

IMAGE PROCESSING MODEL TRAINING METHOD, AND IMAGE PROCESSING METHOD AND APPARATUS

[0001] This application claims the priority to Chinese Patent Application
5 No.202311813675.3, titled "IMAGE PROCESSING MODEL TRAINING METHOD, AND
IMAGE PROCESSING METHOD AND APPARATUS", filed on December 27, 2023 with the
China National Intellectual Property Administration, which is incorporated herein by reference
in its entirety.

10 FIELD

[0002] The present disclosure relates to the field of image processing, and in particular, in a
possible implementation, to a method for training an image processing model, an image
processing method, an apparatus for training an image processing model, an image processing
apparatus, an electronic device and a storage medium.

15 BACKGROUND

[0003] With the development of technology, image processing technology is increasingly
applied in various fields. Since text in images carries more explicit information, text processing
in images has always been a focus of attention. In the related art, some artificial intelligence
20 models have been applied to text processing in images, offering advantages such as high
accuracy and fast speed.

[0004] However, due to variations in font, glyph, size, and the like of text, it is usually
required to train a dedicated artificial intelligence model for a specific application scenario.
During model training, a conventional method relies on manual annotation of a large amount
25 of collected image data, and then training is performed based on the annotation results. The
manual annotation process is time-consuming and labor-intensive, resulting in a long training
period.

SUMMARY

30 [0005] The present disclosure is provided in view of the above issues.

[0006] According to an aspect of the present disclosure, a method for training an image processing model is provided. The image processing model comprises a text detection model and a text determination model. The text detection model is for performing text recognition on an image. The text determination model is for detecting authenticity of text in an image. The method for training the image processing model includes:

step S110: acquiring a first image and a positive sample label corresponding to the first image, where the first image comprises a first text, and the positive sample label comprises authenticity information of the first text;

step S120: obtaining, by using a text generation tool, a second image and a negative sample label corresponding to the second image, where the second image comprises a second text generated by using the text generation tool, and the negative sample label comprises authenticity information of the second text and text recognition information; and

step S130: inputting both the first image and the second image into the text determination model to output a detection result by the text determination model, inputting the second image into the text detection model to output a text recognition result by the text detection model, calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result, and adjusting parameters of the image processing model by using the loss value to train the image processing model.

[0007] In a possible implementation, the loss value comprises a first loss value of the text determination model and a second loss value of the text detection model, and step S130 comprises:

first, inputting both the first image and the second image into the text determination model to output the detection result by the text determination model, calculating the first loss value of the text determination model based on the positive sample label, the negative sample label and the detection result, and adjusting parameters of the text determination model by using the first loss value to train the text determination model; and

then, inputting the second image into the text detection model to output the text recognition result by the text detection model, calculating the second loss value of the text detection model based on the negative sample label and the text recognition result, and adjusting parameters of

the text detection model by using the second loss value to train the text detection model.

[0008] In a possible implementation, the adjusting parameters of the image processing model by using the loss value to train the image processing model comprises:

simultaneously adjusting parameters of the text determination model and parameters of the text detection model by using the loss value to train the image processing model.

[0009] In a possible implementation, the calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result comprises:

calculating a first loss value of the text determination model based on the positive sample label, the negative sample label and the detection result;

calculating a second loss value of the text detection model based on the negative sample label and the text recognition result; and

calculating the loss value of the image processing model based on the first loss value and the second loss value.

[0010] In a possible implementation, the calculating the loss value of the image processing model based on the first loss value and the second loss value comprises:

performing a weighted summation on the first loss value and the second loss value to determine the calculated sum as the loss value of the image processing model.

[0011] In a possible implementation, the text detection model and the text determination model each comprises a network, and the networks share weights with each other.

[0012] In a possible implementation, the obtaining a second image and a negative sample label corresponding to the second image comprises:

acquiring a background image;

generating the second text and obtaining the negative sample label corresponding to the second text by using the text generation tool; and

mapping the second text onto the background image to generate the second image.

[0013] According to another aspect of the present disclosure, an image processing method is provided. The image processing method comprises:

acquiring a to-be-processed image, where the to-be-processed image comprises text; and

inputting the to-be-processed image into the image processing model trained by using the method described above to output a text recognition result of the to-be-processed image and/or an authenticity detection result of the text in the to-be-processed image.

[0014] According to another aspect of the present disclosure, an apparatus for training an image processing model is provided. The image processing model comprises a text detection model and a text determination model, the text detection model is for performing text recognition on an image, and the text determination model is for detecting authenticity of text in an image. The apparatus comprises a first acquisition module, a second acquisition module and a training module.

[0015] The first acquisition module is configured to acquire a first image and a positive sample label corresponding to the first image, where the first image comprises a first text, and the positive sample label comprises authenticity information of the first text.

[0016] The second acquisition module is configured to obtain, by using a text generation tool, a second image and a negative sample label corresponding to the second image, where the second image comprises a second text generated by using the text generation tool, and the negative sample label comprises authenticity information of the second text and text recognition information.

[0017] The training module is configured to input both the first image and the second image into the text determination model to output a detection result by the text determination model, input the second image into the text detection model to output a text recognition result by the text detection model, calculate a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result, and adjust parameters of the image processing model by using the loss value to train the image processing model.

