

METHOD, APPARATUS AND DEVICE FOR DETECTING AND ANALYZING TEXT

TECHNICAL FIELD

[0001] The present invention relates to the field of image processing technologies, in particular to a method, an apparatus and a device for detecting and analyzing a text.

BACKGROUND

[0002] Object detection is a kind of computer technology related to Computer Vision (CV) and image processing, which detects semantic object instances with specific categories (such as humans, buildings, cars, etc.) from digital images and videos. The object detection is well developed in the fields of face detection and pedestrian detection. The object detection has a large number of application scenarios in the field of the Computer Vision, including image retrieval and video surveillance.

[0003] Template analysis is a kind of computer technology related to optical character recognition, which maps a corresponding text area in a digital document image to a semantic object instance, thereby transforming unstructured data such as the digital document image into structured data related to the semantic object instance. The template analysis maps each text area in a document to a semantic object instance to structure document image content. In the existing methods, the matching of text area is based on the traditional image feature matching methods. The brightness, contrast, and resolution of the image seriously affect the accuracy of the text area matching. In the existing methods, the matching of text area is to determine the text area by a fixed relative position, but a handwritten text, a needle-printed text and so on may have unpredictable position offsets, and the locating of the text area is often wrong.

SUMMARY

[0004] In order to solve the problems of the prior art, embodiments of the present invention provide a method, an apparatus and a device for detecting and analyzing a text. The technical schemes are as follows:

[0005] In a first aspect, embodiments of the present invention provide a method for

detecting and analyzing a text, including: detecting and classifying a text area of an image to be detected by using a preset deep neural network detection model to generate text area information with categories; and performing template matching based on labeled template information and the text area information with the categories to generate structured information data.

[0006] In some embodiments of the present invention, the method in the first aspect further includes: generating the labeled template information by performing template labeling on a template image. The template image is the same as a master of the image to be detected, and the text area of the image to be detected includes a first anchor text area and a first non-anchor text area, and a text area of the template image includes a second anchor text area and a second non-anchor text area.

[0007] In some embodiments of the present invention, the generating the labeled template information by performing template labeling on a template image includes: generating the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the second non-anchor text area.

[0008] In some embodiments of the present invention, the detecting and classifying a text area of an image to be detected by using a preset deep neural network detection model to generate text area information with categories includes: detecting and classifying the text area of the image to be detected by using the preset deep neural network detection model to obtain text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

[0009] In some embodiments of the present invention, the performing template matching based on labeled template information and the text area information with the categories to generate structured information data includes: determining positions of the first anchor text area and the first non-anchor text area on the image to be detected by matching the second anchor text area and the first anchor text area, and the second non-anchor text area and the first non-anchor text area; and determining a mapping relationship between an entity and the first anchor text area and the first non-anchor text area according to a mapping relationship between the entity and the second anchor text area and the second non-anchor text area, and removing, based on the positions of the first anchor text area and the first non-anchor text area, a redundant part including an

non-text area and an irrelevant text in the text area of the image to be detected to generate the structured information.

[0010] In some embodiments of the present invention, the method in the first aspect further includes: performing image preprocessing on the image to be detected. The image preprocessing includes image correction and/or scaling to a preset size.

[0011] In some embodiments of the present invention, the method in the first aspect further includes: training the preset deep neural network detection model.

[0012] In some embodiments of the present invention, the training the preset deep neural network detection model includes: generating samples by using a sample generation tool; and training the preset deep neural network detection model by using the samples.

[0013] In some embodiments of the present invention, the image to be detected is a new sample, and the method in the first aspect further includes: fine-tuning the preset deep neural network detection model by using the new sample.

[0014] In a second aspect, embodiments of the present invention provide an apparatus for detecting and analyzing a text, including: a text area detection module, configured to detect and classify a text area of an image to be detected by using a preset deep neural network detection model to generate text area information with categories; and a matching module, configured to perform template matching based on labeled template information and the text area information with the categories to generate structured information data.

[0015] In some embodiments of the present invention, the apparatus in the second aspect further includes: a labeling module, configured to generate the labeled template information by performing template labeling on a template image. The template image is the same as a master of the image to be detected, and the text area of the image to be detected includes a first anchor text area and a first non-anchor text area, and a text area of the template image includes a second anchor text area and a second non-anchor text area.

[0016] In some embodiments of the present invention, the labeling module is configured to: generate the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the second non-anchor text area.

[0017] In some embodiments of the present invention, the text area detection module is configured to: detect and classify the text area of the image to be detected by using the preset deep neural network detection model to obtain text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

5 [0018] In some embodiments of the present invention, the matching module is configured to: determine positions of the first anchor text area and the first non-anchor text area on the image to be detected by matching the second anchor text area and the first anchor text area, and the second non-anchor text area and the first non-anchor text area; and determine a mapping relationship between an entity and the first anchor text area and

10 the first non-anchor text area according to a mapping relationship between the entity and the second anchor text area and the second non-anchor text area, and remove, based on the positions of the first anchor text area and the first non-anchor text area, a redundant part including an non-text area and an irrelevant text in the text area of the image to be detected to generate the structured information.

15 [0019] In some embodiments of the present invention, the apparatus in the second aspect further includes: an image preprocessing module, configured to perform image preprocessing on the image to be detected. The image preprocessing includes image correction and/or scaling to a preset size.

