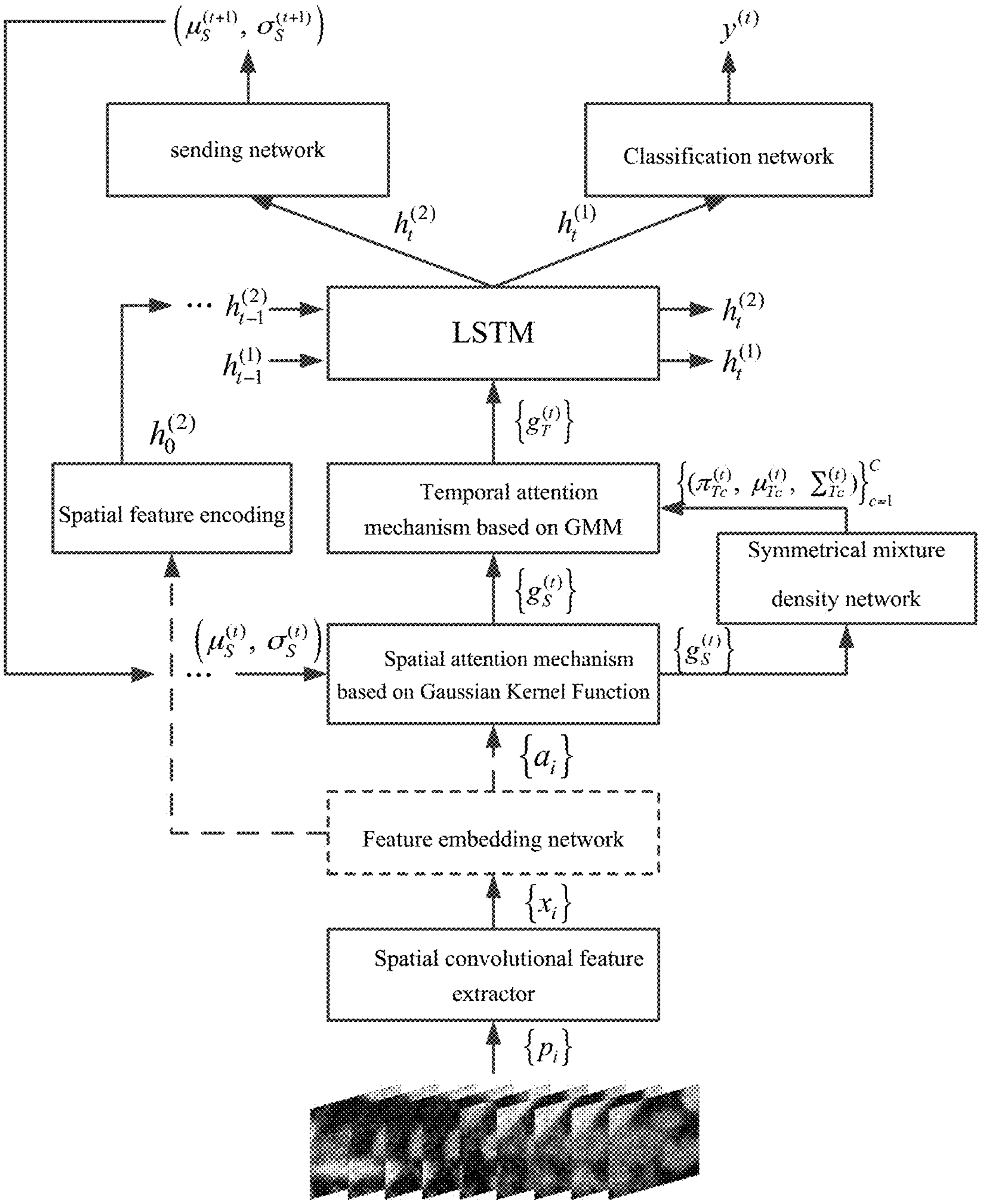

$\left(\mu_S^{(t+1)}, \sigma_S^{(t+1)}\right)$
$y^{(t)}$
sending network
Classification network
$h_t^{(2)}$
$h_t^{(1)}$
$\cdots$ $h_{t-1}^{(2)}$
$h_{t-1}^{(1)}$
LSTM
$h_t^{(2)}$
$h_t^{(1)}$
$\left\{g_T^{(t)}\right\}$
$h_0^{(2)}$
Spatial feature encoding
Temporal attention mechanism based on GMM
$\left\{\left(\pi_{Tc}^{(t)}, \mu_{Tc}^{(t)}, \Sigma_{Tc}^{(t)}\right)\right\}_{c=1}^{C}$
Symmetrical mixture density network
$\left\{g_S^{(t)}\right\}$
$\left(\mu_S^{(t)}, \sigma_S^{(t)}\right)$
$\cdots$
Spatial attention mechanism based on Gaussian Kernel Function
$\left\{g_S^{(t)}\right\}$
$\left\{a_i\right\}$
Feature embedding network
$\left\{x_i\right\}$
Spatial convolutional feature extractor
$\left\{p_i\right\}$

FIG.2

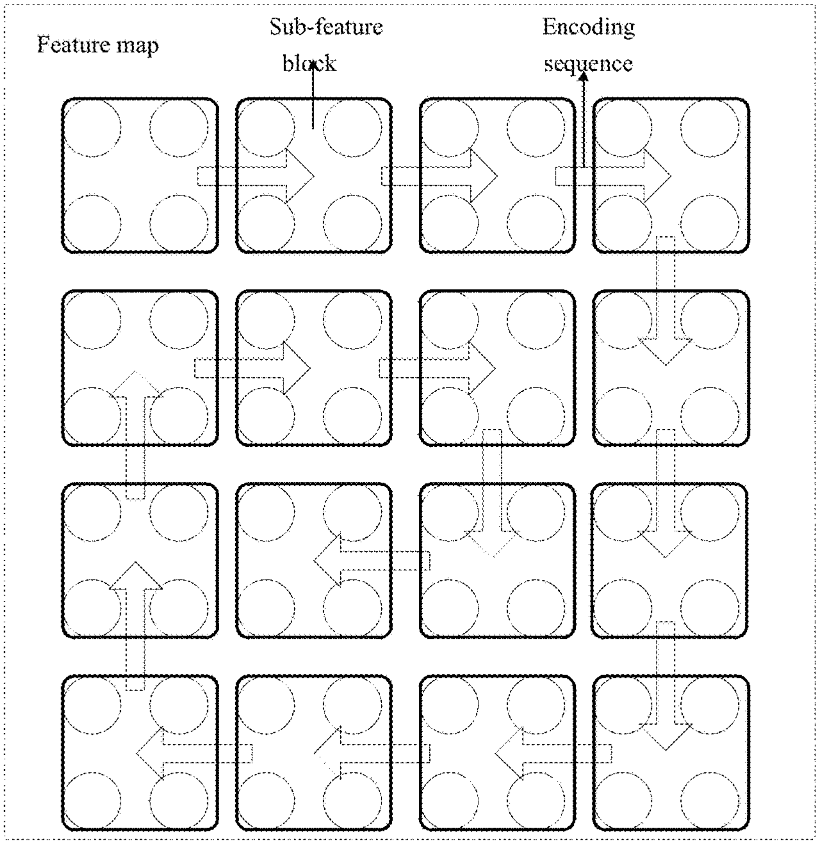

Feature map
Sub-feature block
Encoding sequence

FIG.3

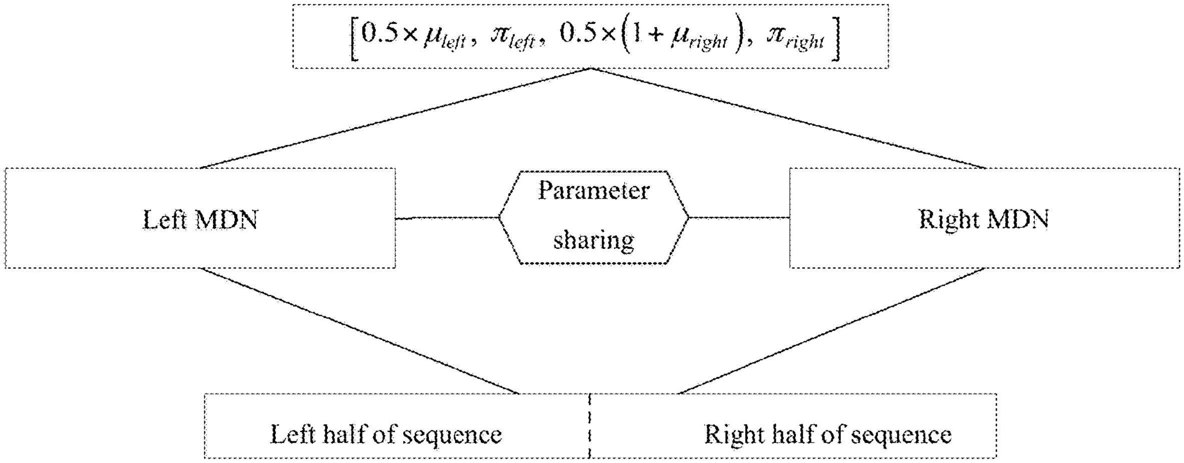

$\left[0.5\times\mu_{left},\ \pi_{left},\ 0.5\times\left(1+\mu_{right}\right),\ \pi_{right}\right]$
Left MDN
Parameter sharing
Right MDN
Left half of sequence
Right half of sequence