[0018] According to another aspect of the present disclosure, an image processing apparatus is provided. The image processing apparatus comprises: a third acquisition module, configured to acquire a to-be-processed image, where the to-be-processed image comprises text; and a processing module, configured to input the to-be-processed image into the image processing model trained by the method described above to output a text recognition result of the to-be-

processed image and/or an authenticity detection result of the text in the to-be-processed image.

[0019] According to another aspect of the present disclosure, an electronic device is provided, comprising a processor and a memory. Computer program instructions are stored in the memory, and the computer program instructions, when run by the processor, are configured to execute the method for training the image processing model and/or the image processing method described above.

[0020] According to another aspect of the present disclosure, a storage medium is provided. Program instructions are stored on the storage medium, and the program instructions, when run, are configured to execute the method for training the image processing model and/or the image processing method described above.

[0021] In the process of training the image processing model described above, not only the second image including artificial characters is utilized, but also the first image including authentic characters is utilized. If model training is performed only by using the second image, the trained model exhibits an ability to recognize authentic images but with low recognition accuracy. In the above embodiments, the second image and the corresponding negative sample label are obtained by using a text generation tool, and the image processing model is trained based on the first image and the corresponding positive sample label as well as the second image and the corresponding negative sample label. Thus, the training of the image processing model is performed based on semantic labels generated by the text generation tool. This solution effectively reduces the manual effort and accelerates model training.

[0022] The above description is merely an overview of the technical solutions according to the present disclosure. To enable a clearer understanding of the technical solutions according to the present disclosure, implement the technical solutions based on the content of the specification, and make the above and other objectives, features, and advantages of the present disclosure more apparent and comprehensible, specific embodiments of the present disclosure are explicitly described below.

BRIEF DESCRIPTION OF THE DRAWINGS

[0023] The above and other purposes, features and advantages of the present disclosure will

become more apparent by describing the embodiments of the present disclosure in more detail with reference to the accompanying drawings. The drawings are used to provide a further understanding of the embodiments of the present disclosure and constitute a part of the specification. They are used together with the embodiments of the present disclosure to explain
5 the present disclosure and do not constitute a limitation on the present disclosure. In the drawings, the same reference numerals generally represent the same components or steps.

[0024] FIG.1 shows a schematic flowchart of a method for training an image processing model according to an embodiment of the present disclosure;

[0025] FIG.2 shows a second image according to an embodiment of the present disclosure;

10 **[0026]** FIG.3 shows a schematic diagram of performing image processing by using an image processing model according to an embodiment of the present disclosure;

[0027] FIG.4 shows a second image according to another embodiment of the present disclosure;

15 **[0028]** FIG.5 shows a schematic block diagram of an apparatus for training an image processing model according to an embodiment of the present disclosure;

[0029] FIG.6 shows a schematic block diagram of an image processing apparatus according to an embodiment of the present disclosure; and

[0030] FIG.7 shows a schematic block diagram of an electronic device according to an embodiment of the present disclosure.

20

DETAILED DESCRIPTION

[0031] In order to make the purposes, technical solutions and advantages of the present disclosure more apparent, exemplary embodiments according to the present disclosure will be described in detail below with reference to the accompanying drawings. Apparently, the
25 described embodiments are only a part of the embodiments of the present disclosure, not all of the embodiments. It should be understood that the present disclosure is not limited by the exemplary embodiments described herein. Based on the embodiments of the present disclosure described in the present disclosure, all other embodiments obtained by those skilled in the art without creative work shall fall within the protection scope of the present disclosure.

[0032] In order to at least partially address the above technical issues, a method for training an image processing model is provided according to an aspect of the present disclosure. The image processing model is configured for processing an image including text. The image processing model includes a text detection model for text recognition and a text determination model for text detection.

[0033] The text detection model is configured for performing text recognition on an image. In a possible implementation, the text detection model performs text localization on an image and recognize the localized text. The text detection model may include a Convolutional Recurrent Neural Network (CRNN) and a Connectionist Temporal Classification model.

[0034] The text determination model is configured for detecting the authenticity of text in an image. The text determination model may be used to detect whether an input image is a synthetic fake image. The text determination model may include network models such as Pointer Generation Network (PGNet), Differentiable Binarization Network (DBNet), Character Region Awareness for Text detection (CRAFT), or may further be object detection network models such as Fully Convolutional One-Stage object detection (FCOS), You Only Look Once (YOLO) for real-time object detection. The text determination model assists the text detection model in performing more accurate text recognition.

[0035] FIG.1 shows a schematic flowchart of a method for training an image processing model according to an embodiment of the present disclosure. As shown in FIG.1, the method includes step S110, step S120 and step S130.

[0036] In step S110, a first image and a positive sample label corresponding to the first image are acquired. The first image includes a first text. The positive sample label includes authenticity information of the first text.

[0037] The first image may be an originally captured image or an image obtained by processing an originally captured image, which includes a limited character set in the industrial field, for example, an image with a production date or an image with a product number, etc. The first image may be an RGB image or a grayscale image. The first image may be an original image directly captured by an image acquisition apparatus, or may be an image after preprocessing operations on the original image. The preprocessing operations may include all

operations for improving the visual effect of the image, enhancing the clarity of the image, or highlighting the text in the image. Exemplarily but not restrictively, the preprocessing operations may include operations such as digitization, geometric transformation, normalization, and filtering of the original image. The first image includes a first text. Since the first image is originally captured or obtained by processing an originally captured image, the first text therein includes authentic characters. The authentic characters may include numbers, letters, etc.