[0020] In some embodiments of the present invention, the apparatus in the second

20 aspect further includes: a model training module, configured to train the preset deep neural network detection model.

[0021] In some embodiments of the present invention, the model training module is configured to: generate samples by using a sample generation tool; and train the preset deep neural network detection model by using the samples.

25 [0022] In some embodiments of the present invention, the image to be detected is a new sample, and the apparatus in the second aspect further includes: a fine-tuning module, configured to fine-tune the preset deep neural network detection model by using the new sample.

[0023] In a third aspect, embodiments of the present invention provide a device for

30 detecting and analyzing a text, including: a processor; and a memory, configured to store executable instructions of the processor. The processor is configured to execute the method for detecting and analyzing a text described in the first aspect by the executable instructions.

[0024] In a fourth aspect, embodiments of the present invention provide a computer readable storage medium, which stores a computer program, and the computer program is configured to execute the method for detecting and analyzing a text described in the first aspect.

5 [0025] A method, an apparatus and a device for detecting and analyzing a text are provided by the embodiments of the present invention, by detecting and classifying a text area of an image by using a preset deep neural network detection model, and then performing template matching by combining labeled template information and text area information with categories obtained from detecting and classifying to generate structured
10 information data, which may realize quick and accurate detection and analysis of various fields in a document image, and have characteristics of real-time, accurate, general, robust and scalable.

BRIEF DESCRIPTION OF DRAWINGS

[0026] In order to more clearly illustrate the technical solutions in the embodiments of
15 the present invention, the accompanying drawings used in the description of the embodiments will be briefly described below. It is obvious that the accompanying drawings in the following description are only some embodiments of the present invention. For those of ordinary skill in the art, other accompanying drawings may also be obtained according to these accompanying drawings without any creative works.

20 [0027] FIG. 1 is a schematic flowchart illustrating a method for detecting and analyzing a text according to a first embodiment of the present invention.

[0028] FIG. 2 is a schematic diagram illustrating a template matching process according to a preferred embodiment.

[0029] FIG. 3 is a schematic flowchart illustrating a method for detecting and analyzing
25 a text according to a second embodiment of the present invention.

[0030] FIG. 4 is a schematic structural diagram illustrating an apparatus for detecting and analyzing a text according to a third embodiment of the present invention.

[0031] FIG. 5 is a schematic composition diagram illustrating a device for detecting and analyzing a text according to an embodiment of the present invention.

30 [0032] FIG. 6 to FIG. 9 are bill processing effect diagrams illustrating an application example of a method, an apparatus and a device for detecting and analyzing a text

according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0033] In order to make the objects, technical solutions and advantages of the present invention clearer, the technical solutions in the embodiments of the present invention are
5 clearly and completely described in the following with reference to the accompanying drawings in the embodiments of the present invention. It is obvious that the described embodiments are only a part of the embodiments of the present invention, but not all of the embodiments. All other embodiments obtained by those of ordinary skill in the art based on the embodiments of the present invention without creative works are within the
10 protection scope of the present invention. It should be noted that, in the description of the present invention, the meaning of "a plurality" is two or more, unless it is specifically defined otherwise.

[0034] A method, an apparatus and a device for detecting and analyzing a text provided by the embodiments of the present invention will be described in detail below with
15 reference to specific embodiments and accompanying drawings.

[0035] First Embodiment

[0036] FIG. 1 is a schematic flowchart illustrating a method for detecting and analyzing a text according to a first embodiment of the present invention. As shown in FIG. 1, the method for detecting and analyzing a text provided by the embodiment of the present
20 invention includes the following contents.

[0037] 101: detecting and classifying a text area of an image to be detected by using a preset deep neural network detection model to generate text area information with categories.

[0038] A neural network is a kind of network structure which is artificially designed,
25 whose essence is Multi-layer Perceptron. The perceptron is composed of several neurons. Each neuron receives an input signal from an external or other node and obtains an output signal by an activation function, just like the signal transmission between the neurons in the brain. The neurons are connected in layers to form the network structure. Unlike neural cells, signals from artificial neurons can propagate backwards. This feedback
30 mechanism allows the perceptron to have a learning function. In addition to the learning function, the Multi-layer Perceptron may represent a non-linear mapping, so the neural

network can help people solve relatively complex problems, such as pattern recognition, automatic control, decision evaluation, prediction, and the like.

[0039] A convolutional neural network is a kind of feed-forward neural network whose neurons may respond to the surrounding cells within a part of the coverage to form a receptive field, which has a better performance on image processing. At the same time, by weight sharing, the convolutional neural network saves space compared to the traditional fully connected neural network. Features learned with the convolutional neural network have certain invariance in position and scale. An Optical Character Recognition (OCR) model trained with the convolutional neural network is robust to changes in the font and the word size.

[0040] A detection algorithm of deep learning mainly includes three parts: the selection of detection windows, the extraction of features and the selection of classifiers. In the early days of the deep learning, sliding windows were mostly used for window extraction. Recent detection algorithms, such as Faster R-CNN, directly use Region Proposal Networks (RPNs) to calculate candidate frames. The RPN takes a picture of any size as an input and outputs a batch of rectangular areas, each area corresponds to one target score and position information. With the help of strong non-linear modeling capabilities of the convolutional neural network, it is possible to extract the features that fit various subtasks. The detection algorithm of the deep learning also breaks the framework of the traditional sliding frame and uses a regression method to solve detection problems.