FIG.4

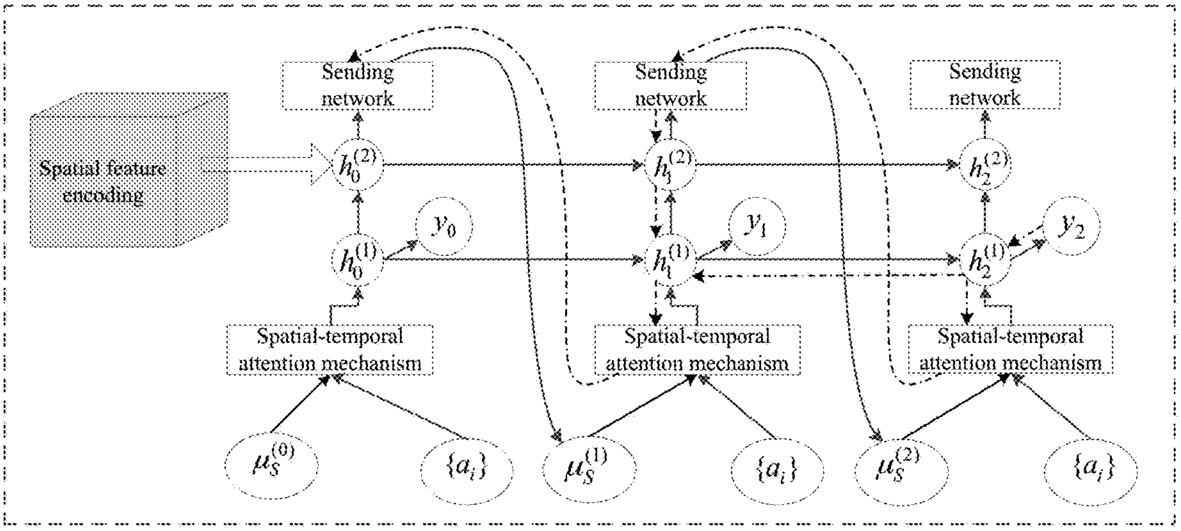

Sending network
Sending network
Sending network
Spatial feature encoding
$h_0^{(2)}$
$h_1^{(2)}$
$h_2^{(2)}$
$h_0^{(1)}$
$h_1^{(1)}$
$h_2^{(1)}$
y_0
y_1
y_2
Spatial-temporal attention mechanism
Spatial-temporal attention mechanism
Spatial-temporal attention mechanism
$\mu_S^{(0)}$
$\{a_i\}$
$\mu_S^{(1)}$
$\{a_i\}$
$\mu_S^{(2)}$
$\{a_i\}$

FIG.5

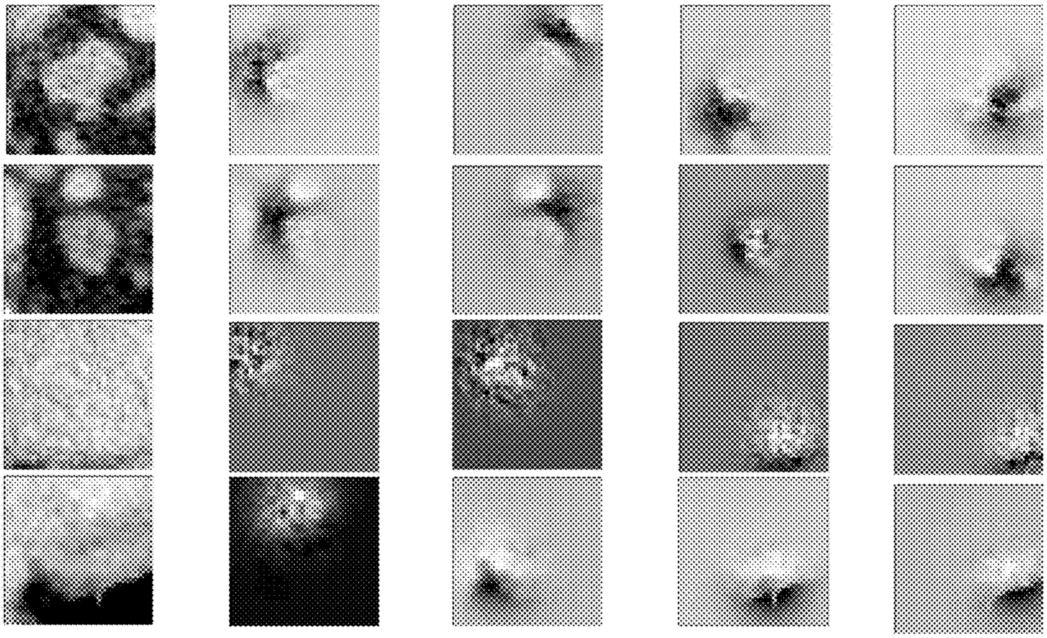


FIG.6

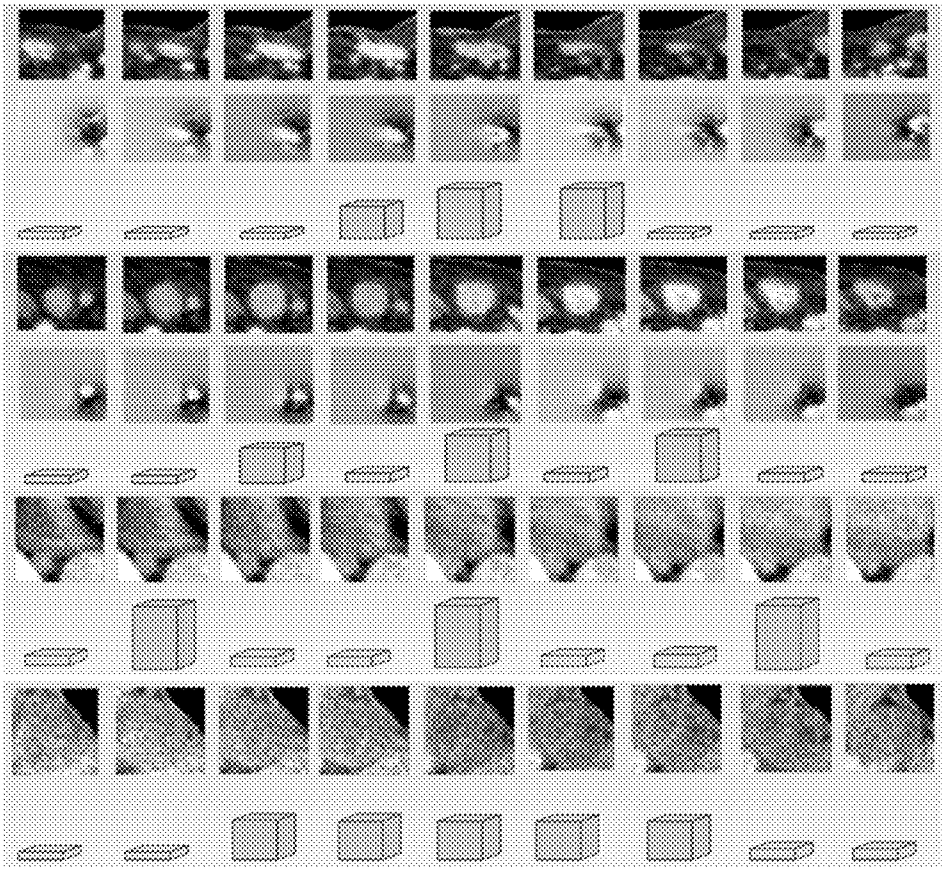
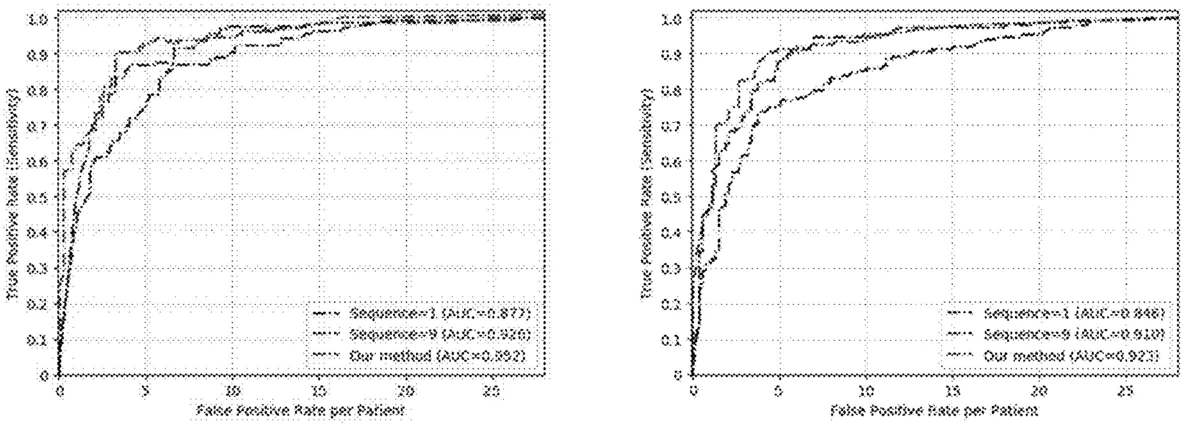