[0038] The first image has a positive sample label corresponding to the first image, and the positive sample label indicates the authenticity of the first text. As mentioned above, because the first image is an authentic image, the first text is authentic text.

[0039] In step S120, a second image and a negative sample label corresponding to the second image are obtained by using a text generation tool, where the second image includes a second text generated by using the text generation tool. The negative sample label includes authenticity information of the second text and text recognition information.

[0040] In the embodiments of the present disclosure, a small number of clean background images that do not include any text may be captured in an industrial scene. Then, various types of text content may be automatically generated at specified positions in the images by using a text generation tool. FIG.2 shows a second image according to an embodiment of the present disclosure. The text in the second image shown in FIG.2 is characters generated by a text generation tool. For convenience of description, the characters in the second text generated by using the text generation tool are hereinafter referred to as artificial characters. The text generation tool may include, but is not limited to, a TTF text generation tool. In the embodiments of the present disclosure, the type of text generation tool is not limited in any way. Any tool that can automatically generate text content falls within the protection scope of the present disclosure.

[0041] The artificial characters in the second image may correspond to the authentic characters in the first text of the first image. For example, the artificial characters in the second image have the same style, such as font, shape, or size, as the authentic characters in the first image. This provides a basis for the text determination model and the text detection model to

respectively perform text detection and text recognition.

[0042] The second image has a negative sample label corresponding to the second image. The negative sample label indicates the authenticity information of the second text and text recognition information. Since the second image is generated by using a text generation tool, the second text therein may be considered synthetic, i.e., the second image is also synthetic. The text recognition information may include text position information, text content information, and the like in the image. It can be understood that when the text generation tool generates artificial characters in the second image, the text generation tool generates the corresponding negative sample label simultaneously. Thus, the process of manually collecting a large number of images including text and annotating the images is avoided.

[0043] In step S130, both the first image and the second image are input into the text determination model to output a detection result by the text determination model, the second image is input into the text detection model to output a text recognition result by the text detection model, a loss value of the image processing model is calculated based on the positive sample label, the negative sample label, the text recognition result and the detection result, and parameters of the image processing model is adjusted by using the loss value to train the image processing model.

[0044] The loss value of a model may be used to measure the degree of difference between a prediction result of the model and a true label. A smaller loss value indicates that the prediction result of the model is closer to the true label, that is, a better performance of the model. When the loss value of the model is large, the parameters of the model are adjusted to improve the accuracy and generalization ability of the model.

[0045] The detection result output by the text determination model may indicate whether the input image is an authentic image or a synthetic image. The authentic image includes authentic characters, and the synthetic image includes artificial characters. The positive sample label corresponding to the first image may indicate that the first image is an authentic image, and the negative sample label corresponding to the second image may indicate that the second image is a synthetic image. Therefore, the loss value of the loss function of the text determination model is calculated based on the positive sample label, the negative sample label, and the detection

result output by the text determination model. The parameters of the text determination model may be adjusted based on the loss value until the detection result output by the text determination model is highly likely to be consistent with the label of the input image, that is, the text determination model can accurately detect the authenticity of the text in the image.

5 **[0046]** The text detection model outputs a text recognition result of the image, which may specifically include the position of the text in the image and the specific content of the text, etc. The negative sample label corresponding to the second image may indicate the position and the specific content of the artificial characters in the second image. The loss value of the loss function of the text detection model may be calculated based on the negative sample label and
10 the text recognition result output by the text detection model. The parameters of the text detection model may be adjusted based on the loss value until the text recognition result output by the text detection model is highly likely to be consistent with the negative sample label of the second image, that is, the text detection model can accurately recognize the text in the image.

[0047] The trained text determination model and the trained text detection model constitute
15 the trained image processing model.

[0048] In the process of training the image processing model described above, not only the second image including artificial characters is utilized, but also the first image including authentic characters is utilized. If model training is performed only by using the second image, the trained model exhibits an ability to recognize authentic images but with low recognition
20 accuracy. In the above embodiments, the second image and the corresponding negative sample label are obtained by using a text generation tool, and the image processing model is trained based on the first image and the corresponding positive sample label as well as the second image and the corresponding negative sample label. Thus, the training of the image processing model is performed based on semantic labels generated by the text generation tool. This solution
25 effectively reduces the manual effort and accelerates model training.

[0049] In some embodiments, the text detection model and the text determination model in the image processing model each include a network, and these networks share weights with each other.

[0050] Both the text detection model and the text determination model are neural network

models. A neural network is formed by a large number of neurons interconnected. After receiving an input, each neuron performs linear weighting processing on the input. These linear weighting processes may be represented by weights. When a neuron receives two inputs, each input is multiplied by an associated weight, and the products are then summed as the output result. These weights may be randomly initialized during the neural network training process and updated during the model training process.

[0051] The text detection model and the text determination model each include a network, and these networks share weights. In other words, the weights remain consistent between these networks. During the training process, if a weight of a neuron in one of the networks sharing weight changes, a weight of a corresponding neuron in the other network changes accordingly. For example, the networks sharing weight may be referred to as backbone networks and may be used to extract features of an image.