[0041] In this embodiment, the preset deep neural network detection model may adopt a Faster R-CNN network model, a Mask R-CNN network model, or any other possible deep neural network model, which is not particularly limited in the embodiments of the present invention.

[0042] Specifically, the image to be detected may be an image with text information, such as an invoice, a receipt, a bill and the like. The image may be obtained by shooting or scanning a paper document, or may be an electronic version directly.

[0043] For example, the image to be detected is an invoice, and the invoice is marked with some item categories, such as "Invoice Code", "Invoice Number", "Payer Name", "Payee Name", "Billing Date", and so on. A version when the invoice is not filled in may be called a master of the invoice. The text information on the master may be entered by an electronic device; and the image obtained after the invoice is filled in (may be filled in electronically or handwritten) may be called the image to be detected.

[0044] The text area of the image to be detected may include original text information on the master and text information filled in by the user.

[0045] Adopting the deep neural network detection model can accurately find the text areas of different scales (word sizes) and perform preliminary classification on text contents, such as amount of money, item name, entry, bill number and other categories, so that the text in the document may be detected accurately.

[0046] In an embodiment, the categories of the text area information may be divided into: the original text information on the master and the text information filled in by the user. Further, the categories of the text area information may also be divided into: the text information of date, Chinese character, English and so on. By using the preset deep neural network detection model to detect the text area of the image to be detected, the text area information with the categories may be obtained. Specifically, the text area information with the categories may include categories and position information of the text, and the like.

[0047] The preset deep neural network detection model may detect and classify an image to be detected of one master, and may also detect and classify images to be detected of a variety of different masters separately, which depends on sample categories during training.

[0048] 102: performing template matching based on labeled template information and the text area information with the categories to generate structured information data. The labeled template information includes information of an entity that the user needs to obtain, and the entity is information that the user needs to obtain from the image to be detected.

[0049] The labeled template information may be obtained by labeling the image to be detected or the master of the image to be detected. The information of the entity may include a category and position information of the entity, and so on. For example, the category of the entity that the user needs to obtain is "Payer Name". The labeled template information includes the category of the entity and the position information of the entity (payer's specific name) in the image to be detected. The position information may be absolute position information or relative position information with respect to other text information.

[0050] When the position information is the relative position information with respect to other text information, the position of the entity that the user needs to obtain may be

accurately obtained by performing the template matching on the labeled template information and the text area information with the categories, i.e., the corresponding text area in the image to be detected is mapped to a semantic object instance, so as to realize the structuring of the content of the image to be detected.

5 [0051] Herein, the entity that the user needs to obtain may be one or more. The recognition process of the specific content of the entity may be carried out after the matching of the labeled template information and the text area information with categories is completed, or may be carried out in 101. The recognition process involved herein may use a recognition model such as tesseract or CRNN, which is not particularly
10 limited in the embodiments of the present invention.

[0052] A method for detecting and analyzing a text is provided by the embodiments of the present invention, by detecting and classifying a text area of an image by using a preset deep neural network detection model, and then performing template matching by combining labeled template information and text area information with categories
15 obtained from detecting and classifying to generate structured information data, which may realize quick and accurate detection and analysis of various fields in a document image, and have characteristics of real-time, accurate, general, robust and scalable.

[0053] According to an embodiment of the present invention, the method in FIG. 1 further includes: generating the labeled template information by performing template
20 labeling on a template image. The template image is the same as a master of the image to be detected, and the text area of the image to be detected includes a first anchor text area and a first non-anchor text area, and a text area of the template image includes a second anchor text area and a second non-anchor text area.

[0054] Specifically, the anchor text area may be an area corresponding to the text
25 information of the master; and the non-anchor text area may be an area corresponding to other text information except the text information of the master, for example, a text area requiring the user to fill in.

[0055] The anchor text area may include one or more anchor points, each anchor point may be a field, and a plurality of anchor points may be distributed in different positions.
30 The plurality of anchor points are stable features in each template image, and contents and relative positions of the plurality of anchor points may not change, for example, a header in a table. By setting a plurality of anchor points, the relative position relationship of each area in the current template image may be completely determined.

[0056] The template image may be a blank master, an image to be detected, or another image (may be filled in by another user) that has the same master as the image to be detected. When the template image is the blank master, the second non-anchor text area may be a blank area that the user has not yet filled in.

5 [0057] Further, the generating the labeled template information by performing template labeling on a template image includes: generating the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the second non-anchor text area.

10 [0058] Specifically, the template labeling process is used to label positions and categories of all the fields that need to be recognized. The categories include: whether or not the fields are the anchor points, and whether the text information is dates, Chinese characters, English, and so on. The generated labeled template information is used for the subsequent template matching. A relative position relationship between a second anchor
15 point and a second non-anchor point, and a size of a text area corresponding to each second anchor point and each second non-anchor point may be determined by performing the template labeling. A size of a text area corresponding to a first anchor point may refer to a height and a width of the text area corresponding to the first anchor point, i.e., a shape of the text area corresponding to the first anchor point may be rectangular, of
20 course, may also be circular, oval, or other shapes, which is not specifically limited in the embodiments of the present invention.

[0059] Further, when there are a plurality of second anchor points and second non-anchor points, a relative position relationship between each second anchor point and a relative position relationship between each second non-anchor point may also be
25 determined by performing the template labeling.

