# DATA PROCESSING METHOD AND APPARATUS, AND DEVICE AND READABLE STORAGE MEDIUM

## RELATED APPLICATION

**[0001]** This application claims priority to Chinese Patent Application No. 202411748954.0, filed with the China National Intellectual Property Administration on November 29, 2024 and entitled "DATA PROCESSING METHOD AND APPARATUS, DEVICE, AND READABLE STORAGE MEDIUM", which is incorporated herein by reference in its entirety.

## FIELD OF THE TECHNOLOGY

**[0002]** This application relates to the field of computer technologies, and in particular, to a data processing method and apparatus, a device, and a readable storage medium.

## BACKGROUND OF THE DISCLOSURE

**[0003]** In the conventional field of visual question answering, images uploaded by users often include private content. When a user uploads an image including private content to a model for question answering, the model may cause leakage of the private content in the process of learning the image and answering image questions, resulting in low security.

## SUMMARY

**[0004]** Embodiments of this application provide a data processing method and apparatus, a device, and a readable storage medium, so as to appropriately enhance security of content protection information while ensuring accuracy of visual question answering.

**[0005]** An aspect of embodiments of this application provides a data processing method, including:

determining a question answering attention region from an image based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text;

obtaining content protection information, and determining a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information; and

determining a first to-be-processed region from the content protection region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping, and performing content protection processing on first image content in the first to-be-processed region in the image.

**[0006]** The determining a first to-be-processed region from the content protection region based on the question answering attention region includes:

determining a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determining the first to-be-processed region as a region in the content protection region other than the service overlap region.

**[0007]** The method further includes:

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content, and replacing the first image content with the first processed content in the image, to obtain a protected question-answering image including the first processed content, the protected question-answering image being configured for generating an answer text corresponding to the question text.

**[0008]** The method further includes:

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content;

determining a second to-be-processed region from the question answering attention region, determining second image content as image content in the second to-be-processed region, and performing content protection processing on the second image content to obtain second processed content; and

replacing, in the image, the first processed content with the first image content, and replacing the second processed content with the second image content, to obtain a protected question-answering image including the first processed content and the second processed content. The determining a second to-be-processed region from the question answering attention region includes:

generating T pieces of key question information based on the question text, T being a positive integer, and the key question information being key information configured for performing recognition to obtain the question text;

performing attention processing on the image content in the question answering attention region and the T pieces of key question information, to obtain a question attention result vector, and determining, from the question answering attention region based on the question attention result vector, a key question region associated with the T pieces of key question information; and

determining the second to-be-processed region as a region in the question answering attention region other than the key question region.

**[0009]** The determining a question answering attention region from an image based on the image and a question text corresponding to the image includes:

performing connected component labeling on each pixel point in the image to obtain M connected components, M being a positive integer, and each of the connected components being configured for representing an image region in which pixel values of pixel points in the image meet a connected component labeling condition and positions of the pixel points are adjacent to each other;

respectively performing feature extraction on the M connected components to obtain an image feature vector, and performing feature extraction on the question text corresponding to the image to obtain a question text feature vector; and

performing attention processing on the image feature vector and the question text feature vector, to obtain the question answering attention result vector, and determining the question answering attention region from the image based on the question answering attention result vector.

**[0010]** The determining the question answering attention region from the image based on the question answering attention result vector includes:

generating attention scores respectively corresponding to the M connected components based on the question answering attention result vector; and

determining a target connected component as a connected component having a maximum attention score, generating boundary contour points based on position information of pixel points in the target connected component, and determining the question answering attention region as a region defined by the boundary contour points in the image.

**[0011]** The performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content includes:

obtaining a protection level score associated with the content protection region, the

protection level score being generated based on a content protection result vector corresponding to the content protection information, the content protection result vector being an attention vector obtained by performing attention processing on the content protection information and the image;

determining a target re-drawing level based on the protection level score, and determining, from N content protection processing modes, a target content protection processing mode indicated by the target re-drawing level, N being a positive integer; and

performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content.

**[0012]** The target content protection processing mode is image concealment re-drawing. The performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content includes:

determining re-drawing source content as image content in the image other than the first to-be-processed region;

obtaining a background image in the re-drawing source content, and performing feature extraction on the background image, to obtain a background feature vector; and

performing re-drawing processing on the image content in the first to-be-processed region based on the background feature vector, to obtain the first processed content, a content feature vector of the first processed content having a similarity relation with the background feature vector.

**[0013]** The target content protection processing mode is image blurring re-drawing. The performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content includes:

obtaining a blurring unit value P, and dividing the image content in the first to-be-processed region into P to-be-blurred image units based on the blurring unit value P, P being a positive integer, and the P to-be-blurred image units including a target to-be-blurred image unit;

generating average color channel information of the target to-be-blurred image unit based on color channel information corresponding to each pixel point in the target to-be-blurred image unit, and performing blurring processing on the pixel point in the target to-be-blurred image unit based on the average color channel information, to obtain a blurred image unit; and

determining the first processed content as the P blurred image units when blurred image units respectively corresponding to the P to-be-blurred image units are obtained.

**[0014]** The target content protection processing mode is same-category replacement re-drawing, and the performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content includes:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information;

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, and determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information;

obtaining a content prompt text associated with the content protection information, determining, from the content prompt text, a target prompt text associated with the key object regions, and performing feature extraction on the target prompt text to obtain a text prompt feature vector;

performing corruption processing respectively on image content corresponding to the key object regions to obtain corrupted content, determining re-drawing source content as image content in the image other than the first to-be-processed region, and performing feature extraction on the re-drawing source content to obtain a re-drawing source feature vector; and

performing image inpainting on the corrupted content based on the text prompt feature vector and the re-drawing source feature vector to obtain the first processed content.

**[0015]** The target content protection processing mode is feature blurring re-drawing, and the performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content includes:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information; and

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector,

determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information, and performing image blurring re-drawing on image content corresponding to the key object regions, to obtain the first processed content.

**[0016]** An aspect of embodiments of this application provides a data processing apparatus, including:

a question answering region processing module, configured to determine a question answering attention region from an image based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text;

a protection region processing module, configured to obtain content protection information, and determine a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information; and

a re-drawing region processing module, configured to determine a first to-be-processed region from the content protection region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping, and perform content protection processing on first image content in the first to-be-processed region in the image.