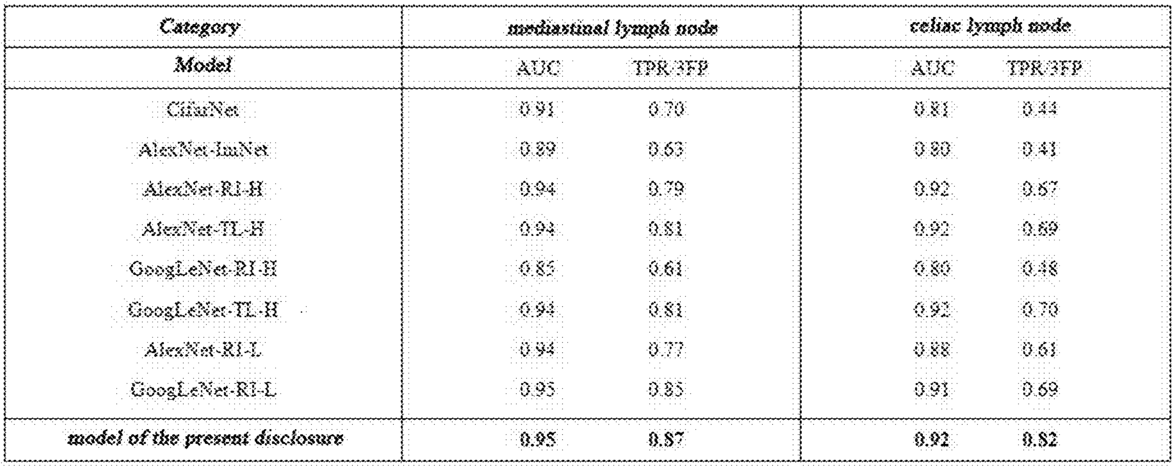
FIG.7


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
True Positive Rate (Sensitivity)
0
5
10
15
20
25
False Positive Rate per Patient
Sequence=1 (AUC=0.877)
Sequence=9 (AUC=0.920)
Our method (AUC=0.892)
Sequence=1 (AUC=0.848)
Sequence=9 (AUC=0.910)
Our method (AUC=0.923)

(a)

(b)

FIG.8



Category	*mediastinal lymph node*		*celiac lymph node*	
Model	AUC	TPR/3FP	AUC	TPR/3FP
CifarNet	0.91	0.70	0.81	0.44
AlexNet-ImNet	0.89	0.63	0.80	0.41
AlexNet-RI-H	0.94	0.79	0.92	0.67
AlexNet-TL-H	0.94	0.81	0.92	0.69
GoogLeNet-RI-H	0.85	0.61	0.80	0.48
GoogLeNet-TL-H	0.94	0.81	0.92	0.70
AlexNet-RI-L	0.94	0.77	0.88	0.61
GoogLeNet-RI-L	0.95	0.85	0.91	0.69
model of the present disclosure	**0.95**	**0.87**	**0.92**	**0.82**

FIG.9

CT LYMPH NODE DETECTION SYSTEM BASED ON SPATIAL-TEMPORAL RECURRENT ATTENTION MECHANISM

TECHNICAL FIELD

[0001] The present disclosure relates to the field of medical image analysis technologies, and in particular to a CT lymph node detection system based on spatial-temporal recurrent attention mechanism.

BACKGROUND

[0002] Considering the special imaging features of medical images, when a deep learning model is applied to the medical image analysis field, the model is usually influenced by different sizes, shapes, dimensions, imaging qualities, and background tissue organs and so on of medical analysis target. Therefore, the analysis accuracy is to be improved.

[0003] A visual attention mechanism may simulate a specific attention mechanism of human eyes for visual information and can accurately locate a region of interest, thereby further solving the influence of irrelevant information on the model. In view of this, the present disclosure applies a conventional deep visual attention model oriented to natural images to lymph node CT sequence images to perform true and false positive classification for a lymph node so as to improve the accuracy of the deep learning method in detection of lesions.

SUMMARY

[0004] To overcome the shortcomings as above, the present disclosure provides a CT lymph node detection system based on recurrent attention mechanism in a visual attention process fusing two dimensions, i.e. space direction and slice direction based on Gaussian Kernel Function and mixture density network in accordance with sequence features of CT images.

[0005] The present disclosure adopts the following technical solution.

[0006] The CT lymph node detection system based on spatial-temporal recurrent attention mechanism includes a training sample extracting module, a deep layer feature extracting network, a feature embedding network, and a spatial-temporal recurrent attention target detection module. The detection includes the following steps.

[0007] At step 1, position coordinate information is marked for the obtained lymph node dcm-format file and a corresponding lymph node by use of the training sample extracting module, and a CT slice image block sequence I_i (i=1,2, . . . ,L), $I_i \in^{W\times H}$ with CT slice image blocks being length L, width W and height H is extracted for each lymph node by using a pydicom module in python.

[0008] At step 2, a high-level spatial feature map sequence corresponding to the CT slice image block sequence of each lymph node is extracted by using a VGG-16 model pre-trained by the natural image according to the deep layer feature extracting network, and denoted as $\{X_0, \ldots, X_L\}$.

[0009] At step 3, the feature embedding network is constructed to perform dimension reduction for the input high-level feature map sequence and output a feature map A_i.

[0010] At step 4, a spatial-temporal recurrent attention frame is constructed, and a spatial attention mechanism is performed based on a recurrent neural network and the Gaussian Kernel Function to obtain a spatial attention result $g_S^{(t)}$.

[0011] At step 5, a temporal attention mechanism is performed for the spatial attention result $g_S^{(t)}$ obtained at step 4 to obtain a spatial-temporal attention feature $\hat{g}^{(t)}$.

[0012] At step 6, a lymph node positive score $\hat{y}_t$ of the recurrent attention iteration step is predicted by using the recurrent neural network constructed at step 4 in combination with the spatial-temporal attention feature $\hat{g}^{(t)}$ obtained at step 5.

[0013] At step 7, a loss function of the model is constructed to perform steps 4-6 for T times, and supervised training is performed for the model by using a gradient back propagation algorithm.

[0014] At step 8, iterative training is performed for the model by repeating steps 3-7, until a trained model is obtained at the end of training.

[0015] At step 9, the lymph node CT sequence to be detected is input to perform a model reasoning process, and a positive score $\hat{y}_T$ output by the final recurrent attention is taken as a CT lymph node detection result.

[0016] Preferably, the step 4 specifically includes the following steps:

[0017] At step 4.1, a long short-term memory network of two layers is constructed.

[0018] At step 4.2, the state of the long short-term memory network is initialized by constructing an encoding process of the feature map.

[0019] At step 4.3, a spatial attention position within a range of the feature map is predicted by using a sending network.

[0020] At step 4.4, an attention matrix $1^{(t)}$ is constructed based on a two-dimension Gaussian Kernel Function.

[0021] At step 4.5, $1^{(t)}$ is multiplied by A, element by element and then added up so as to obtain the spatial attention result $g_S^{(t)}$.

[0022] Preferably, at step 4.1, the recurrent neural network is constructed based on the long short-term memory network of two layers and a hidden layer is expressed as in the formula (1) below.

$$h_t^{(1)}=R_{recur}(\hat{g}^{(t)},h_{t-1}^{(1)}|W_{r1})$$

$$h_t^{(2)}=R_{recur}(h_t^{(1)},h_{t-1}^{(2)}|W_{r2}) \tag{1}$$

[0023] In the above formula, $R_{recur}(\bullet)$ corresponding to $h_t^{(1)}$ is expanded as the following formula (2):

$$\begin{pmatrix} i_t \\ f_t \\ o_t \\ g_t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} M\begin{pmatrix} h_{t-1}, \\ \hat{g}^{(t)} \end{pmatrix}, c_t = f_t e c_{t-1} + i_t e g_t, \tag{2}$$

$$h_t = o_t e \tanh(c_t)$$

[0024] In the above formula, i_t, f_t and o_f represent an input gate, a forget gate and an output gate respectively; c_t and li_t represent a cell state and a hidden layer state respectively; $\hat{g}^{(t)}$ represents a feature vector input into the long short-term memory network by the t-th step of attention mechanism; M: $a \rightarrow b$ is an affine transformation composed of trainable parameters, a=d+E, b=4d; d is the same vector dimension corresponding to i_t, f_t, o_t, g_t, c_t and h_t, E is a dimension of an input feature subjected to dimension reduction, and $R_{recur}(\bullet)$ corresponding to $h_t^{(2)}$ is obtained by replacing $\hat{g}^{(t)}$ in the above $R_{recur}(\bullet)$ with $h_t^{(1)}$.