[0060] Similar to the preset deep neural network detection model, the template images to be labeled herein are not limited to one or two types. According to actual needs, a variety of template images may be set in advance to obtain a variety of labeled template information.

30 [0061] In this embodiment, 101 includes: detecting and classifying the text area of the image to be detected by using the preset deep neural network detection model to obtain text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

[0062] Specifically, the text area information with categories may include text area information of the first anchor text area and the first non-anchor text area.

[0063] Further, 102 includes: determining positions of the first anchor text area and the first non-anchor text area on the image to be detected by matching the second anchor text area and the first anchor text area, and the second non-anchor text area and the first non-anchor text area; and determining a mapping relationship between an entity and the first anchor text area and the first non-anchor text area according to a mapping relationship between the entity and the second anchor text area and the second non-anchor text area, and removing, based on the positions of the first anchor text area and the first non-anchor text area, a redundant part including an non-text area and an irrelevant text in the text area of the image to be detected to generate the structured information.

[0064] Specifically, an absolute position of each first anchor point and each first non-anchor point on the image to be detected may be determined by using the preset deep neural network detection model. However, in the actual detection process, since an image itself is tilted or individual fields are unclear, some first anchor points and some first non-anchor points may not be detected or a detected position of each first anchor point and each first non-anchor point may be shifted from the actual position. In this embodiment, by combining the labeled template information and the text area information with categories, i.e., combining the absolute position and the relative position, the positions of the first anchor point and the first non-anchor point with positioning offsets may be adjusted, or the positions of the first anchor point and the first non-anchor point that have not been detected may be located, thereby accurately locating the position and content information of the desired text area.

[0065] In addition, the positions of the first anchor point and the first non-anchor point may also be determined by matching the size of the text area. Especially when the image to be detected is an image obtained based on a handwritten invoice, when different users fill in the information, the filling position of the information may have different deviations. At this time, a matching may be performed between the size of the text area corresponding to the second non-anchor point and the size of the text area corresponding to the first non-anchor point. When a matching degree satisfies a first threshold, the first non-anchor point is considered to be matched successful.

[0066] Based on the relative position relationship between each anchor point and each

non-anchor point, the positions of other anchor points and non-anchor points that have not been successfully matched may be determined by the positions of the anchor points and non-anchor points that have been successfully matched, thereby improving the accuracy of a matching result and the recognition rate.

5 [0067] In this embodiment, accurate matching can be achieved for the anchor points with stable position and stable content information, which improves the accuracy of matching steps. In addition, by excluding the non-text area and the irrelevant text, it is easier to grasp key information and further improve the accuracy of a detection result.

[0068] According to an embodiment of the present invention, the method in FIG. 1
10 further includes: performing image preprocessing on the image to be detected. The image preprocessing includes image correction and/or scaling to a preset size.

[0069] Specifically, since the actual detected image may have problems of inclination or inconsistency in size, which may affect the accuracy of locating the positions of the anchor point and the non-anchor point during detection. Therefore, in this embodiment,
15 the image to be detected may be corrected or scaled before 101. When the image to be detected is scaled to a size of a sample used in the training of the preset deep neural network detection model, the accuracy of the matching process may be improved. In this embodiment, a size of the template image may be consistent with the size of the sample used in the training of the preset deep neural network detection model. Of course,
20 corresponding settings of the image preprocessing operation may also be performed according to the actual situations.

[0070] FIG. 2 is a schematic diagram illustrating a template matching process according to a preferred embodiment. As shown in FIG. 2, the above-described 102 may further include the following content.

25 [0071] 1021: performing the template matching by combining the labeled template information and the text area information with categories.

[0072] Specifically, the labeled template information is matched with the text area information with categories to determine the positions of the first anchor text area and the first non-anchor text area, i.e., a plurality of first anchor points and a plurality of first
30 non-anchor points are located. In this embodiment, the matching process may make up for the first anchor point and the first non-anchor point that are not detected by the preset deep neural network detection model, or adjust the positions of the first anchor point and the first non-anchor point whose locating results are not accurate.

[0073] 1022: removing the redundant part by the mapping of the entity and the text area to generate the structured information data.

[0074] The text area with categories is mapped to a corresponding entity by the first anchor point and the first non-anchor point to remove the redundant part including the non-text area and the irrelevant text, and then the anchor point and non-anchor point which are corresponding to the entity are determined. The text information at the non-anchor point may be determined by recognition, and the category of the text information (i.e., text information of the anchor point such as name, gender, age, etc.) may be determined according to the relative position between the recognized text information and the anchor point. The structured information data may be generated by combining the text information and the category of the text information. The structured information data may be output for corresponding applications.

[0075] According to an embodiment of the present invention, the method in FIG. 1 further includes: training the preset deep neural network detection model.

[0076] Specifically, a sample for training the preset deep neural network detection model may be labeled with a position of an anchor point in advance. That is, during training, the text area in the sample is classified (including but not limited to classification of anchor points and non-anchor points), and then the detection model is trained.

[0077] In an embodiment, the training the preset deep neural network detection model includes: generating samples by using a sample generation tool; and training the preset deep neural network detection model by using the samples. In this embodiment, a large number of samples may be obtained quickly by using the sample generation tool.

[0078] According to an embodiment of the present invention, the image to be detected is a new sample. The method in FIG. 1 further includes: fine-tuning the preset deep neural network detection model by using the new sample.

