CLAIMS

What is claimed is:

1. A method for generating a medical report, comprising:

receiving a medical image to be recognized; if the medical image is acquired by scanning, collecting ambient light intensity at the time of scanning and performing differential processing on the medical image based on the ambient light intensity;

importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image;

importing the visual feature vector and the keyword sequence into a preset diagnostic item recognition model to determine diagnostic items corresponding to the medical image;

respectively constructing a paragraph for describing each of the diagnostic items based on a diagnostic item extension model;

generating a medical report for the medical image based on the paragraph, the keyword sequence and the diagnostic items;

wherein the respectively constructing a paragraph for describing each of the diagnostic items based on a diagnostic item extension model comprises:

aggregating all the diagnostic items to form a diagnostic item vector;

importing the diagnostic item vector into a LSTM neural network, outputting paragraphs of the diagnostic items, wherein the number of layers of the LSTM neural network is the same as the number of the diagnostic items;

wherein, after receiving a medical image to be recognized, the method further comprises:

performing binaryzation on the medical image to acquire a binarized medical image;

identifying a boundary of the binarized medical image, and dividing the medical image into a plurality of medical sub-images; if boundaries of a plurality of objects in the medical image are overlapping or adjacent, dividing the plurality of objects into the same one medical sub-image;

wherein the step of importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image comprises:

respectively importing the medical sub-images into the VGG neural network to acquire visual feature components and keyword sub-sequences of the medical sub-images;

generating the visual feature vector based on the visual feature components, and constructing the keyword sequence based on the keyword sub-sequences.

2. The method according to claim 1, wherein the step of importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image comprises:

constructing a pixel matrix of the medical image based on pixel values of pixels in the medical image and position coordinates of the pixel values;

performing dimensionality reduction on the pixel matrix through five pooling layers of the VGG neural network to acquire the visual feature vector;

importing the visual feature vector into a fully connected layer of the VGG neural network and outputting an index sequence corresponding to the visual feature vector;

determining the keyword sequence corresponding to the index sequence according to a keyword index table.

3. The method according to claim 1, wherein the step of importing the visual feature vector and the keyword sequence into a preset diagnostic item recognition model to determine diagnostic items corresponding to the medical image comprises:

generating a keyword feature vector corresponding to the keyword sequence based on sequence numbers of keywords in a preset text corpus;

respectively importing the keyword feature vector and the visual feature vector into a preprocessing function to acquire a preprocessed keyword feature vector and a preprocessed visual feature vector; wherein the preprocessing function is specifically as:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{M} e^{z_i}}$$

where $\sigma(z_j)$ is a value of j-th element in the preprocessed keyword feature vector or in the preprocessed visual feature vector, z_j is a value of j-th element in the keyword feature vector or in the visual feature vector, *M* is the number of elements corresponding to the keyword feature

vector or the visual feature vector;

determining the preprocessed keyword feature vector and the preprocessed visual feature vector as an input of the diagnostic item recognition model, and outputting the diagnostic items.

4. The method according to any one of claims 1-3, wherein the method further comprises:

acquiring training visual vectors, training keyword sequences and training diagnostic items of a plurality of training images;

determining the training visual vectors and the training keyword sequences as an input of a LSTM neural network, determining the training diagnostic items as an output of the LSTM neural network, and adjusting learning parameters in the LSTM neural network so that the LSTM neural network meets a convergence condition; wherein the convergence condition is as:

$$\theta^* = arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$$

where θ^* is the adjusted learning parameter, *Visual* is the training visual vector, *Keyword* is the training keyword sequence, *Stc* is the training diagnostic item, $p(Visual, Keyword|Stc; \theta)$ represents an output result of a probability value of the training diagnostic item when the training visual vector and the training keyword sequence are imported into the LSTM neural network with the value of the learning parameter is θ, and $arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$ is the value of the learning parameter when the probability value takes a maximum value;

determining the adjusted LSTM neural network as the diagnostic item recognition model.