[0052] FIG.3 shows a schematic diagram of performing image processing by using an image processing model according to an embodiment of the present disclosure. In FIG.3, a first shared-weight network and a first sub-network constitute the text detection model. It can be understood that the first sub-network may be a downstream task network of the text detection model. A second shared-weight network and a second sub-network constitute the text determination model. Since the text determination model is configured for detecting the authenticity of text in an image, the second sub-network may be implemented using a discriminator. As shown in FIG.3, the second image and the first image are input into the first shared-weight network and the second shared-weight network, respectively. The second image after calculated through the first shared-weight network is input to both the first sub-network and the second sub-network. The first image after calculated through the second shared-weight network is only input to the second sub-network. From the perspective of the text determination model, it receives the first image and the second image and outputs the authenticity of the first image and the second image respectively. From the perspective of the text detection model, it only receives the second image and outputs the text recognition result of the second image.

[0053] In a possible implementation, the text recognition result may include position information of text in the image, position information of each character in the text, and text

content information. It can be understood that the position information of the text may be represented by a bounding rectangle of the text. For convenience of description, this rectangle is hereinafter referred to as a text box. In a possible implementation, the text box may be represented by the coordinates of opposite vertices of the text box, for example, the coordinates of the upper-left vertex and the lower-right vertex. Alternatively, the text box may further be represented by its center point and its length and width. The position information of a character may be represented by the coordinates of the center position of the character. The text content information indicates characters the text may include, such as numeric characters like 1, 2, 3, and English characters like a, b, c, etc. Based on the output results of the image processing model described above, the positive sample label and negative sample label described above, the image processing model is trained.

[0054] In the above technical solution, the image processing model includes not only a text detection model but also a text determination model, and the two models have shared-weight networks. The shared-weight networks extract features of both the first image and the second image. Thus, the accuracy of image processing by the image processing model is ensured. In addition, the shared-weight networks effectively reduce the number of parameters required by the image processing model, improve the generalization ability of the model and apply the knowledge learned from one training task to another related task, further improving the training efficiency of the image processing model.

[0055] In some embodiments, the obtaining a second image and a negative sample label corresponding to the second image in step S120 includes: acquiring a background image; generating the second text and obtaining the negative sample label corresponding to the second text by using the text generation tool; and mapping the second text onto the background image to generate the second image.

[0056] The background image may be an image without any printed text, and the second text may be at any position of the background image. By using the text generation tool, the generated artificial characters are randomly mapped onto the background image to form the second image, and the corresponding negative sample label may further be generated simultaneously. The second image may include strong interference signals and further include features similar to

images in authentic scenes. FIG.4 shows a second image according to another embodiment of the present disclosure. As shown in FIG.4, the text in the second image is generated by mapping artificial characters generated by a text generation tool onto a background image. The second image has strong randomness and further possesses features that authentic images have. The second image generated in this way effectively improves the robustness of the image processing model.

[0057] In the above embodiments, because the second image has strong randomness and is provided with label information, the operation of manually collecting a large number of images with text and annotating the images is avoided, effectively improving the speed of collecting second images and further improving the training efficiency of the image processing model.

[0058] In some embodiments, the loss value includes a first loss value of the text determination model and a second loss value of the text detection model. Step S130 includes the following steps. Both the first image and the second image are input into the text determination model to output the detection result by the text determination model. The first loss value of the text determination model is calculated based on the positive sample label, the negative sample label and the detection result. Parameters of the text determination model are adjusted by using the first loss value to train the text determination model. The second image is input into the text detection model to output the text recognition result by the text detection model. The second loss value of the text detection model is calculated based on the negative sample label and the text recognition result. Parameters of the text detection model are adjusted by using the second loss value to train the text detection model.

[0059] The first image and the second image are input into the text determination model to output respective detection results by the text determination model. The first loss value of the text determination model is calculated based on the detection result output by the text determination model, the positive sample label corresponding to the first image and the negative sample label corresponding to the second image. The parameters of the text determination model are updated based on the first loss value. The training of the text determination model is implemented through multiple repeated steps of calculation, updating, etc.

[0060] After the training of the text determination model is completed, only the second image

is input into the text detection model to output a text recognition result by the text detection model. Then, the second loss value of the text detection model is calculated based on the text recognition result output by the text detection model and the negative sample label, and the training of the text detection model is implemented by using the second loss value.

5 **[0061]** In the above embodiments, the text determination model is trained first, and then the text detection model is trained, ensuring the training effects of both. Particularly, in the embodiments where the text determination model and the text detection model have shared-weight networks, after the text determination model is trained, the text detection model may be trained by using some parameters of the text determination model, thereby accelerating the
10 training process of the image processing model.

[0062] In some embodiments, the adjusting parameters of the image processing model by using the loss value to train the image processing model in step S130 includes: simultaneously adjusting parameters of the text determination model and parameters of the text detection model by using the loss value to train the image processing model.

15 **[0063]** In the above embodiments, the text determination model and the text detection model may be trained simultaneously on different computing units. In other words, the image processing model is trained as a whole. Training the text determination model and the text detection model simultaneously effectively simplifies the training process of the image processing model, accelerates the training process of the image processing model, and improves
20 training efficiency.

[0064] In some embodiments, the calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result in step S130 includes the following steps S131 to S133. In step S131, a first loss value of the text determination model is calculated based on the positive sample label, the
25 negative sample label and the detection result. In step S132, a second loss value of the text detection model is calculated based on the negative sample label and the text recognition result. In step S133, the loss value of the image processing model is calculated based on the first loss value and the second loss value.

[0065] The first loss value represents the differences between the detection result and the

positive sample label, and between the detection result and the negative sample label, respectively, thereby representing the detection accuracy of the text determination model. In a possible implementation, a cross-entropy loss function may be used to calculate the above first loss value.