[0079] Specifically, by forming a data backflow in a detection application, more new samples may be obtained. The preset deep neural network detection model may be fine-tuned by using the new samples. In this way, at the initial stage of training, it does not need a large number of labeled samples to train the detection model, and the recognition rate of the model may be gradually improved after the data backflow is formed.

[0080] Second Embodiment

[0081] FIG. 3 is a schematic flowchart illustrating of a method for detecting and

analyzing a text according to a second embodiment of the present invention. As shown in FIG. 3, the method for detecting and analyzing a text provided by the embodiment of the present invention includes the following contents.

[0082] 201: training the preset deep neural network detection model.

5 [0083] Specifically, the training process may refer to the description in FIG. 1.

[0084] In the detection application, it is possible to form the data backflow to obtain more new samples, and the preset deep neural network detection model is fine-tuned by using the new samples.

10 [0085] During this training or fine-tuning process, the text lines in the sample are classified (including but not limited to classification of anchor points and non-anchor points), and then the detection model is trained.

[0086] It is worth noting that, in addition to the foregoing implementation manners, 201 may also be implemented in other manners, and a specific manner is not limited in the embodiments of the present invention.

15 [0087] 202: generating the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the second non-anchor text area.

20 [0088] This process is used to label positions and categories of all the fields that need to be recognized on the template image. The categories may include: whether or not the fields are the anchor points, and whether the text information is dates, Chinese characters, English, and so on. The generated labeled template information is used for the subsequent template matching. It should be noted that the template images to be labeled herein are not limited to one or two types. According to actual needs, a variety of template images
25 may be set in advance to obtain a variety of labeled template information.

[0089] It is worth noting that 202 may be implemented in other manners besides the above-mentioned implementation manners, and a specific manner is not limited in the embodiments of the present invention.

30 [0090] 203: performing image preprocessing on the image to be detected, the image preprocessing including image correction and/or scaling to a preset size. The content of the image preprocessing herein includes, but is not limited to, the image correction and/or the scaling to the preset size, and corresponding settings of the image preprocessing operation may also be performed according to the actual situations.

[0091] It is worth noting that 203 may be implemented in other manners besides the above-mentioned implementation manners, and a specific manner is not limited in the embodiments of the present invention.

[0092] 204: detecting and classifying the text area of the image to be detected by using
5 the preset deep neural network detection model to obtain the text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

[0093] The preset deep neural network detection model may adopt a Faster R-CNN network model, a Mask R-CNN network model, or any other possible deep neural network model in the prior art, which is not particularly limited in the embodiments of the
10 present invention.

[0094] Adopting a deep neural network object detection framework can accurately find the text areas of different scales (word sizes) and perform preliminary classification on text contents, such as amount of money, item name, entry, bill number and other categories.

[0095] It is worth noting that 204 may be implemented in other manners besides the
15 above-mentioned implementation manners, and a specific manner is not limited in the embodiments of the present invention.

[0096] 202 may be executed before 201 or simultaneously with 201, which is not limited in the embodiments of the present invention.

[0097] 205: performing template matching on the labeled template information and the
20 text area information with categories to locate a first anchor point and a first non-anchor point, mapping the text area with categories to a corresponding entity by the first anchor point and the first non-anchor point, and removing a redundant part including a non-text area and an irrelevant text to generate the structured information data.

[0098] Successful anchor locating is equivalent to the relative position of each field of
25 an entire layout is determined, and then the content information and position of the desired field may be accurately located by combining the relative positions of the layout and previous anchor text area information, and the previous non-anchor text area information. Therefore, stable anchor points can achieve accurate matching and improve
30 the accuracy of the matching steps.

[0099] The text information at the non-anchor point may be determined by recognition, and the category of the text information (i.e., text information of the anchor point such as name, gender, age, etc.) may be determined according to the relative position between the

recognized text information and the anchor point. The structured information data may be generated by combining the text information and the category of the text information. The structured information data may be output for corresponding applications.

5 [00100] It is worth noting that 205 may be implemented in other manners besides the above-mentioned implementation manners, and a specific manner is not limited in the embodiments of the present invention.

[00101] It should be noted that the operation order of specific steps in the method for detecting and analyzing a text provided in the embodiments of the present invention is not limited to the order of steps in the above methods, and without departing from the
10 inventive concept of the present invention, a corresponding transformation design may be performed according to the actual situations.

[00102] Third Embodiment

[00103] FIG. 4 is a schematic structural diagram illustrating an apparatus 30 for detecting and analyzing a text according to a third embodiment of the present invention. As shown
15 in FIG. 4, the apparatus 30 for detecting and analyzing a text provided in the embodiments of the present invention includes a text area detection module 31 and a matching module 32.

[00104] The text area detection module 31 is configured to detect and classify a text area of an image to be detected by using a preset deep neural network detection model to
20 generate text area information with categories; and the matching module 32 is configured to perform template matching based on labeled template information and the text area information with the categories to generate structured information data. The labeled template information includes information of an entity that the user needs to obtain. The entity is information that the user needs to obtain from the image to be detected.

25 [00105] Specifically, structured information may be output finally, and a unified interface may be provided to facilitate the call of other external modules.

[00106] An apparatus for detecting and analyzing a text is provided by the embodiments of the present invention, by detecting and classifying a text area of an image by using a preset deep neural network detection model, and then performing template matching by
30 combining labeled template information and text area information with categories obtained from detecting and classifying to generate structured information data, which may realize quick and accurate detection and analysis of various fields in a document image, and have characteristics of real-time, accurate, general, robust and scalable.