5. A device for generating a medical report, comprising:

a medical image receiving unit, configured to: receive a medical image to be identified; if the medical image is acquired by scanning, collect ambient light intensity at the time of scanning and perform differential processing on the medical image based on the ambient light intensity;

a feature vector acquisition unit, configured to import the medical image into a preset VGG neural network, to acquire a visual feature vector and a keyword sequence of the medical image;

a diagnostic item recognition unit, configured to import the visual feature vector and the keyword sequence into a preset diagnostic item recognition model to determine diagnostic items corresponding to the medical image;

a paragraph determination unit, configured to construct a paragraph for describing the diagnostic items based on a diagnostic item extension model;

a medical report generation unit, configured to generate a medical report for the medical image according to the paragraph, the keyword sequence, and the diagnostic items;

wherein the construct a paragraph for describing the diagnostic items based on a diagnostic item extension model comprises:

aggregating all the diagnostic items to form a diagnostic item vector;

importing the diagnostic item vector into a LSTM neural network, outputting paragraphs of the diagnostic items, wherein the number of layers of the LSTM neural network is the same as the number of the diagnostic items;

wherein the device further comprises:

a binarization unit, configured to perform binarization on the medical image to acquire a binarized medical image;

a boundary division unit, configured to: identify a boundary of the binarized medical image, and to divide the medical image into a plurality of medical sub-images; if boundaries of a plurality of objects in the medical image are overlapping or adjacent, dividing the plurality of objects into the same one medical sub-image;

wherein the feature vector acquisition unit comprises:

a medical sub-image recognition unit, configured to respectively import the medical sub-images into the VGG neural network to acquire visual feature components and keyword sub-sequences of the medical sub-images; and

a feature vector combination unit, configured to generate the visual feature vector based on the visual feature components, and to construct the keyword sequence based on the keyword sub-sequences.

6. The device according to claim 5, wherein the feature vector acquisition unit comprises:

a pixel matrix construction unit, configured to construct a pixel matrix of the medical image based on pixel values of pixels in the medical image and position coordinates of the pixel values;

a visual feature vector generation unit, configured to perform dimensionality reduction on the pixel matrix through five pooling layers of the VGG neural network to acquire the visual feature vector;

an index sequence generation unit, configured to import the visual feature vector into a fully connected layer of the VGG neural network, and to output an index sequence corresponding to the visual feature vector;

a keyword sequence generation unit, configured to determine the keyword sequence corresponding to the index sequence according to a keyword index table.

7. The device according to claim 5, wherein the diagnostic item recognition unit comprises:

a keyword feature vector construction unit, configured to generate the keyword feature vector corresponding to the keyword sequence based on sequence numbers of keywords in a preset text corpus;

a preprocessing unit, configured to respectively import the keyword feature vector and the visual feature vector into a preprocessing function to acquire a preprocessed keyword feature vector and a preprocessed visual feature vector; wherein the preprocessing function is specifically as:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{M} e^{z_i}}$$

where $\sigma(z_j)$ is a value of j-th element in the preprocessed keyword feature vector or in the preprocessed visual feature vector, z_j is a value of j-th element in the keyword feature vector or in the visual feature vector, *M* is the number of elements corresponding to the keyword feature vector or the visual feature vector;

a preprocessed vector importing unit, configured to determine the preprocessed keyword feature vector and the preprocessed visual feature vector as an input of the diagnostic item recognition model, and to output the diagnosis items.

8. The device according to any one of claims 5-7, wherein the device further comprises:

a training parameter acquisition unit, configured to acquire training visual vectors, training

keyword sequences, and training diagnostic items of a plurality of training images;

a learning parameter training unit, configured to determine the training visual vectors and the training keyword sequences as an input of a LSTM neural network, and to determine the training diagnostic items as an output of the LSTM neural network, and to adjust learning parameters in the LSTM neural network so that the LSTM neural network meets a convergence condition; wherein the convergence condition is as:

$$\theta^* = arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$$

where θ^* is the adjusted learning parameter, *Visual* is the training visual vector, *Keyword* is the training keyword sequence, *Stc* is the training diagnostic item, $p(Visual, Keyword|Stc; \theta)$ represents an output result of a probability value of the training diagnostic item when the training visual vector and the training keyword sequence are imported into the LSTM neural network with the value of the learning parameter is θ, and $arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$ is the value of the learning parameter when the probability value takes a maximum value;

a unit for generating a diagnostic item recognition model, configured to determine the adjusted LSTM neural network as the diagnostic item recognition model.