5 **[0066]** The second loss value represents the difference between the text recognition result and the negative sample label, thereby representing the text recognition accuracy of the text detection model. As mentioned above, the text recognition result may include a text box, which is used to represent the position of the text in the image. Alternatively, the text box may further be represented by its center point and its length and width. The second loss value may include
10 text box classification loss, text box regression loss, and/or text content loss, etc. The text box classification loss is mainly used to determine whether the text box contains text. The text box regression loss is used to optimize the position and size of the text box. An offset of each text box relative to the authentic text box is calculated. The text box regression loss may be calculated by using a regression loss function. The text content loss is used to determine the
15 text recognition capability of the text detection model. In a possible implementation, the second loss value of the text detection model may be obtained by performing a weighted summation on the text box classification loss, the text box regression loss and the text content loss. The loss function of the text detection model may be determined according to the model. In a possible implementation, if PGNet is used as the text detection model, the loss function in
20 PGNet may be used.

[0067] In the above embodiments, by calculating the first loss value and the second loss value and calculating the loss value of the image processing model based on the above two loss values, the loss value of the image processing model can be calculated precisely, effectively improving the stability and robustness of the image processing model.

25 **[0068]** In some embodiments, step S133, the calculating the loss value of the image processing model based on the first loss value and the second loss value, includes: performing a weighted summation on the first loss value and the second loss value to determine the calculated sum as the loss value of the image processing model.

[0069] The respective weights of the first loss value and the second loss value may be set by

comprehensively considering the relative importance of the first loss value and the second loss value in the image processing model. If one of the loss values has a greater impact on the performance of the image processing model, the loss value may be given a higher weight, thereby paying more attention to the loss value. This balances the relationship between different loss values, so that the image processing model has the desired performance.

[0070] In the above embodiments, a weighted summation is performed according to the weights of the first loss value and the second loss value, and the summation result is determined as the loss value of the image processing model to train the image processing model, improving the reliability and robustness of the image processing model.

[0071] According to another aspect of the present disclosure, an image processing method is further provided. The processing method includes the following steps. A to-be-processed image is acquired, where the to-be-processed image includes text. The to-be-processed image is input into the image processing model trained by using the training method described above to output a text recognition result of the to-be-processed image and/or an authenticity detection result of the text in the to-be-processed image.

[0072] It can be understood that the to-be-processed image may be an RGB image or a grayscale image. The to-be-processed image may be an original image directly captured by an image acquisition apparatus, or may be an image after preprocessing operations on the original image. The preprocessing operations may include all operations for improving the visual effect of the image, enhancing the clarity of the image, or highlighting the text in the image. Exemplarily but not restrictively, the preprocessing operations may include operations such as digitization, geometric transformation, normalization, and filtering of the original image.

[0073] In some embodiments of the present disclosure, according to the image processing method described above, the text recognition result of the to-be-processed image is determined.

In this embodiment, the presence of the text determination model in the image processing model facilitates training a more accurate text detection model. Thus, when the to-be-processed image is processed by using the image processing model, a more accurate text recognition result can be obtained.

[0074] In some embodiments of the present disclosure, according to the image processing

method described above, the authenticity of the text in the to-be-processed image is determined. In this embodiment, the presence of the text detection model in the image processing model facilitates training a more accurate text determination model. Thus, when the to-be-processed image is processed by using the image processing model, a more accurate text determination
5 result can be obtained.

[0075] In the above technical solution, the text recognition result of the to-be-processed image and/or the authenticity detection result of the text in the to-be-processed image is output by using the image processing model which is trained by the training method described above. This solution ensures the accuracy of the image processing results.

10 **[0076]** According to another aspect of the present disclosure, an apparatus for training an image processing model is further provided. The image processing model is for processing an image including text. The image processing model includes a text detection model and a text determination model. The text detection model is for performing text recognition on an image. The text determination model is for detecting the authenticity of text in an image. FIG.5 shows
15 a schematic block diagram of an apparatus for training an image processing model according to an embodiment of the present disclosure. As shown in FIG.5, the apparatus 500 includes a first acquisition module 510, a second acquisition module 520 and a training module 530.

[0077] The first acquisition module 510 is configured to acquire a first image and a positive sample label corresponding to the first image, where the first image includes a first text, and
20 the positive sample label includes authenticity information of the first text. The second acquisition module 520 is configured to obtain, by using a text generation tool, a second image and a negative sample label corresponding to the second image, where the second image includes a second text generated by using the text generation tool, and the negative sample label includes authenticity information of the second text and text recognition information. The
25 training module 530 is configured to input both the first image and the second image into the text determination model to output a detection result by the text determination model, input the second image into the text detection model to output a text recognition result by the text detection model, calculate a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result, and

adjust parameters of the image processing model by using the loss value to train the image processing model.

[0078] According to another aspect of the present disclosure, an image processing apparatus is further provided. FIG.6 shows a schematic block diagram of an image processing apparatus according to an embodiment of the present disclosure. As shown in FIG.6, the image processing apparatus 600 includes a third acquisition module 610 and a processing module 620.

[0079] The third acquisition module 610 is configured to acquire a to-be-processed image, where the to-be-processed image includes text. The processing module 620 is configured to input the to-be-processed image into the image processing model trained by the training method described above to output a text recognition result of the to-be-processed image and/or an authenticity detection result of the text in the to-be-processed image.