[00107] According to an embodiment of the present invention, the apparatus 30 further includes: a labeling module 33, configured to generate the labeled template information by performing template labeling on a template image. The template image is the same as a master of the image to be detected, and the text area of the image to be detected includes a first anchor text area and a first non-anchor text area, and a text area of the template image includes a second anchor text area and a second non-anchor text area.

[00108] According to an embodiment of the present invention, the labeling module 33 is configured to: generate the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the second non-anchor text area.

[00109] According to an embodiment of the present invention, the text area detection module 31 is configured to: detect and classify the text area of the image to be detected by using the preset deep neural network detection model to obtain text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

[00110] According to an embodiment of the present invention, the matching module 32 is configured to: determine positions of the first anchor text area and the first non-anchor text area on the image to be detected by matching the second anchor text area and the first anchor text area, and the second non-anchor text area and the first non-anchor text area; and determine a mapping relationship between an entity and the first anchor text area and the first non-anchor text area according to a mapping relationship between the entity and the second anchor text area and the second non-anchor text area, and remove, based on the positions of the first anchor text area and the first non-anchor text area, a redundant part including an non-text area and an irrelevant text in the text area of the image to be detected to generate the structured information.

[00111] According to an embodiment of the present invention, the apparatus 30 further includes: an image preprocessing module 34, configured to perform image preprocessing on the image to be detected. The image preprocessing includes image correction and/or scaling to a preset size.

[00112] According to an embodiment of the present invention, the apparatus 30 further includes: a model training module 35, configured to train the preset deep neural network detection model.

[00113] According to an embodiment of the present invention, the model training module

35 is configured to: generate samples by using a sample generation tool; and train the preset deep neural network detection model by using the samples.

[00114] According to an embodiment of the present invention, the image to be detected is a new sample, and the apparatus 30 further includes: a fine-tuning module 36, configured to fine-tune the preset deep neural network detection model by using the new sample.

[00115] It should be understood that the operations and functions of the text area detection module 31, the matching module 32, the labeling module 33, the image preprocessing module 34, the model training module 35, and the fine-tuning module 36 in the above embodiments may be referred to the descriptions of the method for detecting and analyzing a text provided in FIG. 1, FIG. 2 and FIG. 3. In order to avoid repetition, details are not repeated herein again.

[00116] Fourth Embodiment

[00117] FIG. 5 is a schematic composition diagram illustrating a device 40 for detecting and analyzing a text according to an embodiment of the present invention. As shown in FIG. 5, the device 40 for detecting and analyzing a text provided in the embodiments of the present invention includes a processor 41 and a memory 42.

[00118] Specifically, the memory 42 is used to store executable instructions of the processor 41. The processor 41 is configured to execute the steps of the method for detecting and analyzing a text described in the first embodiment and the second embodiment by the executable instructions. For the specific implementation process, please refer to the corresponding content in the first embodiment and the second embodiment, and details are not repeated herein again.

[00119] Application Example

[00120] In the aspect of system construction, it is divided into hardware requirements and software requirements.

[00121] In the aspect of hardware, one server, one computing device, and one image input device are required.

[00122] In the aspect of software, a web service set up on the server for bill recognition, a page for terminal display, and a database for storing images and detection results are required.

[00123] In the aspect of process, it is divided into a model generation process and a document detection process.

[00124] In the aspect of model generation, firstly, training samples are generated by a

sample generation module. If there is sufficient manpower, real samples may be labeled manually. After the training samples are sufficient (usually on the order of magnitude of 100,000), a bill recognition model is trained on the server by using a deep neural network, and stored on the server to set up the web service.

5 [00125] In the aspect of detection, firstly, a document is scanned with the image input device. After a document image is obtained, a page for the recognition service is opened from the terminal, and the image is uploaded to the server. After receiving the image, the server obtains structured detection results by using an image preprocessing module, a text detection module, and a template matching module, and then recognizes specific text
10 content of a corresponding area by using a recognition module, and then post-processes recognition results according to the specific situations. Finally, the recognition results are sent back to the terminal, so as to display to a user.

[00126] The following specifically takes the recognition of a medical bill as an example. Suppose we need to recognize the hospital name, the hospital type, the medical insurance
15 type, the corresponding item, and the amount of money on the bill, and the hospital name, the hospital type, the medical insurance type, the corresponding item, and the amount of money are the entities that need to be obtained.

[00127] In the first step, obtaining an original bill image shown in FIG. 6.

[00128] In the second step, as shown in FIG. 7, performing layout analysis on the bill
20 image.

[00129] Labeled template information is generated by performing layout analysis on the bill image, that is, performing template labeling. Here, the image to be detected and the template image are the same image. Specifically, for the recognition of a plurality of bills with the same master, template labeling can be performed only once.

25 [00130] In the third step, as shown in FIG. 8, finding text areas of different categories.

[00131] Specifically, the text areas of the image to be detected are detected and classified by using a preset deep neural network detection model to generate text area information with categories, so as to distinguish the text areas of different categories.

[00132] In the fourth step, as shown in FIG. 9, extracting the area to be recognized
30 according to the layout analysis and classification of the text areas.

[00133] Specifically, the template labeling is performed based on the labeled template information and the text area information with categories to extract the areas that need to be recognized.