[0025] Preferably, step 4.2 specifically includes the following steps:

[0026] At step 4.2.1, a new double-layer long short-term memory network (LSTM) having the same structure as formula (1) is constructed.

[0027] At step 4.2.2, the feature map A_{mid} corresponding to the exact center of a CT slice sequence of each lymph node at step 3 is divided according to a spatial neighborhood; specifically, 8×8×200 are divided into 16 sub-feature blocks with 2×2×200 based on adjacent four positions as one group.

[0028] At step 4.2.3, the 16 sub-feature blocks are input into the new double-layer long short-term memory network sequentially clockwise from outside to inside and subjected to 16 cycles to obtain a cell state $c'_T{}^{(2)}$ corresponding to the second layer of the LSTM at the last moment so as to initialize the cell state $c_0^{(2)}$ of the second layer of the long short-term memory network at step 4.1.

[0029] Preferably, the step 4.3 specifically includes the following steps:

[0030] At step 4.3.1, a feature vector $h_t^{(2)}$ output by the first hidden layer of the long short-term memory network and a feature result $g_{S,Center}^{(t)}$ corresponding to the center of the slice sequence in the recurrent attention iteration step are concatenated to obtain $[h_t^{(2)}, g_{S,Center}^{(t)}]$.

[0031] At step 4.3.2, $[h_t^{(2)}, g_{S,Center}^{(t)}]$ is input to the sending network composed of one fully-connected layer to perform regression for the spatial attention position of the next recurrent iteration step as shown in the formula (3):

$$(\mu_S^{(t+1)}, \sigma_S^{(t+1)})=\sigma(W_S[h_t^{(2)}, g_{S,Center}^{(t)}]+b_X) \tag{3}$$

[0032] In the above formula, $\mu_S^{(t+1)}=(\mu_{S,X}^{(t+1)}, \mu_{S,y}^{(t+1)})$ represents an attention position coordinate predicted at the t+1 moment, where a variance $\sigma_S^{(t+1)}$ of the Gaussian Kernel Function is set to a fixed value 0.1, thus, $W_S\in^{2\times(R+E)}$, $b_S\in^{2\times1}$ only corresponds to two output nodes.

[0033] Preferably, at step 4.4, after an attention position coordinate $\mu_S^{(t)}$ is obtained, an attention template matrix is constructed based on the two-dimension Gaussian Kernel Function and softmax as shown in the following formula (4):

$$\begin{aligned} l_i^{(t)} &= p(L_t = i|\mu_S^{(t)}, \sigma_S^{(t)}) \\ &= softmax(\phi(l_i \mid \mu_S^{(t)}, \sigma_S^{(t)})) \\ &= \frac{\exp(\phi(l_i \mid \mu_S^{(t)}, \sigma_S^{(t)}))}{\sum_{j=1}^{K^2} \exp(\phi(l_j \mid \mu_S^{(t)}, \sigma_S^{(t)}))} \end{aligned} \tag{4}$$

[0034] In the above formula, $\phi(l_i|\mu_S^{(t)}, \sigma_S^{(t)})=C\cdot\exp(l_i-\mu_S^{(t)})/2\sigma_S^{(t)}$, $l_i=\{(x_i, y_l)\}_{j=1}^{K\times K}$ is discrete equidistant position coordinates normalized to the interval [0, 1] within the range A_i; C is Gaussian normalized constant which is a fixed value 10.

[0035] Preferably, the step 5 specifically includes the following steps:

[0036] At step 5.1, a mixture density network is constructed to predict an attention position $\mu_T^{(t)}$ of a slice direction.

[0037] At step 5.2, an attention weight vector $l'^{(t)}$ obtained based on Gaussian Mixture Distribution.

[0038] At step 5.3, $l'^{(t)}$ and the input feature $g_S^{(t)}$ are multiplied element by element and added up to obtain the spatial-temporal attention feature $\hat{g}^{(t)}$.

[0039] Preferably, the step 5.1 specifically includes the following steps:

[0040] At step 5.1.1, the sequence feature $g_S^{(t)}$ corresponding to each lymph node is equally divided into left and right halves, that is $g_{Left}^{(t)}$ and $g_{Right}^{(t)}$.

[0041] At step 5.1.2, the mixture density network is formed by one fully-connected hidden layer to perform regression for the temporal attention position coordinate and the Mixture Gaussian Function parameters as shown in the following formula (5):

$$Z^{(t)}=\sigma(W_T g_{S/2}^{(t)}+b_T) \tag{5}$$

[0042] In the above formula, $g_{S/2}^{(t)}\in^{E\cdot L/2}$ represents the left half or the right half of $g_S^{(t)}$; $W_T\in^{2C\times E\cdot L/2}$, $b_T\in^{2C\times1}$ represents a training parameter of the mixture density network, C represents a component number of the Gaussian Mixture Model respectively corresponding to the left half and the right half of the sequence feature, and σ is sigmoid function.

[0043] At step 5.1.3, respective regression coefficients $Z_{Left}^{(t)}\in^{2C\times1}$ and $Z_{Right}^{(t)}\in^{2C\times1}$ of the left half and the right half are both distributed to the Gaussian Mixture Model according to a rule of $Z_{0:\,C-1}^{(t)}\rightarrow\pi_T^{(t)}$ and $Z_{C:\,2C-1}^{(t)}\rightarrow\mu_T^{(t)}$.

[0044] At step 5.1.4, with the formula (6), μ_{Left} is limited to [0, 0.5) and μ_{Right} is limited to (0.5, 1].

$$\mu_{Left}\leftarrow\mu_{Left}\times0.5,\ \mu_{Right}\leftarrow0.5\times(1+\mu_{Right}) \tag{6}$$

[0045] Preferably, the step 5.2 specifically includes the following steps:

[0046] At step 5.2.1, the coefficient $\pi_T^{(t)}$ of each Gaussian component is normalized with softmax, that is, $\pi_T^{(t)}\leftarrow softmax(\pi_T^{(t)})$, and the variance $\Sigma_T^{(t)}$ of all Gaussian components is fixed as 0.1.

[0047] At step 5.2.2, in combination with the Gaussian Mixture Model parameters $(\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})$ derived from the slice feature vectors of the left and right halves, a corresponding temporal attention weight vector $l'_i{}^{(t)}$ is obtained based on the formula (7).

$$\begin{aligned} l_i'^{(t)} &= p(L_t = i \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}) \\ &= softmax(N(l_i \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})) \\ &= \frac{\exp(N(l_i \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}))}{\sum_{j=1}^{L} \exp(N(l_j \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}))} \end{aligned} \tag{7}$$

[0048] In the above formula, N is a Gaussian mixture density function.

$$N(l_i|\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})=\Sigma_{c=1}^{C}\pi_{T,c}^{(t)}\phi_c(l_i,\mu_{T,c}^{(t)},\Sigma_{T,c}^{(t)}),\ s.t.\Sigma_{c=1}^{C}\pi_{T,c}^{(t)}=1\phi_c(l_i|\mu_{T,c}^{(t)},\Sigma_{T,c}^{(t)})=C\cdot\exp(l_i-\mu_{T,c}^{(t)})/2\Sigma_{T,c}^{(t)} \tag{8}$$

[0049] Preferably, at step 7, a target function for the model to receive supervised training is obtained in the following steps:

[0050] At step 7.1, a classification loss function is constructed; and cross-entropy loss functions corresponding to all recurrent attention iteration steps are investigated as shown in the following formula (9):

$$F_c = -\sum_{t=1}^{T}\sum_{i=1}^{2} y_i \log\hat{y}_{t,i} \tag{9}$$

[0051] In the above formula, y_i and $\hat{y}_{t,i}$ are a real lymph node positive score and a predicted lymph node positive score, which are a total number of the recurrent attention iterations.

[0052] At step 7.2, a constraint term of a predicted position of the spatial attention mechanism is constructed in the following two steps:

[0053] At step 7.2.1, a "convergent" constraint term of the position is constructed, and the predicted spatial attention position is constrained to around the center of the feature map based on Batch Normalization and the prior information of the lymph node being at the center of the slice, as shown in the following formula (10):

$$\min(\Sigma_{i=1}{}^{d}(\gamma_i-1)^2+\beta_i{}^2) \tag{10}$$

[0054] In the above formula, $\hat{y}_i \leftarrow \gamma\hat{x}_i+\beta$, (i=1, . . . , m),

$$\hat{x}_i \leftarrow \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}} \Leftrightarrow \hat{X} = A^T(X-B)$$

[0055] In the above formulas, β and γ are trainable parameters introduced by the Batch Normalization; d is an output dimension of the sending network; and m is a batch sample capacity.

[0056] At step 7.2.2, a "divergent" constraint term of the position is constructed; the uniformly-distributed cross-entropy loss functions are constructed according to different positions predicted in the entire recurrent attention process, as shown in the following formula (11):

$$F_S=\text{Const}-\Sigma_{t=1}{}^{T}P_{Unif}\log[\text{softmax}(\mu_S{}^{(t)})] \tag{11}$$

[0057] In the above formula, P_{Unif}: Uniform $(\mathbf{0}, K^2)$ represents a uniform distribution within a spatial range of the feature map; the constant Const is set to 1.