[0080] According to another aspect of the present disclosure, an electronic device 700 is further provided. FIG.7 shows a schematic block diagram of an electronic device 700 according to an embodiment of the present disclosure. As shown in FIG.7, the electronic device 700 includes a processor 710 and a memory 720. Computer program instructions are stored in the memory 720, and the computer program instructions, when run by the processor, are configured to execute the method for training the image processing model and/or the image processing method described above.

[0081] In addition, according to another aspect of the present disclosure, a storage medium is further provided. Program instructions are stored on the storage medium. The program instructions, when run by a computer or processor, causes the computer or processor to execute the corresponding steps of the method for training the image processing model and/or the image processing method of the embodiments of the present disclosure described above, and implements the corresponding modules of the apparatus for training the image processing model and/or the image processing apparatus of the embodiments of the present disclosure described above or the corresponding modules in the electronic device described above. The storage medium may include, for example, a storage component of a tablet computer, a hard disk of a personal computer, a read-only memory (ROM), an erasable programmable read-only memory (EPROM), a portable compact disc read-only memory (CD-ROM), a USB memory,

or any combination of the above storage media. The computer-readable storage medium may be any combination of one or more computer-readable storage media.

[0082] Those skilled in the art can understand the specific implementations and beneficial effects of the apparatus for training the image processing model, the image processing apparatus, the electronic device and the storage medium described above by reading the specific descriptions of the method for training the image processing model and/or the image processing method described above. For brevity, they are not repeated here.

[0083] Embodiment 1

A method for training an image processing model is provided, where the image processing model comprises a text detection model and a text determination model, the text detection model is for performing text recognition on an image, and the text determination model is for detecting authenticity of text in an image,

where the method comprises:

step S110: acquiring a first image and a positive sample label corresponding to the first image, where the first image comprises a first text, and the positive sample label comprises authenticity information of the first text;

step S120: obtaining, by using a text generation tool, a second image and a negative sample label corresponding to the second image, where the second image comprises a second text generated by using the text generation tool, and the negative sample label comprises authenticity information of the second text and text recognition information; and

step S130: inputting both the first image and the second image into the text determination model to output a detection result by the text determination model, inputting the second image into the text detection model to output a text recognition result by the text recognition model, calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result, and adjusting parameters of the image processing model by using the loss value to train the image processing model.

[0084] Embodiment 2

The method for training the image processing model described in embodiment 1 is

provided, where the loss value comprises a first loss value of the text determination model and a second loss value of the text detection model,

where step S130 comprises:

first, inputting both the first image and the second image into the text determination model
5 to output the detection result by the text determination model, calculating the first loss value of the text determination model based on the positive sample label, the negative sample label and the detection result, and adjusting parameters of the text determination model by using the first loss value to train the text determination model; and

then, inputting the second image into the text detection model to output the text recognition
10 result by the text detection model, calculating the second loss value of the text detection model based on the negative sample label and the text recognition result, and adjusting parameters of the text detection model by using the second loss value to train the text detection model.

[0085] Embodiment 3

The method for training the image processing model described in any one of embodiments
15 1 to 2 is provided, where the adjusting parameters of the image processing model by using the loss value to train the image processing model comprises:

simultaneously adjusting parameters of the text determination model and parameters of the text detection model by using the loss value to train the image processing model.

[0086] Embodiment 4

The method for training the image processing model described in any one of embodiments
20 1 to 3 is provided, where the calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result comprises:

calculating a first loss value of the text determination model based on the positive sample
25 label, the negative sample label and the detection result;

calculating a second loss value of the text detection model based on the negative sample label and the text recognition result; and

calculating the loss value of the image processing model based on the first loss value and the second loss value.

[0087] Embodiment 5

The method for training the image processing model described in any one of embodiments 1 to 4 is provided, where the calculating the loss value of the image processing model based on the first loss value and the second loss value comprises:

- 5 performing a weighted summation on the first loss value and the second loss value to determine the calculated sum as the loss value of the image processing model.

[0088] Embodiment 6

The method for training the image processing model described in any one of embodiments 1 to 5 is provided, where the text detection model and the text determination model each
10 comprises a network, and the networks share weights with each other.

[0089] Embodiment 7

The method for training the image processing model described in any one of embodiments 1 to 6 is provided, where the obtaining a second image and a negative sample label corresponding to the second image comprises:

- 15 acquiring a background image;
generating the second text and obtaining the negative sample label corresponding to the second text by using the text generation tool; and
mapping the second text onto the background image to generate the second image.

[0090] Embodiment 8

- 20 An image processing method is provided, and the image processing method comprises:
acquiring a to-be-processed image, where the to-be-processed image comprises text; and
inputting the to-be-processed image into the image processing model trained by using the method for training the image processing model described in any one of embodiments 1 to 7 to
output a text recognition result of the to-be-processed image and/or an authenticity detection
25 result of the text in the to-be-processed image.