[00134] In this embodiment, according to the mapping relationship between the entity and the text area, a redundant part including a non-text area and an irrelevant text needs to be removed. For example, as shown in FIG. 8, the text area pointed by arrow A is the irrelevant text, the text area pointed by arrow B is the non-text area, and the text area pointed by arrow A and the text area pointed by arrow B are the redundant part to be removed.

[00135] In the fifth step, recognizing the above areas by using a corresponding recognition model.

[00136] The recognition results are:

[00137] "上海第一妇婴保健院(Shanghai First Maternity and Infant Hospital)"

[00138] "三级甲等专科医院(3-A-Class Specialized Hospital)";

[00139] "医保自费(Medical Insurance at Own Expense)";

[00140] "治疗费（自费）(Treatment Fee (at own expense))";

[00141] "门诊诊疗费(Outpatient Consultation Fee)";

[00142] "6.00";

[00143] "800".

[00144] The above recognition results may be obtained by using the recognition models such as tesseract, CRNN, etc. The recognition results may also deviate from the real results, so the recognition results may be corrected by a post-processing process.

[00145] In the sixth step, post-processing the recognition results.

[00146] The post-processing results are:

[00147] "上海市第一妇婴保健院(Shanghai City First Maternity and Infant Hospital)"
(Check the list of the hospital name for matching);

[00148] "三级甲等专科医院(3-A-Class Specialized Hospital)";

[00149] "医保自费(Medical Insurance at Own Expense)";

[00150] "治疗费（自费）(Treatment Fee (at own expense))";

[00151] "门诊诊疗费(Outpatient Consultation Fee)";

[00152] "6.00";

[00153] "8.00" (matching according to the rule of the amount of money).

[00154] Herein, the hospital name "Shanghai First Maternity and Infant Hospital" in the recognition results may be corrected by querying the list of the hospital name. The

correction result is "Shanghai City First Maternity and Infant Hospital", which means that the hospital name in the recognition results is missing a word of "City". Similarly, each information in the recognition results may be corrected separately. For example, the outpatient consultation fee in the recognition results is "800", but after matching the rule of the amount of money, it is corrected to "8.00".

[00155] In the seventh step, formatting the recognition results to generate structured information data.

[00156] {
[00157] Hospital Name: 上海市第一妇婴保健院(Shanghai First Maternity and Infant Hospital)

[00158] Hospital Type: 三级甲等专科医院(3-A-Class Specialized Hospital)

[00159] Medical Insurance Type: 医保自费(Medical Insurance at Own Expense)

[00160] Cost list:

[00161] {

[00162] 治疗费（自费）(Treatment Fee (at own expense)): 6.00

[00163] 门诊诊疗费 (Outpatient Consultation Fee): 8.00

[00164] }

[00165] }

[00166] It should be noted that an apparatus and a device for detecting and analyzing a text provided in the foregoing embodiments are only illustrated as examples by the division of the above functional modules when performing the text detection and analysis. In actual applications, the above functions may be allocated by different functional modules according to needs, i.e., the internal structures of the apparatus and the device are divided into different functional modules to complete all or part of the functions described above. In addition, the embodiments of the method, the apparatus, and the device for detecting and analyzing a text provided in the foregoing embodiments belong to a same concept, and the specific implementation process is detailed in the embodiments of the method, which will not be described herein.

[00167] A person of ordinary skill in the art may understand that all or part of the steps for implementing the foregoing embodiments may be implemented by hardware, or may be implemented by a program instructing related hardware. The program may be stored in a computer readable storage medium. The storage medium mentioned above may be a

read-only memory, a magnetic disk and an optical disk, etc.

[00168] The embodiments of the present invention are described with reference to flowcharts and/or block diagrams of a method, a device (system), and a computer program product according to the embodiments of the present invention. It should be understood that each flow in the flowcharts and/or each block in the block diagrams, and a combination of the flows in the flowcharts and/or the blocks in the block diagrams may be implemented by computer program instructions. These computer program instructions may be provided to a general purpose computer, a special purpose computer, an embedded processor, or processors of other programmable data processing devices to produce a machine, causing instructions executed by the computer or the processors of other programmable data processing devices to generate an apparatus used to implement the functions specified in one or more flows in the flowcharts and/or one or more blocks in the block diagrams.

[00169] These computer program instructions may also be stored in a computer readable memory that can direct the computer or the other programmable data processing devices to operate in a particular manner, such that the instructions stored in the computer readable memory produce an article of manufacture including an instruction apparatus. The instruction apparatus implements the functions specified in one or more flows in the flowcharts and/or one or more blocks in the block diagrams.

[00170] These computer program instructions may also be loaded onto a computer or other programmable data processing devices, so that a series of operational steps are performed on the computer or the other programmable data devices to produce a computer-implemented processing, and the instructions executed in the computer or the other programmable data devices provide steps for realizing the functions specified in one or more flows in the flowcharts and/or one or more blocks in the block diagrams.

[00171] Although preferred embodiments in the embodiments of the present invention have been described, those of ordinary skill in the art can make further changes and modifications to these embodiments once they know the basic inventive concept. Therefore, the appended claims are intended to be interpreted as including the preferred embodiments and all the changes and modifications falling within the scope of the embodiments of the present invention.

[00172] It is apparent that those of ordinary skill in the art can make various modifications and variations to the present invention without departing from the spirit and

scope of the present invention. Thus, if these modifications and variations of the present invention fall within the scope of the claims of the present invention and its equivalent technologies, the present invention also intends to include these modifications and variations.