[0058] At step 7.3, a constraint term of a predicted position of the temporal attention mechanism is constructed; the "convergent" constraint term is applied to the predicted position, the batch normalization layer is also introduced to the mixture density network and the formula (10) is minimized so that the attention positions of the left and right halves in the direction of slice $\mu_T{}^{(t)}\sim(0.25, 0.75)$.

[0059] At step 7.4, a final entire loss function of the model is constructed as shown in the following formula (12):

$$F = -\sum_{t=1}^{T}\sum_{i=1}^{2} y_i\log\hat{y}_{t,i} + \lambda_1\left(\sum\nolimits_{i=1}^{d}(\gamma_i-1)^2+\beta_i^2\right) + \lambda_2\left(Const-\sum\nolimits_{t=1}^{T} p_{Unif}\log[softmax(\mu_S^{(t)})]\right) \tag{12}$$

[0060] The present disclosure has the following beneficial effects.

[0061] In the present disclosure, the lymph node CT image high-level spatial feature is weight-extracted by use of the recurrent attention mechanism to self-adaptively process 3D lesion voxel information without introducing external position label information. The present disclosure is based on the Gaussian Kernel Function with the model being small in scale, derivable entirely and easy to train with back propagation algorithm. The present disclosure brings an ideal lymph node detection result and thus has good technical value and good application prospect.

BRIEF DESCRIPTION OF THE DRAWINGS

[0062] FIG. **1** is a flowchart of a method according to the present disclosure.

[0063] FIG. **2** is a structural diagram of each part of the present disclosure.

[0064] FIG. **3** is a schematic diagram of a code sequence of a feature map.

[0065] FIG. **4** is a structural diagram of symmetrical mixed density networks.

[0066] FIG. **5** is a schematic diagram of backward and forward propagation of a recurrent neural network during a training stage.

[0067] FIG. **6** is an instance diagram of performing a spatial attention mechanism.

[0068] FIG. **7** is an instance diagram of performing a temporal attention mechanism.

[0069] FIG. **8** is a diagram of comparison of FROCs and AUC of the present disclosure and a reference model.

[0070] FIG. **9** is a diagram of performance comparison of the present disclosure and other models.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0071] The specific implementations of the present disclosure will be further described below in combination with accompanying drawings and specific examples.

[0072] To overcome the influence of different lymph node lesion regions and complex background on a detection result, the present disclosure provides a detection system based on recurrent attention mechanism in a visual attention process fusing two dimensions, i.e. space direction and slice direction based on Gaussian Kernel Function and mixture density network in accordance with sequence features of CT images. Further, the predicted spatial-temporal attention position is constrained based on prior distribution of lesion region positions.

[0073] In combination with the accompanying drawings, there is provided a CT lymph node detection system based on a spatial-temporal recurrent attention mechanism, including a training sample extracting module, a deep layer feature extracting network, a feature embedding network and a spatial-temporal recurrent attention target detection module. The detection process includes the following steps:

[0074] At step 1, position coordinate information is marked for the obtained lymph node dcm-format file and a corresponding lymph node by use of the training sample extracting module, and a CT slice image block sequence I_i (i=1,2, . . . , L), $I_i \in {}^{w\times H}$, with CT slice image blocks being length L, width W and height H is extracted for each lymph node by using a pydicom module in python.

[0075] A data set used herein is a CT lymph node data set provided by The Cancer Imaging Archive (TCIA). The data set contains 389 mediastinal lymph nodes of 90 patients, 3208 false positive samples corresponding to the mediastinal lymph nodes, 595 celiac lymph nodes of 86 patients and 3484 false positive samples corresponding to the celiac lymph nodes. According to the prior statistic information of the CT lymph nodes, the extracted CT slice sequence length corresponding to each lymph node: L=9, that is, a sequence is formed by extracting four adjacent slices forward and backward respectively with the center of each lymph node as a symmetrical axis. Further, W=H=32 is fixed. In the spatial domain, the center of the image block corresponds to the position coordinate marked for the CT lymph node.

[0076] At step 2, a high-level spatial feature map sequence corresponding to the CT slice image block sequence of each lymph node is extracted by using a VGG-16 model pre-trained by a natural image according to the deep layer feature extracting network, and denoted as $\{X_0, \ldots, X_L\}$.

[0077] The pre-trained CNN model is an Inception-V3 model trained based on an ImageNet data set. Further, the penultimate convolutional layer of the high-level feature is selected and the extracted feature dimensions are 8×8×1280.

[0078] At step 3, the feature embedding network is constructed to perform dimension reduction for the input high-level feature map sequence and output a feature map A_i.

[0079] The constructed feature embedding network is composed of one fully-connected layer, i.e. 1×1 convolution, and the output dimension is set to 200. A dimension reduction process of a single feature map is as shown in the following formula (13):

$$A_i=\tanh(W_{Em}X_i+b) \tag{13}$$

[0080] In the above formula, $W_{Em}\in^{E\times D}$, $b\in^{E\times 1}$ are training $A_i \in^{K^2\times E}$ is a feature map subjected to dimension reduction, where E=200 and K=8.

[0081] At step 4, a spatial-temporal recurrent attention frame is constructed, and a spatial attention mechanism is performed based on a recurrent neural network and the Gaussian Kernel Function to obtain a spatial attention result $g(_S{}^{(t)}$.

[0082] Preferably, the step 4 specifically includes the following steps:

[0083] At step 4.1, a long short-term memory network (LSTM) of two layers is constructed.

[0084] At step 4.1, the recurrent neural network is constructed based on the long short-term memory network of two layers and a hidden layer is expressed as in the formula (1) below.

$$h_t^{(1)}=R_{recur}(\hat{g}^{(t)},h_{t-1}^{(1)}|W_{r1})$$

$$h_t^{(2)}=R_{recur}(h_t^{(1)},h_{t-1}^{(2)}|W_{r2}) \tag{1}$$

[0085] In the above formula, $R_{recur}(\bullet)$ corresponding to $h_t^{(1)}$ is expanded as the following formula (2):

$$\begin{pmatrix} i_t \\ f_t \\ o_t \\ g_t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} M\begin{pmatrix} h_{t-1}, \\ \hat{g}^{(t)} \end{pmatrix}, \tag{2}$$

$$c_t = f_t e c_{t-1} + i_t e g_t,$$

$$h_t = o_t e \tanh(c_t)$$

[0086] In the above formula, i_t, f_t and o_t represent an input gate, a forget gate and an output gate respectively; c_t and h_t represent a cell state and a hidden layer state respectively; $\hat{g}^{(t)}$ represents a feature vector input into the long short-term memory network by the t-th step of attention mechanism; M: $^a\rightarrow{}^b$ is an affine transformation composed of trainable parameters, a=d+E,b=4d; d is the same vector dimension corresponding to i_t, f_t, o_t, g_t, c_t and h_t, E is a dimension of an input feature subjected to dimension reduction, and $R_{recur}(\bullet)$ corresponding to $h_t^{(2)}$ is obtained by replacing $\hat{g}^{(t)}$ in the above $R_{recur}(\bullet)$ with $h_t^{(t)}$.

[0087] At step 4.2, the state of the long short-term memory network is initialized by constructing an encoding process of the feature map.

[0088] Preferably, step 4.2 specifically includes the following steps:

[0089] At step 4.2.1, a new double-layer long short-term memory network (LSTM) having the same structure as formula (1) is constructed.

[0090] At step 4.2.2, the feature map A_{mid} corresponding to the exact center of a CT slice sequence of each lymph node at step 3 is divided according to a spatial neighborhood; specifically, 8×8×200 are divided into 16 sub-feature blocks with 2×2×200 based on adjacent four positions as one group.

[0091] At step 4.2.3, the 16 sub-feature blocks are input into the new double-layer long short-term memory network sequentially clockwise from outside to inside and subjected to 16 cycles to obtain a cell state $c'_T{}^{(2)}$ corresponding to the second layer of the LSTM at the last moment so as to initialize the cell state $c_0^{(2)}$ of the second layer of the long short-term memory network at step 4.1.

[0092] At step 4.3, a spatial attention position within a range of the feature map is predicted based on a sending-network.

[0093] The step 4.3 specifically includes the following steps:

[0094] At step 4.3.1, the feature vector $h_t^{(2)}$ output by the first hidden layer of the long short-term memory network and the feature result $g_{S,Center}{}^{(t)}$ corresponding to the center of the slice sequence in the recurrent attention iteration step are concatenated to obtain $[h_t^{(2)}, g_{S,Center}{}^{(t)}]$. Herein, let $g_{S,Center}{}^{(0)}=A_{Center}$.