[0091] Embodiment 9

An apparatus for training an image processing model is provided, where the image processing model comprises a text detection model and a text determination model, the text detection model is for performing text recognition on an image, and the text determination

model is for detecting authenticity of text in an image,

where the apparatus comprises:

a first acquisition module, configured to acquire a first image and a positive sample label corresponding to the first image, where the first image comprises a first text, and the positive
5 sample label comprises authenticity information of the first text;

a second acquisition module, configured to obtain, by using a text generation tool, a second image and a negative sample label corresponding to the second image, where the second image comprises a second text generated by using the text generation tool, and the negative sample label comprises authenticity information of the second text and text recognition information;

10 and

a training module, configured to input both the first image and the second image into the text determination model to output a detection result by the text determination model, input the second image into the text detection model to output a text recognition result by the text detection model, calculate a loss value of the image processing model based on the positive
15 sample label, the negative sample label, the text recognition result and the detection result, and adjust parameters of the image processing model by using the loss value to train the image processing model.

[0092] Embodiment 10

An image processing apparatus is provided, and the image processing apparatus comprises:

20 a third acquisition module, configured to acquire a to-be-processed image, where the to-be-processed image comprises text; and

a processing module, configured to input the to-be-processed image into the image processing model trained by the method for training the image processing model described in any one of embodiments 1 to 7 to output a text recognition result of the to-be-processed image
25 and/or an authenticity detection result of the text in the to-be-processed image.

[0093] Embodiment 11

An electronic device, comprising a processor and a memory, is provided. Computer program instructions are stored in the memory, and the computer program instructions, when run by the processor, are configured to execute the method for training the image processing

model described in any one of embodiments 1 to 7 and/or the image processing method described in embodiment 8.

[0094] Embodiment 12

A storage medium, storing program instructions, is provided. The program instructions, when run, are configured to execute the method for training the image processing model described in any one of embodiments 1 to 7 and/or the image processing method described in embodiment 8.

[0095] Although the embodiments have been described herein with reference to the drawings, it should be understood that the above embodiments are merely exemplary and are not intended to limit the scope of the present disclosure. Various changes and modifications may be made by those skilled in the art without departing from the scope and spirit of the present disclosure. All such changes and modifications are deemed to fall into the scope claimed by the claims of the present disclosure.

[0096] Those skilled in the art may realize that the units and algorithm steps of each example described in conjunction with the embodiments herein may be implemented by electronic hardware, or a combination of computer software and electronic hardware. Whether these functions are performed in hardware or software depends on the specific application and design constraints of the technical solution. Those skilled in the art may use different methods to implement the described functions for each specific application, but such implementation should not be regarded as exceeding the scope of the present disclosure.

[0097] It should be understood that the devices and methods disclosed in the embodiments of the present disclosure may be implemented in other manners. For example, the device embodiments described hereinabove are only schematic. For example, the division of the units is only based on logical functions, and there may be other division manners in practical implementation, such as multiple units or components may be combined or integrated into another device, or some features may be neglected or not executed.

[0098] In the specification provided herein, a lot of specific details are explained. However, it may be understood that the embodiments of the present disclosure may be practiced without these specific details. In some embodiments, well-known methods, structures and techniques

are not shown in detail in order to avoid obscuring understanding of the specification.

[0099] Similarly, it should be understood that in order to simplify the present disclosure and facilitate understanding one or more of the various aspects of the disclosure, in the description of the exemplary embodiments of the present disclosure, the various features of the present disclosure are sometimes grouped together into a single embodiment, figure, or description thereof. However, the method according to the present disclosure should not be interpreted as reflecting the following intention: the claimed disclosure requires more features than those explicitly stated in each claim. More specifically, as reflected in the corresponding claims, the inventive point is that the corresponding technical problem can be solved with features fewer than all the features of a single embodiment being disclosed. Therefore, the claims following the detailed description are hereby expressly incorporated into the detailed description, where each claim itself serves as an independent embodiment of the present disclosure.

[0100] It can be understood by those skilled in the art that, except for mutually exclusive features, all features disclosed in this specification (including the claims, abstract and drawings) and all processes or units of any method or device disclosed in the specification may be combined in any way. Unless otherwise expressly stated, each feature disclosed in this specification (including the claims, abstract and drawings) may be replaced by an alternative feature that provides the same, equivalent or similar purpose.

[0101] In addition, it can be understood by those skilled in the art that, although some embodiments described herein include certain features included in other embodiments instead of other features, the combination of features of different embodiments is meant to be within the scope of the present disclosure and form different embodiments. For example, in the claims, any one of the claimed embodiments may be used in any combination.

[0102] The various component embodiments of the present disclosure may be implemented in hardware, or in software modules running on one or more processors, or in a combination thereof. It can be understood by those skilled in the art that a microprocessor or a digital signal processor (DSP) may be adopted in practice to realize some or all of the functions of some of the modules of the apparatus for training the image processing model and the image processing apparatus according to the embodiments of the present disclosure. The present disclosure may alternatively be implemented as a device program (such as a computer program and computer

program product) for performing part or all of the methods described herein. Such a program implementing the present disclosure may be stored on a computer-readable medium, or may be in the form of one or more signals. Such a signal may be downloaded from an Internet website, or provided on a carrier signal, or provided in any other form.

5 **[0103]** It should be noted that the above embodiments illustrate rather than limit the present disclosure, and alternative embodiments may be designed by those skilled in the art without departing from the scope of the claims. In the claims, any reference symbol between parentheses shall not be construed as a limitation of the claims. The word "comprising/including" does not exclude the presence of elements or steps not listed in the claims. The word "a" or "an" preceding an element does not exclude the presence of multiple of such elements. The present disclosure may be implemented by means of hardware including several different elements and by means of a suitably programmed computer. In a unit claim enumerating several apparatuses, several of these apparatuses may be embodied by a same hardware item. The use of the words "first", "second", "third" and the like does not indicate any order. These words may be
10
15 interpreted as names.