**[0017]** In one embodiment, when determining a first to-be-processed region from the content protection region based on the question answering attention region, the re-drawing region processing module is specifically configured to perform the following operations:

determining a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determining the first to-be-processed region as a region in the content protection region other than the service overlap region.

**[0018]** In one embodiment, the re-drawing region processing module is further configured to perform the following operations:

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content, and replacing the first image content with the first processed content in the image, to obtain a protected question-answering image including

the first processed content, the protected question-answering image being configured for generating an answer text corresponding to the question text.

**[0019]** In one embodiment, the re-drawing region processing module is further configured to perform the following operations:

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content;

determining a second to-be-processed region from the question answering attention region, determining second image content as image content in the second to-be-processed region, and performing content protection processing on the second image content to obtain second processed content; and

determining a protected question-answering image as an image in which the first image content has been replaced with the first processed content and the second image content has been replaced with the second processed content.

**[0020]** In one embodiment, when determining a second to-be-processed region from the question answering attention region, the re-drawing region processing module is specifically configured to perform the following operations:

generating T pieces of key question information based on the question text, T being a positive integer, and the key question information being key information configured for performing recognition to obtain the question text;

performing attention processing on the image content in the question answering attention region and the T pieces of key question information, to obtain a question attention result vector, and determining, from the question answering attention region based on the question attention result vector, a key question region associated with the T pieces of key question information; and

determining the second to-be-processed region as a region in the question answering attention region other than the key question region.

**[0021]** In one embodiment, when determining the question answering attention region from the image based on the image and the question text corresponding to the image, the question answering region processing module is specifically configured to perform the following operations:

performing connected component labeling on each pixel point in the image to obtain M connected components, M being a positive integer, and each of the connected components being

configured for representing an image region in which pixel values of pixel points in the image meet a connected component labeling condition and positions of the pixel points are adjacent to each other;

respectively performing feature extraction on the M connected components to obtain an image feature vector, and performing feature extraction on the question text corresponding to the image to obtain a question text feature vector; and

performing attention processing on the image feature vector and the question text feature vector, to obtain the question answering attention result vector, and determining the question answering attention region from the image based on the question answering attention result vector.

**[0022]** In one embodiment, when determining the question answering attention region from the image based on the question answering attention result vector, the question answering region processing module is specifically configured to perform the following operations:

generating attention scores respectively corresponding to the M connected components based on the question answering attention result vector; and

determining a target connected component as a connected component having a maximum attention score, generating boundary contour points based on position information of pixel points in the target connected component, and determining the question answering attention region as a region defined by the boundary contour points in the image.

**[0023]** In one embodiment, when performing content protection processing on the image content in the first to-be-processed region to obtain the first processed content, the re-drawing region processing module is specifically configured to perform the following operations:

obtaining a protection level score associated with the content protection region, the protection level score being generated based on a content protection result vector corresponding to the content protection information, the content protection result vector being an attention vector obtained by performing attention processing on the content protection information and the image;

determining a target re-drawing level based on the protection level score, and determining, from N content protection processing modes, a target content protection processing mode indicated by the target re-drawing level, N being a positive integer; and

performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content.

**[0024]** In one embodiment, the target content protection processing mode is image concealment re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module is specifically configured to perform the following operations:

determining re-drawing source content as image content in the image other than the first to-be-processed region;

obtaining a background image in the re-drawing source content, and performing feature extraction on the background image, to obtain a background feature vector; and

performing re-drawing processing on the image content in the first to-be-processed region based on the background feature vector, to obtain the first processed content, a content feature vector of the first processed content having a similarity relation with the background feature vector.

**[0025]** In one embodiment, the target content protection processing mode is image blurring re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module is specifically configured to perform the following operations:

obtaining a blurring unit value P, and dividing the image content in the first to-be-processed region into P to-be-blurred image units based on the blurring unit value P, P being a positive integer, and the P to-be-blurred image units including a target to-be-blurred image unit;

generating average color channel information of the target to-be-blurred image unit based on color channel information corresponding to each pixel point in the target to-be-blurred image unit, and performing blurring processing on the pixel point in the target to-be-blurred image unit based on the average color channel information, to obtain a blurred image unit; and

determining the first processed content as the P blurred image units when blurred image units respectively corresponding to the P to-be-blurred image units are obtained.

**[0026]** In one embodiment, the target content protection processing mode is same-category replacement re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module is specifically configured to perform the following operations:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information;

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, and determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information;

obtaining a content prompt text associated with the content protection information, determining, from the content prompt text, a target prompt text associated with the key object regions, and performing feature extraction on the target prompt text to obtain a text prompt feature vector;

performing corruption processing respectively on image content corresponding to the key object regions to obtain corrupted content, determining re-drawing source content as image content in the image other than the first to-be-processed region, and performing feature extraction on the re-drawing source content to obtain a re-drawing source feature vector; and

performing image inpainting on the corrupted content based on the text prompt feature vector and the re-drawing source feature vector to obtain the first processed content.

**[0027]** In one embodiment, the target content protection processing mode is feature blurring re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module is specifically configured to perform the following operations:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information; and

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information, and performing image blurring re-drawing on image content corresponding to the key object regions, to obtain the first processed content.

**[0028]** An aspect of embodiments of this application provides a computer device, including

a processor, a memory, and a network interface,

the processor being connected to the memory and the network interface, the network interface being configured to provide a data communication function, the memory being configured to store a computer program, the computer program, when executed by the processor, causing the computer device to perform the method provided in embodiments of this application.

**[0029]** An aspect of the embodiments of this application provides a computer-readable storage medium, the computer-readable storage medium having a computer program stored therein, the computer program being adapted to be loaded and executed by a processor, to cause a computer device having the processor to perform the method provided in embodiments of this application.

**[0030]** An aspect of embodiments of this application provides a computer program product, including a computer program, the computer program being stored in a computer-readable storage medium. A processor of a computer device reads the computer program from the computer-readable storage medium, and the processor executes the computer program, to cause the computer device to perform the method provided in embodiments of this application.