9. A device for generating a medical report, comprising a memory, a processor, and a computer-readable instruction stored in the memory and executable on the processor, wherein the processor, when executing the computer-readable instruction, implements the following steps of:

receiving a medical image to be recognized; if the medical image is acquired by scanning, collecting ambient light intensity at the time of scanning and performing differential processing on the medical image based on the ambient light intensity;

importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image;

importing the visual feature vector and the keyword sequence into a preset diagnostic item recognition model to determine diagnostic items corresponding to the medical image;

constructing a paragraph for describing each of the diagnostic items respectively based on a

diagnostic item extension model;

generating a medical report for the medical image based on the paragraph, the keyword sequence and the diagnostic items;

wherein the constructing a paragraph for describing each of the diagnostic items respectively based on a diagnostic item extension model comprises:

aggregating all the diagnostic items to form a diagnostic item vector;

importing the diagnostic item vector into a LSTM neural network, outputting paragraphs of the diagnostic items, wherein the number of layers of the LSTM neural network is the same as the number of the diagnostic items;

wherein, after receiving a medical image to be recognized, the processor, when executing the computer-readable instruction, further implements the following steps of:

performing binaryzation on the medical image to acquire a binarized medical image;

identifying a boundary of the binarized medical image, and dividing the medical image into a plurality of medical sub-images; if boundaries of a plurality of objects in the medical image are overlapping or adjacent, dividing the plurality of objects into the same one medical sub-image;

wherein the step of importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image comprises:

importing the medical sub-images into the VGG neural network respectively to acquire visual feature components and keyword sub-sequences of the medical sub-images;

generating the visual feature vector based on the visual feature components, and constructing the keyword sequence based on the keyword sub-sequences.

10. The device according to claim 9, wherein the step of importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image comprises:

constructing a pixel matrix of the medical image based on pixel values of pixels in the medical image and position coordinates of the pixel values;

performing dimensionality reduction on the pixel matrix through five pooling layers of the

VGG neural network to acquire the visual feature vector;

importing the visual feature vector into a fully connected layer of the VGG neural network and outputting an index sequence corresponding to the visual feature vector;

determining the keyword sequence corresponding to the index sequence according to a keyword index table.

11. The device according to claim 10, wherein the step of importing the visual feature vector and the keyword sequence into a preset diagnostic item recognition model to determine diagnostic items corresponding to the medical image comprises:

generating a keyword feature vector corresponding to the keyword sequence based on sequence numbers of keywords in a preset text corpus;

respectively importing the keyword feature vector and the visual feature vector into a preprocessing function to acquire a preprocessed keyword feature vector and a preprocessed visual feature vector; wherein the preprocessing function is specifically as:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{M} e^{z_i}}$$

where $\sigma(z_j)$ is a value of j-th element in the preprocessed keyword feature vector or in the preprocessed visual feature vector, z_j is a value of j-th element in the keyword feature vector or in the visual feature vector, *M* is the number of elements corresponding to the keyword feature vector or the visual feature vector;

determining the preprocessed keyword feature vector and the preprocessed visual feature vector as an input of the diagnostic item recognition model, and outputting the diagnostic items.

12. The device according to any one of claims 9-11, wherein the processor, when executing the computer-readable instruction, further implements the following steps of:

acquiring training visual vectors, training keyword sequences and training diagnostic items of a plurality of training images;

determining the training visual vectors and the training keyword sequences as an input of a LSTM neural network, determining the training diagnostic items as an output of the LSTM neural network, and adjusting learning parameters in the LSTM neural network so that the LSTM neural

network meets a convergence condition; wherein the convergence condition is as:

$$\theta^* = arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$$

where θ^* is the adjusted learning parameter, *Visual* is the training visual vector, *Keyword* is the training keyword sequence, *Stc* is the training diagnostic item, $p(Visual, Keyword|Stc; \theta)$ represents an output result of a probability value of the training diagnostic item when the training visual vector and the training keyword sequence are imported into the LSTM neural network with the value of the learning parameter is θ, and $arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$ is the value of the learning parameter when the probability value takes a maximum value;

determining the adjusted LSTM neural network as the diagnostic item recognition model.