[0104] Described above are only specific embodiments of the present disclosure or explanation for the specific embodiments of the present disclosure, and the protection scope of the present disclosure is not limited thereto. Any changes or alternations that can be easily envisaged by those familiar with the art within the technical scope disclosed in the present disclosure shall fall within the protection scope of the present disclosure. The protection scope of the present disclosure shall be in accordance with the protection scope of the claims.
20

CLAIMS

1. A method for training an image processing model, wherein the image processing model comprises a text detection model and a text determination model, the text detection model is
5 for performing text recognition on an image, and the text determination model is for detecting authenticity of text in an image,

wherein the method comprises:

step S110: acquiring a first image and a positive sample label corresponding to the first image, wherein the first image comprises a first text, and the positive sample label comprises
10 authenticity information of the first text;

step S120: obtaining, by using a text generation tool, a second image and a negative sample label corresponding to the second image, wherein the second image comprises a second text generated by using the text generation tool, and the negative sample label comprises authenticity information of the second text and text recognition information; and

15 step S130: inputting both the first image and the second image into the text determination model to output a detection result by the text determination model, inputting the second image into the text detection model to output a text recognition result by the text recognition model, calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result, and adjusting
20 parameters of the image processing model by using the loss value to train the image processing model.

2. The method for training the image processing model according to claim 1, wherein the loss value comprises a first loss value of the text determination model and a second loss value of the text detection model,

25 wherein step S130 comprises:

first, inputting both the first image and the second image into the text determination model to output the detection result by the text determination model, calculating the first loss value of the text determination model based on the positive sample label, the negative sample label and the detection result, and adjusting parameters of the text determination model by using the first
30 loss value to train the text determination model; and

then, inputting the second image into the text detection model to output the text recognition result by the text detection model, calculating the second loss value of the text detection model based on the negative sample label and the text recognition result, and adjusting parameters of the text detection model by using the second loss value to train the text detection model.

5 3. The method for training the image processing model according to claim 1, wherein the adjusting parameters of the image processing model by using the loss value to train the image processing model comprises:

simultaneously adjusting parameters of the text determination model and parameters of the text detection model by using the loss value to train the image processing model.

10 4. The method for training the image processing model according to claim 3, wherein the calculating a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result comprises:

calculating a first loss value of the text determination model based on the positive sample label, the negative sample label and the detection result;

15 calculating a second loss value of the text detection model based on the negative sample label and the text recognition result; and

calculating the loss value of the image processing model based on the first loss value and the second loss value.

20 5. The method for training the image processing model according to claim 4, wherein the calculating the loss value of the image processing model based on the first loss value and the second loss value comprises:

performing a weighted summation on the first loss value and the second loss value to determine the calculated sum as the loss value of the image processing model.

25 6. The method for training the image processing model according to any one of claims 1 to 5, wherein the text detection model and the text determination model each comprises a network, and the networks share weights with each other.

7. The method for training the image processing model according to any one of claims 1 to 5, wherein the obtaining a second image and a negative sample label corresponding to the second image comprises:

acquiring a background image;

generating the second text and obtaining the negative sample label corresponding to the second text by using the text generation tool; and

mapping the second text onto the background image to generate the second image.

5 8. An image processing method, comprising:

acquiring a to-be-processed image, wherein the to-be-processed image comprises text; and

inputting the to-be-processed image into the image processing model trained by using the method according to any one of claims 1 to 7 to output a text recognition result of the to-be-processed image and/or an authenticity detection result of the text in the to-be-processed image.

10 9. An apparatus for training an image processing model, wherein the image processing model comprises a text detection model and a text determination model, the text detection model is for performing text recognition on an image, and the text determination model is for detecting authenticity of text in an image,

wherein the apparatus comprises:

15 a first acquisition module, configured to acquire a first image and a positive sample label corresponding to the first image, wherein the first image comprises a first text, and the positive sample label comprises authenticity information of the first text;

20 a second acquisition module, configured to obtain, by using a text generation tool, a second image and a negative sample label corresponding to the second image, wherein the second image comprises a second text generated by using the text generation tool, and the negative sample label comprises authenticity information of the second text and text recognition information; and

25 a training module, configured to input both the first image and the second image into the text determination model to output a detection result by the text determination model, input the second image into the text detection model to output a text recognition result by the text detection model, calculate a loss value of the image processing model based on the positive sample label, the negative sample label, the text recognition result and the detection result, and adjust parameters of the image processing model by using the loss value to train the image processing model.

10. An image processing apparatus, comprising:

a third acquisition module, configured to acquire a to-be-processed image, wherein the to-be-processed image comprises text; and

5 a processing module, configured to input the to-be-processed image into the image processing model trained by the method according to any one of claims 1 to 7 to output a text recognition result of the to-be-processed image and/or an authenticity detection result of the text in the to-be-processed image.

10 11. An electronic device, comprising a processor and a memory, wherein computer program instructions are stored in the memory, and the computer program instructions, when run by the processor, are configured to execute the method for training the image processing model according to any one of claims 1 to 7 and/or the image processing method according to claim 8.

15 12. A storage medium, storing program instructions, wherein the program instructions, when run, are configured to execute the method for training the image processing model according to any one of claims 1 to 7 and/or the image processing method according to claim 8.