**[0031]** In embodiments of this application, a question answering attention region and a content protection region are determined, so that image content required by a model to process question answering can be distinguished, and then a first to-be-processed region is determined from the content protection region based on the question answering attention region, and content protection processing is performed on first image content in the first to-be-processed region, thereby reducing the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model. Meanwhile, since the first to-be-processed region and the question answering attention region do not overlap, the model can still recognize, by using the question answering attention region, the image content associated with the question text. Accordingly, this application can appropriately improve security of the content protection information while ensuring accuracy of question answering.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0032]** To describe the technical solutions in embodiments of this application or in the related art more clearly, the accompanying drawings required for describing embodiments or the related art are to be briefly described below. Apparently, the accompanying drawings in the following description show only some embodiments of this application, and a person of ordinary skill in the art may still derive other accompanying drawings from the accompanying drawings

without creative efforts.

**[0033]** FIG. 1 is a schematic diagram of a network architecture according to an embodiment of this application;

**[0034]** FIG. 2 is a schematic diagram 1 of a data processing scenario according to an embodiment of this application;

**[0035]** FIG. 3 is a schematic flowchart 1 of a data processing method according to an embodiment of this application;

**[0036]** FIG. 4 is a schematic flowchart 2 of a data processing method according to an embodiment of this application;

**[0037]** FIG. 5 is a schematic diagram 2 of a data processing scenario according to an embodiment of this application;

**[0038]** FIG. 6 is a schematic diagram 3 of a data processing scenario according to an embodiment of this application;

**[0039]** FIG. 7 is a schematic structural diagram of a data processing apparatus according to an embodiment of this application; and

**[0040]** FIG. 8 is a schematic structural diagram of a computer device according to an embodiment of this application.

## DESCRIPTION OF EMBODIMENTS

**[0041]** The technical solutions in embodiments of this application are clearly and completely described below with reference to accompanying drawings in embodiments of this application. Apparently, the described embodiments are merely some rather than all of embodiments of this application. All other embodiments obtained by a person of ordinary skill in the art based on embodiments of this application without creative efforts fall within the protection scope of this application.

**[0042]** FIG. 1 is a schematic diagram of a network architecture according to an embodiment of this application. As shown in FIG. 1, the network architecture may include a terminal device 100 and a service server 200. A communication connection may exist between the terminal device 100 and the service server 200. The foregoing communication connection does not limit a connection manner, which may be a direct or indirect connection in a wired communication manner, may be a direct or indirect connection in a wireless communication manner, or may be in another manner. This is not limited herein in this application.

**[0043]** The terminal device may include a smart terminal having a data processing function, such as a smart phone, a tablet computer, a notebook computer, a desktop computer, a smart voice interaction device, a smart home appliance (for example, a smart television), a wearable device, an on-board terminal, or an aircraft. The on-board terminal may be a terminal device in a smart traffic scenario or an assisted driving scenario. The terminal device 100 shown in FIG. 1 may be installed with an application client having a data processing function. When running in each terminal device, each application client may perform data interaction with the foregoing service server 200 shown in FIG. 1.

**[0044]** The application client may specifically include an on-board client, a smart home client, an entertainment client (for example, a game client), a multimedia client (for example, a video client), a social client, an information client (for example, a news client), or the like. The application client in embodiments of this application may be integrated in a client (for example, the social client), or may be an independent client (for example, the news client). A type of the application client is not limited in embodiments of this application.

**[0045]** The service server 200 may be a server corresponding to the application client. The service server 200 may be an independent physical server, a server cluster composed of a plurality of physical servers, a distributed system, or a cloud server that provides basic cloud computing services such as a cloud service, a cloud database, cloud computing, a cloud function, cloud storage, a network service, cloud communication, a middleware service, a domain name service, a security service, a content delivery network (CDN), and a big data and artificial intelligence platform.

**[0046]** As shown in FIG. 1, an object (a user) may upload, in a terminal device 100, an image, a question text corresponding to the image, and content protection information by using an application client. An on-device multimodal model may be deployed in the terminal device 100. The on-device multimodal model may refer to an artificial intelligence model that runs on a terminal device and is capable of simultaneously processing data of a plurality of modalities (such as text, images, and audio). The content protection information may be description information for private image content in the image, and the content protection information may be text data, image data, or audio data.

**[0047]** The terminal device 100 may perform attention processing on the image and the question text by using the on-device multimodal model (to be specific, the terminal device may learn a correlation between the image and the question text), to determine, from the image, a question answering attention region associated with the question text. Similarly, the terminal device 100 may perform attention processing on the image and the content protection information by using

the on-device multimodal model (to be specific, the terminal device may learn a correlation between the image and the content protection information), to determine, from the image, a content protection region associated with the content protection information. Both the question answering attention region and the content protection region may be image masks defined by polygon vertices.

**[0048]** If an overlapping region exists between the question answering attention region and the content protection region, the terminal device 100 may determine a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determine a first to-be-processed region as a region in the content protection region other than the service overlap region. Content protection processing is performed on the first image content in the first to-be-processed region by using the on-device multimodal model, to obtain first processed content. The content protection processing includes various manners of eliminating the image indicated by the content protection information, such as content hiding processing and re-drawing processing. Content hiding processing may refer to adding an opaque mask layer to an image region, thereby hiding the image region. The image region may alternatively be directly erased (eliminated). Re-drawing processing may include image concealment re-drawing, image blurring re-drawing, same-category replacement re-drawing, feature blurring re-drawing, and the like. This is not limited herein in embodiments of this application.

**[0049]** A protected question-answering image is determined as an image in which the first image content has been replaced with the first processed content. The first image content is the image content in the first to-be-processed region of the image that has not been redrawn.

**[0050]** The terminal device 100 may transmit the protected question-answering image obtained through content protection processing and the question text to the service server 200. The service server recognizes the protected question-answering image by using a cloud-based multimodal model, to generate an answer text corresponding to the question text, and returns the answer text to the terminal device 100.

**[0051]** The on-device multimodal model in the terminal device 100 may also be deployed in the service server 200. The service server 200 may obtain an image, perform content protection processing on the image by using the on-device multimodal model, to obtain a protected question-answering image, and then recognizes the protected question-answering image by using the cloud-based multimodal model, to generate an answer text corresponding to the question text. The on-device multimodal model and the cloud-based multimodal model may ensure the security of the content protection information through model isolation. The model isolation may be implemented through physical isolation or trusted environment isolation (a trusted environment is

accessible only upon authorization). This is not limited herein in embodiments of this application.