- 5 [00173] The above are only better embodiments of the present invention and not intended to limit the present invention. Any modification, equivalent replacement, improvement, etc. made within the spirit and principles of the present invention shall be included in the protection scope of the present invention.

What is claimed is:

1. A method for detecting and analyzing a text, comprising:

detecting and classifying a text area of an image to be detected by using a preset deep neural network detection model to generate text area information with categories;
5 and

performing template matching based on labeled template information and the text area information with the categories to generate structured information data.

2. The method according to claim 1, further comprising:

generating the labeled template information by performing template labeling on a
10 template image,

wherein the template image is the same as a master of the image to be detected, and the text area of the image to be detected comprises a first anchor text area and a first non-anchor text area, and a text area of the template image comprises a second anchor text area and a second non-anchor text area.

3. The method according to claim 2, wherein the generating the labeled template information by performing template labeling on a template image comprises:

generating the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the
20 second non-anchor text area.

4. The method according to claim 2 or 3, wherein the detecting and classifying a text area of an image to be detected by using a preset deep neural network detection model to generate text area information with categories comprises:

detecting and classifying the text area of the image to be detected by using the preset
25 deep neural network detection model to obtain text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

5. The method according to any one of claims 2 to 4, wherein the performing template matching based on labeled template information and the text area information with the categories to generate structured information data comprises:

determining positions of the first anchor text area and the first non-anchor text area
30 on the image to be detected by matching the second anchor text area and the first anchor text area, and the second non-anchor text area and the first non-anchor text area; and

determining a mapping relationship between an entity and the first anchor text area

and the first non-anchor text area according to a mapping relationship between the entity and the second anchor text area and the second non-anchor text area, and removing, based on the positions of the first anchor text area and the first non-anchor text area, a redundant part comprising an non-text area and an irrelevant text in the text area of the image to be
5 detected to generate the structured information.

6. The method according to any one of claims 1 to 5, further comprising:
performing image preprocessing on the image to be detected,
wherein the image preprocessing comprises image correction and/or scaling to a
preset size.

10 7. The method according to any one of claims 1 to 6, further comprising:
training the preset deep neural network detection model.
8. The method according to claim 7, wherein the training the preset deep neural
network detection model comprises:

generating samples by using a sample generation tool; and
15 training the preset deep neural network detection model by using the samples.
9. The method according to claim 7 or 8, wherein the image to be detected is a new
sample, and
the method further comprises:
fine-tuning the preset deep neural network detection model by using the new sample.

20 10. An apparatus for detecting and analyzing a text, comprising:
a text area detection module, configured to detect and classify a text area of an image
to be detected by using a preset deep neural network detection model to generate text area
information with categories; and
a matching module, configured to perform template matching based on labeled
25 template information and the text area information with the categories to generate
structured information data.

11. The apparatus according to claim 10, further comprising:
a labeling module, configured to generate the labeled template information by
performing template labeling on a template image,
30 wherein the template image is the same as a master of the image to be detected, and
the text area of the image to be detected comprises a first anchor text area and a first
non-anchor text area, and a text area of the template image comprises a second anchor
text area and a second non-anchor text area.

12. The apparatus according to claim 11, wherein the labeling module is configured to:

generate the labeled template information by labeling sizes and relative positions of the second anchor text area and the second non-anchor text area on the template image, and a mapping relationship between an entity and the second anchor text area and the second non-anchor text area.

13. The apparatus according to claim 11 or 12, wherein the text area detection module is configured to:

detect and classify the text area of the image to be detected by using the preset deep neural network detection model to obtain text area information of the first anchor text area and the first non-anchor text area on the image to be detected.

14. The apparatus according to any one of claims 11 to 13, wherein the matching module is configured to:

determine positions of the first anchor text area and the first non-anchor text area on the image to be detected by matching the second anchor text area and the first anchor text area, and the second non-anchor text area and the first non-anchor text area; and

determine a mapping relationship between an entity and the first anchor text area and the first non-anchor text area according to a mapping relationship between the entity and the second anchor text area and the second non-anchor text area, and remove, based on the positions of the first anchor text area and the first non-anchor text area, a redundant part comprising an non-text area and an irrelevant text in the text area of the image to be detected to generate the structured information.

15. The apparatus according to any one of claims 10 to 14, further comprising:

an image preprocessing module, configured to perform image preprocessing on the image to be detected,

wherein the image preprocessing comprises image correction and/or scaling to a preset size.

16. The apparatus according to any one of claims 10 to 15, further comprising:

a model training module, configured to train the preset deep neural network detection model.

17. The apparatus according to claim 16, wherein the model training module is configured to:

generate samples by using a sample generation tool; and

train the preset deep neural network detection model by using the samples.

18. The apparatus according to claim 16 or 17, wherein the image to be detected is a new sample, and

the apparatus further comprises:

5 a fine-tuning module, configured to fine-tune the preset deep neural network detection model by using the new sample.

19. A device for detecting and analyzing a text, comprising:

a processor; and

a memory, configured to store executable instructions of the processor,

10 wherein the processor is configured to execute the method for detecting and analyzing a text according to any one of claims 1 to 9 by the executable instructions.

20. A computer readable storage medium on which a computer program is stored, wherein the computer program is configured to execute the method for detecting and analyzing a text according to any one of claims 1 to 9.

15