[0095] At step 4.3.2, $[h_t^{(2)}, g_{S,Center}{}^{(t)}]$ is input to the sending network composed of one fully-connected layer to perform regression for the spatial attention position of the next recurrent iteration step as shown in the formula (3):

$$(\mu_S{}^{(t+1)}, \sigma_S{}^{(t+1)})=\sigma(W_S[h_t^{(2)}, g_{S,Center}{}^{(t)}]+b_S) \tag{3}$$

[0096] In the above formula, $\mu_S{}^{(t+1)}=(\mu_{S,x}{}^{(t+1)}, \mu_{S,y}{}^{(t+1)})$ represents an attention position coordinate predicted at the t+1 moment, where the variance $\sigma_S{}^{(t+1)}$ of the Gaussian Kernel Function is set to a fixed value 0.1, thus, $W_S\in^{2\times(R+E)}$, and $b_S\in^{2\times 1}$ only corresponds to two output nodes.

[0097] At step 4.4, an attention matrix $l^{(t)}$ is constructed based on a two-dimension Gaussian Kernel Function.

[0098] At step 4.4, after the attention position coordinate $\mu_S{}^{(t)}$ is obtained, an attention template matrix is constructed based on the two-dimension Gaussian Kernel Function and softmax as shown in the following formula (4):

$$\begin{aligned} l_i^{(t)} &= p(L_t = i \mid \mu_S^{(t)}, \sigma_S^{(t)}) \\ &= softmax(\phi(l_i \mid \mu_S^{(t)}, \sigma_S^{(t)})) \\ &= \frac{\exp(\phi(l_i \mid \mu_S^{(t)}, \sigma_S^{(t)}))}{\sum_{j=1}^{K^2} \exp(\phi(l_j \mid \mu_S^{(t)}, \sigma_S^{(t)}))} \end{aligned} \tag{4}$$

[0099] In the above formula, $\phi(l_i|\mu_X^{(t)}, \sigma_S^{(t)})=C\cdot\exp(l_i-\mu_S^{(t)})/2\sigma_S^{(t)}$, $l_i=\{(x_i, y_i)\}_{j=1}^{K\times K}$ is discrete equidistant position coordinates normalized to the interval [0, 1] within the range A_i; C is Gaussian normalized constant which is a fixed value 10.

[0100] At step 4.5, $1^{(t)}$ is multiplied by A_i element by element and added up so as to obtain the spatial attention result $g_S^{(t)}$.

[0101] At step 5, a temporal attention mechanism is performed for the spatial attention result $g_S^{(t)}$ obtained at step 4 to obtain a spatial-temporal attention feature $\hat{g}^{(t)}$.

[0102] The step 5 specifically includes the following steps:

[0103] At step 5.1, a mixture density network is constructed to predict an attention position $\mu_T^{(t)}$ of a slice direction.

[0104] The step 5.1 specifically includes the following steps:

[0105] At step 5.1.1, the sequence feature $g_S^{(t)}$ corresponding to each lymph node is equally divided into left and right halves, that is, $g_{Left}^{(t)}$ and $g_{Right}^{(t)}$.

[0106] At step 5.1.2, the mixture density network is formed by one fully-connected hidden layer to perform regression for the temporal attention position coordinate and the Mixture Gaussian Function parameters as shown in the following formula (5):

$$Z^{(t)}=\sigma(W_T g_{S/2}^{(t)}+b_T) \tag{5}$$

[0107] In the above formula, $g_{S/2}^{(t)}\in^{E\cdot L/2}$ represents the left half or the right half of $g_S^{(t)}$; $W_T\in^{2C\times E\cdot L/2}$, $b_T\in^{2C\times 1}$ represents a training parameter of the mixture density network, C represents a component number of the Gaussian Mixture Model corresponding to the left half and the right half of the sequence feature, and σ is sigmoid function.

[0108] At step 5.1.3, respective regression coefficients $Z_{Left}^{(t)}\in^{2C\times 1}$ and $Z_{Right}^{(t)}\in^{2C\times 1}$ of the left half and the right half are both distributed to the Gaussian Mixture Model according to a rule of $Z_{0:\,C\text{-}1}^{(t)}\rightarrow\pi_T^{(t)}$ and $Z_{C:\,2C\text{-}1}^{(t)}\rightarrow\mu_T^{(t)}$.

[0109] At step 5.1.4, with the formula (6), μ_{Left} is limited to [0, 0.5) and μ_{Right} is limited to (0.5, 1].

$$\mu_{Left}\leftarrow\mu_{Left}\times 0.5,\ \mu_{Right}\leftarrow 0.5\times(1+\mu_{Right}) \tag{6}$$

[0110] At step 5.2, an attention weight vector $l'^{(t)}$ is obtained based on Gaussian Mixture Distribution.

[0111] The step 5.2 specifically includes the following steps:

[0112] At step 5.2.1, the coefficient $\pi_T^{(t)}$ of each Gaussian component is normalized with softmax, that is, $\pi_T^{(t)}\leftarrow\text{softmax}(\pi_T^{(t)})$, and the variance $\Sigma_T^{(t)}$ of all Gaussian components is fixed as 0.1.

[0113] At step 5.2.2, in combination with the Gaussian Mixture Model parameters $(\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})$ derived from the slice feature vectors of the left and right halves, a corresponding temporal attention weight vector $l'^{(t)}_i$ is obtained based on the formula (7).

$$\begin{aligned} l_t'^{(t)} &= p(L_t = i\,|\,\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}) \\ &= softmax(N(l_i\,|\,\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})) \\ &= \frac{\exp(N(l_i\,|\,\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}))}{\sum_{j=1}^{L}\exp(N(l_j\,|\,\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}))} \end{aligned} \tag{7}$$

[0114] In the above formula, N is a Gaussian mixture density function.

$$N(l_i|\pi_T^{(t)},\mu_T^{(t)},\Sigma_T^{(t)})=\Sigma_{c=1}^{C}\pi_{T,c}^{(t)}\phi_c(l_i,\mu_{T,c}^{(t)},\Sigma_{T,c}^{(t)}),\ s.t.\Sigma_{c=1}^{C}\pi_{T,c}^{(t)}=1\phi_c(l_i|\mu_{T,c}^{(t)},\Sigma_{T,c}^{(t)})=C\cdot\exp((l_i-\mu_{T,c}^{(t)})/2\Sigma_{T,c}^{(t)} \quad (8$$

[0115] At step 5.3, $l'^{(t)}$ and the input feature $g_S^{(t)}$ are multiplied element by element and added to obtain the spatial-temporal attention feature $\hat{g}^{(t)}$.

[0116] At step 6, a lymph node positive score $\hat{y}_t$ of the recurrent attention iteration step is predicted by using the recurrent neural network constructed at step 4 in combination with the spatial-temporal attention feature $\hat{g}^{(t)}$ obtained at step 5.

[0117] At step 7, a loss function of the model is constructed to perform steps 4-6 for T times, and supervised training is performed for the model by using a gradient back propagation algorithm.

[0118] At step 7, the target function for the model to receive supervised training is obtained in the following steps:

[0119] At step 7.1, a classification loss function is constructed; and cross-entropy loss functions corresponding to all recurrent attention iteration steps are investigated as shown in the following formula (9).

$$F_c = -\sum_{t=1}^{T}\sum_{i=1}^{2} y_i \log\hat{y}_{t,i} \tag{9}$$

[0120] In the above formula, y_i and $\hat{y}_{t,i}$ are a real lymph node positive score and a predicted lymph node positive score, which are a total number of the recurrent attention iterations.

[0121] At step 7.2, a constraint term of a predicted position of the spatial attention mechanism is constructed in the following two steps:

[0122] At step 7.2.1, a "convergent" constraint term of the position is constructed, and the predicted spatial attention position is constrained to around the center of the feature map based on Batch Normalization and the prior information of the lymph node being at the center of the slice, as shown in the following formula (10):

$$\begin{aligned} &\min\left(\sum\nolimits_{i=1}^{d}(\gamma_i-1)^2+\beta_i^2\right) \\ &\text{where } \hat{y}_i \leftarrow \gamma\hat{x}_i+\beta,\ (i=1,\ \ldots\ ,m), \\ &\hat{x}_i \leftarrow \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}} \Leftrightarrow \hat{X}=A^T(X-B) \end{aligned} \tag{10}$$

[0123] In the above formula, β and γ are trainable parameters introduced by the Batch Normalization; d is an output dimension of the sending network; and m is a batch sample capacity.

[0124] At step 7.2.2, a "divergent" constraint term of the position is constructed; the uniformly-distributed cross-entropy loss functions are constructed according to different positions predicted in the entire recurrent attention process, as shown in the following formula (11):

$$F_S=\text{Const}-\Sigma_{t=1}^{T}P_{Unif}\log[\text{softmax}(\mu_S^{(t)})] \tag{11}$$

[0125] In the above formula, P_{Unif}: Uniform(0, K^2) represents a uniform distribution within a spatial range of the feature map; the constant Const is set to 1.

[0126] At step 7.3, a constraint term of a predicted position of the temporal attention mechanism is constructed; the "convergent" constraint term is applied to the predicted position, the batch normalization layer is also introduced to the mixture density network and the formula (10) is minimized so that the attention positions of the left and right halves in the direction of slice $\mu_T^{(t)}\approx(0.25, 0.75)$.