**[0052]** In embodiments of this application, a question answering attention region and a content protection region are determined, so that image content required by a model to process visual question answering can be distinguished, and a first to-be-processed region is determined from the content protection region based on the question answering attention region, and content protection processing is performed on first image content in the first to-be-processed region through the on-device multimodal model deployed on the terminal device 100, rendering the image content indicated by the content protection information unrecognizable by the model. Furthermore, when the protected question-answering image and the question text are transmitted to the cloud-based multimodal model of the service server 200, this reduces the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model. Meanwhile, since the first to-be-processed region and the question answering attention region do not overlap, the model can still recognize, by using the question answering attention region, the image content associated with the question text. Accordingly, this application can appropriately improve security of the content protection information while ensuring accuracy of question answering.

**[0053]** FIG. 2 is a schematic diagram 1 of a data processing scenario according to an embodiment of this application. As shown in FIG. 2, an object (a user) may upload, in a terminal device, an image A and a question text corresponding to the image A by using an application client. The question text corresponding to the image A may be a text obtained by performing speech recognition on a voice input by the object, or a text directly input by the object on the application client. This is not limited herein in embodiments of this application. The terminal device may be the terminal device 100 in the foregoing embodiment corresponding to FIG. 1. For example, the question text may be "a type of a car in an image".

**[0054]** The terminal device may obtain content protection information. The content protection information may be description information for private image content in the image A. The content protection information may be a type description or a specific description directly specified, may be configured by selecting from private content types in the application client, may be input by the object, or may be obtained through automatic recognition by a model. The content protection information may be text data, image data, or audio data. Types of the content protection information may include personal identity information, contact information, financial information, vehicle information, medical information, address location information, and the like. The content protection information may be for example a "person/portrait".

**[0055]** An on-device multimodal model may be deployed in the terminal device. The on-device multimodal model may refer to an artificial intelligence model that runs on the terminal device and is capable of simultaneously processing data of a plurality of modalities (such as text, images, and audio), which may be for example a multimodal large language model such as a Phi-3-mini-128k-instruct model (an open-source lightweight model), a Chat-general language model (GLM) model (a natural language generation model), or a Qwen-vision language model (VLM) model (a large-scale visual language model) operable on an edge device.

**[0056]** The terminal device may determine, from the image A by using the on-device multimodal model, the image A, and the question text, a question answering attention region B associated with the question text. Similarly, the terminal device may determine, from the image A by using the on-device multimodal model, the image A, and the content protection information, a content protection region C associated with the content protection information. Both the question answering attention region B and the content protection region C may be image masks defined by polygon vertices.

**[0057]** As shown in FIG. 2, an overlapping region exists between the question answering attention region B and the content protection region C. The terminal device may determine a service overlap region D as the overlapping region between the question answering attention region and the content protection region. The service overlap region D may be a region represented by diagonal lines. The terminal device may determine a first to-be-processed region E as a region in the content protection region C other than the service overlap region D. The first to-be-processed region E may be a region represented by horizontal lines.

**[0058]** The terminal device may perform content protection processing on the first to-be-processed region E. The content protection processing includes various manners of eliminating the image indicated by the content protection information, such as content hiding processing and re-drawing processing. Content hiding processing may refer to adding an opaque mask layer to an image region, thereby hiding the image region. The image region may alternatively be directly erased (eliminated). Re-drawing processing may include image concealment re-drawing, image blurring re-drawing, same-category replacement re-drawing, feature blurring re-drawing, and the like. This is not limited herein in embodiments of this application.

**[0059]** An example in which the content protection processing is re-drawing processing is used. The terminal device may determine a protection level score based on a degree of association between image content in the content protection region C and the content protection information, and determine a content protection processing mode for the content protection processing based on

the protection level score.

**[0060]** An example in which content protection processing is image concealment re-drawing is used. The image concealment re-drawing may be generating first processed content F in the first to-be-processed region E based on a background image (for example, the image may be a tree) in an image A, and determining a protected question-answering image G as the image A in which the first image content has been replaced with the first processed content F. The first image content is image content in the first to-be-processed region E of the image, and the protected question-answering image G is configured for performing image recognition to generate an answer text corresponding to the question text. The protected question-answering image G may provide, by using the first processed content F, security protection for image content associated with the content protection information.

**[0061]** Embodiments of this application may be applied to a visual question answering (VQA) task and a visual re-drawing task. In the image re-drawing task, in this application, localized re-drawing may be performed on indicated image content by using the content protection information, to realize style conversion of partial image content, while retaining other original features in the image. This can achieve diversified processing of the image and can be applied to fields such as image editing and visual design. In the question answering task, in this application, content protection processing may be performed on the first image content in the first to-be-processed region, thereby reducing the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model, and effectively protecting the content protection information, while retaining image content associated with the question text in the image, and ensuring that the accuracy of question answering is not affected.

**[0062]** FIG. 3 is a schematic flowchart 1 of a data processing method according to an embodiment of this application. The data processing method may be performed by a computer device. The computer device may be the terminal device 100 or the service server 200 shown in FIG. 1. A description is provided below by using an example in which the data processing method is performed by a computer device. The data processing method may include at least the following operation S101 to operation S103.

**[0063]** Operation S101: Determine a question answering attention region from an image based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text.

**[0064]** Specifically, the computer device may obtain the image and the question text

corresponding to the image. The question text corresponding to the image may be a text obtained by performing speech recognition on a voice input by an object, or a text directly input by the object on the application client. This is not limited herein in embodiments of this application. The question text may be for example "a type of a vehicle in an image", "whether the image includes a vehicle", or "a time the image was captured".

**[0065]** An on-device multimodal model may be deployed in the computer device. The on-device multimodal model may refer to an artificial intelligence model that runs on the terminal device and is capable of simultaneously processing data of a plurality of modalities (such as text, images, and audio), which may be for example a multimodal large language model such as a Phi-3-mini-128k-instruct model, a Chat-GLM model, or a Qwen-VLM model operable on an edge device.

**[0066]** The computer device may divide an image into a plurality of image regions, for example, M image regions, by using the on-device multimodal model. The image regions may be connected components or shape domains. Attention processing is performed on the plurality of image regions and the question text to obtain a question answering attention result vector. The question answering attention result vector may be an M-dimensional vector. Each dimension in the question answering attention result vector may represent a first attention score between a first image region and the question text. The first attention score may represent a correlation degree between the image region and the question text. The computer device may determine, as the image region having a maximum first attention score, the question answering attention region associated with the question text. The question answering attention region may be an image mask defined by polygon vertices. In some embodiments, the computer device may separately determine question answering attention regions as a plurality of image regions in which first attention scores reach a first score threshold (a preset numerical value). To be specific, a plurality of question answering attention regions may be obtained.