13. A computer readable storage medium, stored with a computer readable instruction, wherein the computer readable instruction, when executed by a processor, implements the following steps of:

receiving a medical image to be recognized; if the medical image is acquired by scanning, collecting ambient light intensity at the time of scanning and performing differential processing on the medical image based on the ambient light intensity;

importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image;

importing the visual feature vector and the keyword sequence into a preset diagnostic item recognition model to determine diagnostic items corresponding to the medical image;

constructing a paragraph for describing each of the diagnostic items respectively based on a diagnostic item extension model;

generating a medical report for the medical image based on the paragraph, the keyword sequence and the diagnostic items;

wherein the constructing a paragraph for describing each of the diagnostic items respectively based on a diagnostic item extension model comprises:

aggregating all the diagnostic items to form a diagnostic item vector;

importing the diagnostic item vector into a LSTM neural network, outputting paragraphs of

the diagnostic items, wherein the number of layers of the LSTM neural network is the same as the number of the diagnostic items;

wherein, after receiving a medical image to be recognized, the computer readable instruction, when executed by the processor, further implements the following steps of:

performing binaryzation on the medical image to acquire a binarized medical image;

identifying a boundary of the binarized medical image, and dividing the medical image into a plurality of medical sub-images; if boundaries of a plurality of objects in the medical image are overlapping or adjacent, dividing the plurality of objects into the same one medical sub-image;

wherein the step of importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image comprises:

importing the medical sub-images into the VGG neural network respectively to acquire visual feature components and keyword sub-sequences of the medical sub-images;

generating the visual feature vector based on the visual feature components, and constructing the keyword sequence based on the keyword sub-sequences.

14. The computer readable storage medium according to claim 13, wherein the step of importing the medical image into a preset VGG neural network to acquire a visual feature vector and a keyword sequence of the medical image comprises:

constructing a pixel matrix of the medical image based on pixel values of pixels in the medical image and position coordinates of the pixel values;

performing dimensionality reduction on the pixel matrix through five pooling layers of the VGG neural network to acquire the visual feature vector;

importing the visual feature vector into a fully connected layer of the VGG neural network and outputting an index sequence corresponding to the visual feature vector;

determining the keyword sequence corresponding to the index sequence according to a keyword index table.

15. The computer readable storage medium according to claim 13, wherein the step of importing the visual feature vector and the keyword sequence into a preset diagnostic item

recognition model to determine diagnostic items corresponding to the medical image comprises:

generating a keyword feature vector corresponding to the keyword sequence based on sequence numbers of keywords in a preset text corpus;

respectively importing the keyword feature vector and the visual feature vector into a preprocessing function to acquire a preprocessed keyword feature vector and a preprocessed visual feature vector; wherein the preprocessing function is specifically as:

$$\sigma(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{M} e^{z_i}}$$

where $\sigma(z_j)$ is a value of j-th element in the preprocessed keyword feature vector or in the preprocessed visual feature vector, z_j is a value of j-th element in the keyword feature vector or in the visual feature vector, *M* is the number of elements corresponding to the keyword feature vector or the visual feature vector;

determining the preprocessed keyword feature vector and the preprocessed visual feature vector as an input of the diagnostic item recognition model, and outputting the diagnostic items.

16. The computer readable storage medium according to any one of claims 13-15, wherein the computer readable instruction, when executed by the processor, further implements the following steps of:

acquiring training visual vectors, training keyword sequences and training diagnostic items of a plurality of training images;

determining the training visual vectors and the training keyword sequences as an input of a LSTM neural network, determining the training diagnostic items as an output of the LSTM neural network, and adjusting learning parameters in the LSTM neural network so that the LSTM neural network meets a convergence condition; wherein the convergence condition is as:

$$\theta^* = arg\ max_\theta \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$$

where θ^* is the adjusted learning parameter, *Visual* is the training visual vector, *Keyword* is the training keyword sequence, *Stc* is the training diagnostic item, $p(Visual, Keyword|Stc; \theta)$

represents an output result of a probability value of the training diagnostic item when the training visual vector and the training keyword sequence are imported into the LSTM neural network with the value of the learning parameter is θ, and $arg\ max_{\theta} \sum_{Stc} logp(Visual, Keyword|Stc; \theta)$ is the value of the learning parameter when the probability value takes a maximum value;

determining the adjusted LSTM neural network as the diagnostic item recognition model.