[0127] At step 7.4, a final entire loss function of the model is constructed as shown in the following formula (12):

$$F = -\sum_{t=1}^{T}\sum_{i=1}^{2} y_i \log \hat{y}_{t,i} + \lambda_1\left(\sum\nolimits_{i=1}^{d} (\gamma_i - 1)^2 + \beta_i^2\right) + \lambda_2\left(Const - \sum\nolimits_{t=1}^{T} p_{Unif} \log[softmax(\mu_S^{(t)})]\right) \tag{12}$$

[0128] At step 8, steps 3-7 are repeated to perform iterative training for the model until a trained model is obtained at the end of training.

[0129] The training data is formed by the true and false positive samples in the same proportion. The model adopts three cross-validations during a test. The parameters of the model are set according to the following values: 1. in the formula (12), $\lambda_1=1$ and $\lambda_2=0.5$; 2. The number of recurrences of the recurrent attention mechanism for each time T=8;(3) in the formula (5), the component number of Gaussian Mixture Model respectively corresponding to the left and right halves of sequence: C=1, that is, there are totally three components, each of which corresponds to three parts, i.e. left, middle and right subparts; (4)the capacity of training batch samples is 20, the initial learning rate is set to 0.1 with reduction of 0.1 for each 25 periods in a total 200 training periods.

[0130] At step 9, the lymph node CT sequence to be detected is input to perform model reasoning process and the positive score $\hat{y}_T$ output by the final recurrent attention is taken as the CT lymph node detection result.

[0131] FIG. 6 is an instance diagram of performing a spatial attention mechanism. It can be seen that the attention position is accurately located near the boundary of a lymph node when the constraint term of the spatial attention position and its reasonable hyper-parameter are set.

[0132] FIG. 7 is an instance diagram of performing a temporal attention mechanism. In the Figure, the first two instances correspond to the true positive nodes and the last two instances correspond to the false positive nodes. The bar charts at the bottom are self-adaptive weight values predicted by the model for each slice in the sequence. Under the action of the constraint term of the temporal attention position, the model can accurately and self-adaptively located to a slice boundary corresponding to a different node size, thereby extracting context feature information relating to a lymph node.

[0133] FIG. 8 is a diagram of comparison of FROCs and AUC of the present disclosure and a reference model. Herein, the reference model is a double-layer LSTM model without adopting the recurrent attention mechanism. For fair comparison, its structure is consistent with that of LSTM in step 4.1. Sequence=1 represents that the reference model can only extract the slices at the center of the lymph node and the model with a recurrence number T=1 is degraded into a classical CNN structure.

[0134] Sequence=1 represents that the reference model performs nine recurrences. It can be seen from the figure that FROCs and AUC in the present disclosure both perform optimally.

[0135] FIG. 9 is a diagram of performance comparison of the present disclosure and other models. In the table, those symbols with "*-H" represents that the input image size corresponds to a high resolution (256×256); "*-L" represents a low resolution (64×64); "*-RI" represents that the model is trained from the beginning; "*-TL" represents that the model parameters are migrated from the ImageNet data set. It can be seen from the table that the present disclosure has the highest accuracies in classification of both chest lymph nodes and celiac lymph nodes.

[0136] Of course, the above descriptions are not intended to limit the present disclosure and the present disclosure is also not limited to these examples. Any modification, changes, addition or replacements made by those skilled in the art within the essential scope of the present disclosure shall all fall in the scope of protection of the present disclosure.

1. A CT lymph node detection system based on spatial-temporal recurrent attention mechanism, comprising a training sample extracting module, a deep layer feature extracting network, a feature embedding network, and a spatial-temporal recurrent attention target detection module, wherein a detection process includes the following steps:

at step 1, position coordinate information is marked for the obtained lymph node dcm-format file and a corresponding lymph node by use of the training sample extracting module, and a CT slice image block sequence I_i (i=1,2, . . . , L), $I_i \in^{W\times H}$, with CT slice image blocks being length L, width W and height H is extracted for each lymph node by using a pydicom module in python;

at step 2, extracting a high-level spatial feature map sequence corresponding to the CT slice image block sequence of each lymph node is extracted by using a VGG-16 model pre-trained by a natural image according to the deep layer feature extracting network and denoting the high-level spatial feature map sequence as $\{X_0, \ldots, X_L\}$.

at step 3, constructing the feature embedding network to perform dimension reduction for the input high-level feature map sequence and output a feature map A_i;

at step 4, constructing a spatial-temporal recurrent attention frame, and performing a spatial attention mechanism is performed based on a recurrent neural network and the Gaussian Kernel Function to obtain a spatial attention result $g_S^{(t)}$.

at step 5, performing a temporal attention mechanism for the spatial attention result $g_S^{(t)}$ obtained at step 4 to obtain a spatial-temporal attention feature $\hat{g}^{(t)}$.

at step 6, predicting a lymph node positive score $\hat{y}_t$ of the recurrent attention iteration step by using the recurrent neural network constructed at step 4 in combination with the spatial-temporal attention feature $\hat{g}^{(t)}$ obtained at step 5;

at step 7, constructing a loss function of the model to perform steps 4-6 for T times, and performing supervised training for the model by using a gradient back propagation algorithm;

at step 8, performing iterative training for the model by repeating steps 3-7, until a trained model is obtained at the end of training; and

at step 9, inputting the lymph node CT sequence to be detected to perform a model reasoning process, and taking a positive score $\hat{y}_T$ output by the final recurrent attention as a CT lymph node detection result.

2. The CT lymph node detection system according to claim **1**, wherein the step 4 specifically comprises the following steps:

at step 4.1, constructing a long short-term memory network (LSTM) of two layers;

at step 4.2, initializing the state of the long short-term memory network by constructing an encoding process of the feature map;

at step 4.3, predicting a spatial attention position within a range of the feature map by using a sending network;

at step 4.4, constructing an attention matrix $1^{(t)}$ based on a two-dimension Gaussian Kernel Function; and

at step 4.5, $1^{(t)}$ is multiplied by A_i element by element and then added up so as to obtain the spatial attention result $g_S^{(t)}$.

3. The CT lymph node detection system according to claim **2**, wherein at step 4.1, the recurrent neural network is constructed based on the long short-term memory network of two layers and a hidden layer is expressed as shown in the following formula (1):

$$h_t^{(1)}=R_{recur}(\hat{g}^{(t)},h_{t\text{-}1}^{(1)}|W_{r1})$$

$$h_t^{(2)}=R_{recur}(h_t^{(1)},h_{t\text{-}1}^{(2)}|W_{r2}) \tag{1}$$

wherein $R_{recur}(\bullet)$ corresponding $h_t^{(1)}$ is expanded as shown in the following formula (2):

$$\begin{pmatrix} i_t \\ f_t \\ o_t \\ g_t \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ \tanh \end{pmatrix} M\begin{pmatrix} h_{t-1}, \\ \hat{g}^{(t)} \end{pmatrix}, \tag{2}$$

$$c_t = f_t e c_{t-1} + i_t e g_t,$$

$$h_t = o_t e \tanh(c_t)$$

wherein, i_t, f_t and o_f represent an input gate, a forget gate and an output gate respectively; c_t and li_t represent a cell state and a hidden layer state respectively; $\hat{g}^{(t)}$ represents a feature vector input into the long short-term memory network by the t-th step of attention mechanism; $M: {}^{a} \rightarrow {}^{b}$ is an affine transformation composed of trainable parameters, a=d+E, b=4d; d is the same vector dimension corresponding to i_t, f_t, o_t, g_t, c_t and h_t, E is a dimension of an input feature subjected to dimension reduction, and $R_{recur}(\bullet)$ corresponding to $h_t^{(2)}$ is obtained by replacing $\hat{g}^{(t)}$ in the above $R_{recur}(\bullet)$ with $h_t^{(1)}$.

4. The CT lymph node detection system according to claim **2**, wherein the step 4.2 specifically comprises the following steps:

at step 4.2.1, constructing a new double-layer long short-term memory network having the same structure as formula (1);

at step 4.2.2, dividing the feature map A_{mid} corresponding to the exact center of a CT slice sequence of each lymph node at step 3 according to a spatial neighborhood; specifically, dividing 8×8×200 into 16 sub-feature blocks with 2×2×200 based on adjacent four positions as one group; and

at step 4.2.3, inputting the 16 sub-feature blocks into the new double-layer long short-term memory network sequentially clockwise from outside to inside to go through 16 cycles and obtain a cell state $c'_T{}^{(2)}$ corresponding to the second layer of the LSTM at the last moment so as to initialize the cell state $c_0^{(2)}$ of the second layer of the long short-term memory network at step 4.1.