**[0067]** Operation S102: Obtain content protection information, and determine a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information.

**[0068]** Specifically, the computer device may obtain the content protection information. The content protection information may be description information for private image content in the image. The content protection information may be a type description or a specific description directly specified, may be configured by selecting from private content types in the application

client, may be input by the object, or may be obtained through automatic recognition by a model. The content protection information may be text data, image data, or audio data. Types of the content protection information may include personal identity information, contact information, financial information, vehicle information, medical information, address location information, and the like.

**[0069]** The computer device may perform attention processing on the plurality of image regions and the content protection information by using the on-device multimodal model, to obtain a content protection result vector. The content protection result vector may also be an M-dimensional vector. Each dimension in the content protection result vector may represent a second attention score between an image region and the content protection information. The second attention score may represent a correlation degree between the image region and the content protection information. The computer device may determine, as the image region having a maximum second attention score, the content protection region associated with the content protection information. The content protection region may be an image mask defined by polygon vertices. In some embodiments, the computer device may separately determine content protection regions as a plurality of image regions in which second attention scores reach a second score threshold (a preset numerical value). To be specific, a plurality of content protection regions may be obtained.

**[0070]** Operation S103: Determine a first to-be-processed region from the content protection region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping, and perform content protection processing on first image content in the first to-be-processed region in the image.

**[0071]** In embodiments of this application, the first to-be-processed region is configured for indicating to perform content protection processing on first image content in the image, the first image content being image content in the first to-be-processed region in the image.

**[0072]** Specifically, when an overlapping region exists between the question answering attention region and the content protection region, the computer device may determine a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determine the first to-be-processed region as a region in the content protection region other than the service overlap region.

**[0073]** The computer device may perform content protection processing on the first to-be-processed region. The content protection processing includes various manners of eliminating the image indicated by the content protection information, such as content hiding processing and

re-drawing processing. Content hiding processing may refer to adding an opaque mask layer to an image region, thereby hiding the image region. The image region may alternatively be directly erased (eliminated). Re-drawing processing may include image concealment re-drawing, image blurring re-drawing, same-category replacement re-drawing, feature blurring re-drawing, and the like. This is not limited herein in embodiments of this application.

**[0074]** The computer device may determine a protection level score by using a second attention score corresponding to the content protection region. For example, the protection level score of the content protection region may be determined as the second attention score, and a content protection processing mode for content protection processing may be determined based on the protection level score.

**[0075]** The computer device may perform content protection processing on the first image content corresponding to the first to-be-processed region by using the on-device multimodal model, to obtain first processed content, and determine a protected question-answering image as an image in which the first image content has been replaced with the first processed content. The first image content is image content in the first to-be-processed region of the image, and the protected question-answering image may be configured to generate an answer text corresponding to the question text.

**[0076]** The computer device may determine one or more first to-be-processed regions from the image. Each first to-be-processed region may correspond to a different protection level score. To be specific, a re-drawing mode for each first to-be-processed region may differ from others.

**[0077]** In embodiments of this application, a question answering attention region and a content protection region are determined, so that image content required by a model to process question answering can be distinguished, and then a first to-be-processed region is determined from the content protection region based on the question answering attention region, and content protection processing is performed on first image content in the first to-be-processed region, thereby reducing the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model. Meanwhile, since the first to-be-processed region and the question answering attention region do not overlap, the model can still recognize, by using the question answering attention region, the image content associated with the question text. Accordingly, this application can appropriately improve security of the content protection information while ensuring accuracy of question answering.

**[0078]** FIG. 4 is a schematic flowchart 2 of a data processing method according to an

embodiment of this application. The data processing method may be performed by a computer device. The computer device may be the terminal device 100 or the service server 200 shown in FIG. 1. A description is provided below by using an example in which the data processing method is performed by a computer device. The data processing method may include at least the following operation S201 to operation S206.

**[0079]** Operation S201: Perform connected component labeling on each pixel point in the image to obtain M connected components, M being a positive integer, and each of the connected components being configured for representing an image region in which pixel values of pixel points in the image meet a connected component labeling condition and positions of the pixel points are adjacent to each other; respectively perform feature extraction on the M connected components to obtain an image feature vector, and perform feature extraction on the question text corresponding to the image to obtain a question text feature vector; and perform attention processing on the image feature vector and the question text feature vector, to obtain the question answering attention result vector.

**[0080]** Specifically, an on-device multimodal model may be deployed in the computer device. The on-device multimodal model may refer to an artificial intelligence model that runs on the terminal device and is capable of simultaneously processing data of a plurality of modalities (such as text, images, and audio), which may be for example a multimodal large language model such as a Phi-3-mini-128k-instruct model, a Chat-GLM model, or a Qwen-VLM model operable on an edge device. The on-device multimodal model may be an image-text multimodal model. The image-text multimodal model is an artificial intelligence model capable of simultaneously processing and understanding image and text data. By integrating visual information and language information, such models can perform tasks such as image description generation, question answering, image classification, and text matching. The core of the models lies in fusion and alignment of data of different modalities, enabling the models to extract useful features from a plurality of information sources and perform comprehensive analysis and reasoning.

**[0081]** The computer device may input an image into an image processing layer of the on-device multimodal model. The image processing layer of the on-device multimodal model may be an image encoder in contrastive language-image pre-training (CLIP) or a vision transformer (ViT) image encoder. This is not limited herein in embodiments of this application. The computer device may perform binarization processing on the image by using the image processing layer, set a grayscale value of a pixel point in the image greater than or equal to a service grayscale threshold to 255, and set a grayscale value of a pixel point in the image less than the service grayscale threshold

to 0, to obtain a grayscale image of the image. The service grayscale threshold may be set according to a service requirement or a specific image. This is not limited herein in embodiments of this application.

**[0082]** The computer device may perform connected component labeling on the grayscale image of the image, and label, as one connected component, pixel points that meet the connected component labeling condition and in which positions of the pixel point are adjacent to each other, to obtain M connected components. The connected component labeling condition may mean that pixel points have the same grayscale value. The connected component is configured for representing an image region in which pixel values of pixel points in the image meet a connected component labeling condition and positions of the pixel points are adjacent to each other. The computer device may separately perform feature extraction on image content in the M connected components, to obtain an image feature vector.

**[0083]** The computer device may input the question text into a text processing layer of the on-device multimodal model. The text processing layer of the on-device multimodal model may be a bidirectional encoder representations from transformers (BERT) text encoder. The computer device may split and encode the question text into tokens by using a tokenizer in the text processing layer. For example, when the question text is "a type of a car in an image", a token sequence corresponding to the question text may be [image, car, type], [image], [car], and [type] being all tokens. A method for text splitting may include word-based splitting, character-based splitting, or subword-based splitting. This is not limited in the embodiments of this application. The computer device may embed all tokens in the question text into vectors of the same size, and then perform vector concatenation on these vectors of the same size to obtain a question text feature vector.

**[0084]** The computer device may input the image feature vector and the question text feature vector into the attention processing layer of the on-device multimodal model, perform a dot product operation on the question text feature vector and a query parameter matrix $W_i^Q$ to obtain a query vector Q, perform a dot product operation on the image feature vector and a key parameter matrix $W_i^K$ to obtain a key vector K, and perform a dot product operation on the image feature vector and a value parameter matrix $W_i^V$ to obtain a value vector V. The query parameter matrix $W_i^Q$, the key parameter matrix $W_i^K$, and the value parameter matrix $W_i^V$ are all matrices formed by learnable parameters for attention processing.

**[0085]** The computer device may perform a dot product operation on a query vector $Q$ and a transpose operation result $K^T$ of the key vector K, to obtain an attention score vector.

Dimensionality reduction processing is performed on the attention score vector based on a square root $\sqrt{d_K}$ of a dimension value $d_K$ of the key vector K, normalization processing is performed on the attention score vector after dimensionality reduction to obtain an attention weight vector, and a dot product operation is performed on the attention weight vector and the value vector to obtain a question answering attention result vector $Attn_Q$. The process thereof may be as shown in Formula (1):

$$Attn_Q = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \quad \text{Formula (1)}$$

**[0086]** The softmax function is configured for normalization processing, and $d_K$ represents the dimension value of the key vector K.

**[0087]** Operation S202: Generate attention scores respectively corresponding to the M connected components based on the question answering attention result vector; and determine a target connected component as a connected component having a maximum attention score, generate boundary contour points based on position information of pixel points in the target connected component, and determine a question answering attention region as a region defined by the boundary contour points in the image.

**[0088]** Specifically, the question answering attention result vector $Attn_Q$ may be an M-dimensional vector, and each dimension in the question answering attention result vector $Attn_Q$ may represent a first attention score between a connected component and the question text. The first attention score may represent a correlation degree between the connected component and the question text. The computer device may determine a target connected component as the connected component having a maximum first attention score, and the computer device may determine S boundary contour points based on position information of pixel points in the target connected component. A number of boundary contour points may be determined based on a range and a shape of the target connected component. The computer device may determine the question answering attention region as a polygonal region formed by the boundary contour points. The question answering attention region may be an image mask defined by polygon vertices. For example, the question answering attention region may represent polygon vertex coordinates in text format, for example, {[100, 50], [50, 200], [150, 200]}, [100, 50], [50, 200], and [150, 200] being all coordinates of boundary contour points in an image. In some embodiments, the computer device may separately determine question answering attention regions as a plurality of image regions in which first attention scores reach a first score threshold (a preset numerical value). To be specific, a plurality of question answering attention regions may be obtained.

**[0089]** Operation S203: Obtain content protection information, and determine a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information.

**[0090]** Specifically, the computer device may obtain the content protection information. The content protection information may be description information for private image content in the image. The content protection information may be a preset type description or a specific description directly specified, may be configured by selecting from private content types in the application client, may be input by the object, or may be obtained through automatic recognition by a model. The content protection information may be text data, image data, or audio data. Types of the content protection information may include personal identity information, contact information, financial information, vehicle information, medical information, address location information, and the like.

**[0091]** The computer device may perform image recognition on the image data or perform speech recognition on the audio data, to obtain text data of the content protection information. An example in which the content protection information is text data is used. The computer device may perform feature extraction on the content protection information by using the text processing layer of the on-device multimodal model, to obtain a content protection feature vector corresponding to the content protection information. For the process, refer to the specific description of the text processing layer in the foregoing operation S201. Details are not described herein again.

**[0092]** The computer device may input the image feature vector and the content protection feature vector into the attention processing layer of the on-device multimodal model, and perform cross-attention calculation on the image feature vector and the content protection feature vector by using the attention processing layer, to obtain a content protection result vector. For the process of cross-attention calculation, refer to the specific description of the attention processing layer in the foregoing operation S201. Details are not described herein again.

**[0093]** The content protection result vector may also be an M-dimensional vector. Each dimension in the content protection result vector may represent a second attention score between an image region and the content protection information. The second attention score may represent a correlation degree between the image region and the content protection information. The computer device may determine, from the M connected components by using the second attention score in the content protection result vector, a service connected component associated with the question text, and determine the content protection region in the image based on the service connected component. The image content in the content protection region is image content in the image

associated with the content protection information. For the process of determining the service connected component and the content protection region, refer to the specific descriptions about the target connected component and the question answering attention region in operation S202. Details are not described herein again. In some embodiments, the computer device may separately determine content protection regions as a plurality of image regions in which second attention scores reach a second score threshold (a preset numerical value). To be specific, a plurality of content protection regions may be obtained.

**[0094]** Operation S204: Determine a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determine the first to-be-processed region as a region in the content protection region other than the service overlap region.

**[0095]** Specifically, the computer device may determine, by using the boundary contour points of the question answering attention region and boundary contour points of the content protection region, whether the question answering attention region and the content protection region overlap. If an overlapping region exists between the question answering attention region and the content protection region, the computer device may determine overlapping contour points based on the boundary contour points of the question answering attention region and the boundary contour points of the content protection region, and determine the overlapping region as a polygonal region formed by the overlapping contour points. The computer device may determine a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determine the first to-be-processed region as a region in the content protection region other than the service overlap region.

**[0096]** Operation S205: Obtain a protection level score associated with the content protection region, the protection level score being generated based on a content protection result vector corresponding to the content protection information, the content protection result vector being an attention vector obtained by performing attention processing on the content protection information and the image; determine a target re-drawing level based on the protection level score, and determine, from N content protection processing modes, a target content protection processing mode indicated by the target re-drawing level, N being a positive integer; and perform content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content.