5. The CT lymph node detection system according to claim **2**, wherein the step 4.3 specifically comprises the following steps:

at step 4.3.1, concatenating a feature vector $h_t^{(2)}$ output by the first hidden layer of the long short-term memory network and a feature result $g_{S,Center}{}^{(t)}$ corresponding to the center of the slice sequence in the recurrent attention iteration step to obtain $[h_t^{(2)}, g_{S,Center}{}^{(t)}]$;

At step 4.3.2, inputting $[h_t^{(2)}, g_{S,Center}{}^{(t)}]$ to a sending network composed of one fully-connected layer to perform regression for the spatial attention position of the next recurrent iteration step as shown in the formula (3):

$$(\mu_S{}^{(t+1)}, \sigma_S{}^{(t+1)})=\sigma(W_S[h_t^{(2)}, g_{S,Center}{}^{(t)}]+b_S) \tag{3}$$

wherein $\mu_S{}^{(t+1)}=(\mu_{S,x}{}^{(t+1)}, \mu_{S,y}{}^{(t+1)})$ represents an attention position coordinate predicted at the t+1 moment, where the variance $\sigma_S{}^{(t+1)}$ of the Gaussian Kernel is set to a fixed value 0.1, thus, $W_S \in{}^{2\times(R+E)}$, and $b_S \in{}^{2\times1}$ only corresponds to two output nodes.

6. The CT lymph node detection system according to claim **2**, wherein at the step 4.4, after an attention position coordinate $\mu_S{}^{(t)}$ is obtained, an attention template matrix is constructed based on the two-dimension Gaussian Kernel Function and softmax as shown in the following formula (4):

$$\begin{aligned} l_i^{(t)} &= p(L_t = i \mid \mu_S^{(t)}, \sigma_S^{(t)}) \\ &= softmax(\phi(l_i \mid \mu_S^{(t)}, \sigma_S^{(t)})) \\ &= \frac{\exp(\phi(l_i \mid \mu_S^{(t)}, \sigma_S^{(t)}))}{\sum_{j=1}^{K^2} \exp(\phi(l_j \mid \mu_S^{(t)}, \sigma_S^{(t)}))} \end{aligned} \tag{4}$$

wherein $\phi(1_i|\mu_S{}^{(t)}, \sigma_S{}^{(t)})=C\cdot\exp(1_i-\mu_S{}^{(t)})/2\sigma_S{}^{(t)}$, $1_i=\{(x_i, y_l)\}_{j=1}{}^{K\times K}$ is discrete equidistant position coordinates normalized to the interval [0, 1] within the range A_i; C is Gaussian normalized constant which is a fixed value 10.

7. The CT lymph node detection system according to claim **1**, wherein the step 5 specifically comprises the following steps:

at step 5.1, constructing a mixture density network to predict an attention position $\mu_T{}^{(t)}$ of a slice direction.

at step 5.2, obtaining an attention weight vector $1'^{(t)}$ obtained based on Gaussian Mixture Distribution; and at step 5.3, multiplying $l'^{(t)}$ by the input feature $g_S^{(t)}$ element by element and performing addition to obtain the spacial-temporal attention feature $\hat{g}^{(t)}$.

8. The CT lymph node detection system according to claim **7**, wherein the step 5.1 specifically comprises the following steps:

at step 5.1.1, equally dividing the sequence feature $g_S^{(t)}$ corresponding to each lymph node into left and right halves, that is $g_{Left}^{(t)}$ and $g_{Right}^{(t)}$;

at step 5.1.2, forming the mixture density network is formed by one fully-connected hidden layer to perform regression for the temporal attention position coordinate and the Mixture Gaussian Function parameters as shown in the following formula (5):

$$Z^{(t)}=\sigma(W_T g_{S/2}{}^{(t)}+b_T) \tag{5}$$

wherein $g_{S/2}{}^{(t)}\in^{E\cdot L/2}$ represents the left half or the right half of $g_S^{(t)}$; $W_T\in^{2C\times E\cdot L/2}$, $b_T\in^{2C\times 1}$ represents a training parameter of the mixture density network, C represents a component number of the Gaussian Mixture Model respectively corresponding to the left half and the right half of the sequence feature, and σ is sigmoid function.

at step 5.1.3, distributing respective regression coefficients $Z_{Left}{}^{(t)}\in^{2C\times 1}$ and $Z_{Right}{}^{(t)}\in^{2C\times 1}$ of the left half and the right half are both distributed to the Gaussian Mixture Model according to a rule of $Z_{0:\ C\text{-}1}{}^{(t)}\rightarrow\pi_T{}^{(t)}$ and $Z_{C:\ 2C\text{-}1}{}^{(t)}\rightarrow\mu_T{}^{(t)}$; and

at step 5.1.4, with the formula (6), limiting μ_{Left} is limited to [0, 0.5) and limiting μ_{Right} to (0.5, 1],

$$\mu_{Left}\leftarrow\mu_{Left}\times 0.5,\ \mu_{Right}\leftarrow 0.5\times(1+\mu_{Right}) \tag{6}$$

9. The CT lymph node detection system according to claim **7**, wherein the step 5.2 specifically comprises the following steps:

at step 5.2.1, normalizing the coefficient $\pi_T^{(t)}$ of each Gaussian component is normalized with softmax, that is, $\pi_T^{(t)}\leftarrow\text{softmax}(\pi_T^{(t)})$, and fixing a variance $\Sigma_T^{(t)}$ of all Gaussian components as 0.1;

at step 5.2.2, in combination with the Gaussian Mixture Model parameters $(\pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})$ derived from the slice feature vectors of the left and right halves, a corresponding temporal attention weight vector $l'_i{}^{(t)}$ is obtained based on the formula (7).

$$\begin{aligned} l_t'^{(t)} &= p(L_t = i \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}) \\ &= softmax(N(l_i \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)})) \\ &= \frac{\exp(N(l_i \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}))}{\sum_{j=1}^{L}\exp(N(l_j \mid \pi_T^{(t)}, \mu_T^{(t)}, \Sigma_T^{(t)}))} \end{aligned} \tag{7}$$

wherein N is a Gaussian mixture density function.

$$N(l_i|\pi_T{}^{(t)}, \mu_T{}^{(t)}, \Sigma_T{}^{(t)})=\Sigma_{c=1}{}^{C}\pi_{T,c}{}^{(t)}\phi_c(l_i,\mu_{T,c}{}^{(t)},\Sigma_{T,c}{}^{(t)}),\ s.t.\Sigma_{c=1}{}^{C}\pi_{T,c}{}^{(t)}=1\phi_c(l_i|\mu_{T,c}{}^{(t)},\Sigma_{T,c}{}^{(t)})=C\cdot\exp(l_i-\mu_{T,c}{}^{(t)})/2\Sigma_{T,c}{}^{(t)} \tag{8}$$

10. The CT lymph node detection system according to claim **1**, wherein at step 7, a target function for the model to receive supervised training is obtained in the following steps:

at step 7.1, constructing a classification loss function; and investigating cross-entropy loss functions corresponding to all recurrent attention iteration steps as shown in the following formula (9):

$$F_c = -\sum_{t=1}^{T}\sum_{i=1}^{2} y_i \log\hat{y}_{t,i} \tag{9}$$

wherein y_i and $\hat{y}_{t,i}$ are a real lymph node positive score and a predicted lymph node positive score, which are a total number of the recurrent attention iterations;

at step 7.2, a constraint term of a predicted position of the spatial attention mechanism is constructed in the following two steps:

at step 7.2.1, constructing a "convergent" constraint term of the position, and constraining the predicted spatial attention position is constrained to around the center of the feature map based on Batch Normalization and the prior information of the lymph node being at the center of the slice, as shown in the following formula (10):

$$\begin{aligned} &\min\left(\sum_{i=1}^{d}(\gamma_i-1)^2+\beta_i^2\right) \\ &\text{where } \hat{y}_i \leftarrow \gamma\hat{x}_i+\beta,\ (i=1,\ \ldots\ ,m), \\ &\hat{x}_i \leftarrow \frac{x_i-\mu_B}{\sqrt{\sigma_B^2+\varepsilon}} \Leftrightarrow \hat{X}=A^T(X-B) \end{aligned} \tag{10}$$

wherein β and γ are trainable parameters introduced by the Batch Normalization; d is an output dimension of the sending network; and m is a batch sample capacity;

at step 7.2.2, constructing a "divergent" constraint term of the position is constructing the uniformly-distributed cross-entropy loss functions according to different positions predicted in the entire recurrent attention process, as shown in the following formula (11):

$$F_S=\text{Const}-\Sigma_{t=1}{}^{T}P_{Unif}\log[\text{softmax}(\mu_S{}^{(t)})] \tag{11}$$

wherein P_{Unif}: Uniform ($\mathbf{0}$, K^2) represents a uniform distribution within a spatial range of the feature map; the constant Const is set to 1.

at step 7.3, constructing a constraint term of a predicted position of the temporal attention mechanism; applying the "convergent" constraint term is applied to the predicted position, further introducing the batch normalization layer to the mixture density network and minimizing the formula (10) so that the attention positions of the left and right halves in the direction of slice $\mu_T{}^{(t)}\sim(0.25, 0.75)$; and

at step 7.4, constructing a final entire loss function of the model is constructed as shown in the following formula (12):

$$\begin{aligned} F = -\sum_{t=1}^{T}\sum_{i=1}^{2} y_i\log\hat{y}_{t,i} + \lambda_1\left(\sum_{i=1}^{d}(\gamma_i-1)^2+\beta_i^2\right) + \\ \lambda_2\left(Const-\sum_{t=1}^{T}p_{Unif}\log[softmax(\mu_S^{(t)})]\right) \end{aligned} \tag{12}$$

* * * * *