**[0097]** Specifically, the computer device may perform content protection processing on the first to-be-processed region. The content protection processing includes various manners of

eliminating the image indicated by the content protection information, such as content hiding processing and re-drawing processing. Content hiding processing may refer to adding an opaque mask layer to an image region, thereby hiding the image region. The image region may alternatively be directly erased (eliminated). re-drawing processing may include image concealment re-drawing, image blurring re-drawing, same-category replacement re-drawing, feature blurring re-drawing, and the like. This is not limited herein in embodiments of this application.

**[0098]** The computer device may perform normalization processing on the M attention scores in the content protection result vector, to obtain a normalized score corresponding to each attention score. The computer device may determine, as the normalized score corresponding to the attention score corresponding to the service connected component in the content protection result vector, the protection level score associated with the content protection region. The protection level score may be a score ranging from 0 to 1, for example, may be 0.7. The computer device may perform recording by setting the protection level score as a grayscale value of the image mask of the content protection region. After the computer device obtains the protection level score, the protection level score may be stored, as a grayscale value, in the content protection region serving as a mask. When determining the content protection processing mode below, the computer device may determine the protection level score by extracting the grayscale value in the mask.

**[0099]** The computer device may determine a target re-drawing level of the content protection region based on the protection level score, and determine, from N content protection processing modes, a target content protection processing mode indicated by the target re-drawing level. An example in which content protection processing is re-drawing processing is used. A mapping relationship between the protection level score and the content protection processing mode is shown in Table 1:

Table 1

| Re-drawing level | Protection degree score | Content protection processing mode |
|---|---|---|
| T0 | [0, 0.2) | Maintain the original image |
| T1 | [0.2, 0.4) | Feature blurring re-drawing |
| T2 | [0.4, 0.6) | Homogeneous replacement re-drawing |
| T3 | [0.6, 0.8) | Image blurring re-drawing |
| T4 | [0.8, 1] | Image concealment re-drawing |

**[00100]** For example, when the protection level score is 0.1, the computer device may determine that the re-drawing level is T0, perform no re-drawing on the image content in the first to-be-processed region, and retain the original image. If the protection level score is 0.3, the

computer device may determine that the re-drawing level is T1, and perform feature blurring re-drawing on the image content in the first to-be-processed region. If the protection level score is 0.5, the computer device may determine that the re-drawing level is T2, and perform same-category replacement re-drawing on the image content in the first to-be-processed region. If the protection level score is 0.7, the computer device may determine that the re-drawing level is T3, and perform image blurring re-drawing on the image content in the first to-be-processed region. If the protection level score is 0.9, the computer device may determine that the re-drawing level is T4, and perform image concealment re-drawing on the image content in the first to-be-processed region. The first image content is the image content in the first to-be-processed region of the image.

**[00101]** FIG. 5 is a schematic diagram 2 of a data processing scenario according to an embodiment of this application. As shown in FIG. 5 An image A0 in which image content in a first to-be-processed region is first image content P0 is used as an example. A computer device may perform content protection processing on the first image content P0 based on a target content protection processing mode.

**[00102]** When the target content protection processing mode is image concealment re-drawing, the process of performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode to obtain first processed content may include: determining re-drawing source content as image content in the image other than the first to-be-processed region; obtaining a background image in the re-drawing source content, and performing feature extraction on the background image, to obtain a background feature vector; and performing re-drawing processing on the image content in the first to-be-processed region based on the background feature vector, to obtain the first processed content, a content feature vector of the first processed content having a similarity relation with the background feature vector.

**[00103]** Specifically, the computer device may determine the re-drawing source content as image content in the image A0 other than the first to-be-processed region, and the computer device may perform background recognition on the re-drawing source content, to obtain a background image in the re-drawing source content. The computer device may input the background image into a denoising diffusion probabilistic model (DDPM) or a latent stable diffusion model (LDM), and determine, by using ControlNet (a neural network configured to add condition control to an image diffusion model) and IPAdapter (Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models, an adapter configured to implement image prompt), a conditional input of a noise map in a backward diffusion process as the background feature vector corresponding to the

background image, thereby generating first processed content P4 that has a similarity relation with the background image. The content feature vector of the first processed content P4 has a similarity relation with the background feature vector. The foregoing process may also be referred to as performing outpainting (image outpainting) on the image content in the first to-be-processed region through the image diffusion model. The computer device may determine a protected question-answering image A4 as an image in which the first image content P0 has been replaced with the first processed content P4.

**[00104]** In some embodiments, the image concealment re-drawing may be directly setting pixel values of pixel points of the first image content P0 to 0 or 255, causing the first image content P0 to be invisible, and directly achieving image concealment. The image content of the first image content P0 may also be replaced in a manner of background replacement. Background replacement is a technique for hiding or removing sensitive information by replacing a background in an image. This technique can effectively protect privacy while retaining main objects or foreground information in the image.

**[00105]** When the target content protection processing mode is image blurring re-drawing, the process of performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode to obtain the first processed content may include: obtaining a blurring unit value P, and dividing the image content in the first to-be-processed region into P to-be-blurred image units based on the blurring unit value P, P being a positive integer, and the P to-be-blurred image units including a target to-be-blurred image unit; generating average color channel information of the target to-be-blurred image unit based on color channel information corresponding to each pixel point in the target to-be-blurred image unit, and performing blurring processing on the pixel point in the target to-be-blurred image unit based on the average color channel information, to obtain a blurred image unit; and determining the first processed content as the P blurred image units when blurred image units respectively corresponding to the P to-be-blurred image units are obtained.

**[00106]** Specifically, the computer device may obtain the blurring unit value P. The blurring unit value P may be determined based on a range size of the first image content P0. A wider range of the first image content P0 indicates a larger blurring unit value P. The computer device may divide the first image content P0 into P to-be-blurred image units based on the blurring unit value P, and the P to-be-blurred image units include a target to-be-blurred image unit.

**[00107]** The modes of image blurring re-drawing may include Gaussian blur, mean blur, bilateral filtering, and the like. Image blurring desensitization technology is a widely used image

processing technology, mainly configured for protecting privacy and sensitive information. Through blurring processing, some portions of an image may be rendered indistinct, thereby preventing unauthorized personnel from obtaining detailed information in the image.

**[00108]** Mean blur is used as an example. The computer device may generate average color channel information of the target to-be-blurred image unit based on color channel information corresponding to each pixel point in the target to-be-blurred image unit, and perform blurring processing on the pixel point in the target to-be-blurred image unit based on the average color channel information, which may be for example applying the average color channel information to each pixel point in the target to-be-blurred image unit, to obtain a blurred image unit. When obtaining the blurred image units respectively corresponding to the P to-be-blurred image units, the computer device may determine first processed content P3 as the P blurred image units. The computer device may determine a protected question-answering image A3 as an image in which the first image content P0 has been replaced with the first processed content P3.

**[00109]** In some embodiments, the mode of image blurring re-drawing may also be image cartoonization or color filling. The image cartoonization is an image processing technique that converts an image into a cartoon style. Through this technology, details and textures in an image may be simplified, causing the image to resemble a hand-drawn cartoon image. This technology may be applied not only to artistic creation but also to privacy protection and data desensitization, because cartoonization processing may blur or simplify sensitive information in the image and cause the image to be unrecognizable. Color filling is a technique for hiding sensitive information by filling some regions in an image with random colors or noise. This method not only alters colors but also introduces randomness, rendering the original information more difficult to recover and thus effectively protecting privacy.

**[00110]** When the target content protection processing mode is same-category replacement re-drawing, the process of performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode to obtain first processed content may include: generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information; performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, and determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information; obtaining a content prompt text associated

with the content protection information, determining, from the content prompt text, a target prompt text associated with the key object regions, and performing feature extraction on the target prompt text to obtain a text prompt feature vector; performing corruption processing respectively on image content corresponding to the key object regions to obtain corrupted content, determining re-drawing source content as image content in the image other than the first to-be-processed region, and performing feature extraction on the re-drawing source content to obtain a re-drawing source feature vector; and performing image inpainting on the corrupted content based on the text prompt feature vector and the re-drawing source feature vector to obtain the first processed content.

**[00111]** Specifically, the computer device may generate K pieces of key object information based on the content protection information, and the key object information may be key information configured for performing recognition to obtain key information of the content protection information. For example, when the content protection information is a "female character", the key object information may be "eyes (iris)", "nose", and "lip (lip wrinkles)". The computer device may perform attention processing on the first image content P0 and the K pieces of key object information, to obtain an object attention result vector, and determine, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information. For example, an image region of the key object information in the first image content P0 may be obtained, and image content in the key object regions may be configured for representing the key object information. For example, the key object regions may be image regions representing eyes of a person, a nose of a person, and a mouth of a person in the first image content P0.

**[00112]** The computer device may obtain a plurality of content prompt texts (prompts) associated with the key object regions. For example, when the key object region is an image region corresponding to the mouth of the person (where the key object information is the mouth), the computer device may obtain a prompt associated with the mouth of the person, which may include, for example, "a smiling mouth" or "an open mouth". The computer device may perform attention processing on several content prompt texts and the key object information represented by the key object region, to obtain a key object result vector, and determine, from the several content prompt texts based on the key object result vector, the target prompt text associated with the key object region. For example, the target prompt text may be "a smiling mouth", and the target prompt text may also be identical with a text of the content protection information. In some embodiments, the computer device may also perform image recognition by using the image content in the content protection region, to generate a target prompt text.

**[00113]** The computer device may perform corruption processing on the image content corresponding to the key object region (which may be for example the image region representing the mouth of a person), to obtain corrupted content. The corruption processing may be adding, to the image content corresponding to the key object region, noise conforming to a Gaussian distribution.

**[00114]** The computer device may input the target prompt text and the corrupted content into the DDPM or the LDM. Through ControlNet and IPAdapter, the conditional input for the corrupted content during reverse diffusion is determined as the text prompt feature vector corresponding to the content prompt text, thereby generating first processed content P2 that has a similarity relation with the target prompt text. For example, the computer device may redraw the mouth of the person in the first image content P0 as a smiling mouth in the first processed content P2, and perform backward diffusion on the first image content P0 based on target prompt texts (for example, eyes, a nose, and a girl) corresponding to a plurality of key object regions, to generate the first processed content P2. The foregoing process may also be referred to as performing inpainting (image inpainting) on the first image content P0 by using the image diffusion model. The computer device may determine a protected question-answering image A2 as an image in which the first image content P0 has been replaced with the first processed content P2.

**[00115]** When the target content protection processing mode is feature blurring re-drawing, the process of performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode to obtain first processed content may include: generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information; and performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, determining, from the first to-be-processed region based on the object attention result vector, K key object regions associated with the K pieces of key object information, performing image blurring re-drawing on image content corresponding to the key object regions, to obtain K pieces of redrawn image content, and determining the first processed content as the first image content including the K pieces of redrawn image content.

**[00116]** Specifically, the computer device may generate K pieces of key object information based on the content protection information, and the key object information may be key information configured for performing recognition to obtain key information of the content protection

information. For example, when the content protection information is a "person/human face", the key object information may be "eyes (iris)", "nose", and "lip (lip wrinkles)". The computer device may perform attention processing on the first image content P0 and the K pieces of key object information, to obtain an object attention result vector, and determine, from the first to-be-processed region based on the object attention result vector, K key object regions associated with the K pieces of key object information. For example, an image region of each piece of key object information in the first image content P0 may be obtained, and image content in each of the key object regions may be configured for representing the key object information.

**[00117]** The computer device may perform image blurring re-drawing on image content corresponding to the K key object regions, to obtain K pieces of redrawn image content, and determine first processed content P1 as the first image content P0 including the K pieces of redrawn image content. The mode of image blurring re-drawing may include Gaussian blur, mean blur, bilateral filtering, and the like. This is not limited herein in embodiments of this application. The computer device may determine a protected question-answering image A1 as an image in which the first image content P0 has been replaced with the first processed content P1.

**[00118]** The computer device may determine one or more first to-be-processed regions from the image. Each first to-be-processed region may correspond to a different protection level score. To be specific, a re-drawing mode for each first to-be-processed region may differ from others.

**[00119]** Operation S206: Determine a second to-be-processed region from the question answering attention region, determine second image content as image content in the second to-be-processed region, and perform content protection processing on the second image content to obtain second processed content; and replace, in the image, the first processed content with the first image content, and replace the second processed content with the second image content, to obtain a protected question-answering image including the first processed content and the second processed content.

**[00120]** Specifically, because an overlapping region exists between the question answering attention region and the content protection region, and the question answering attention region also retains partial image content indicated by the content protection information, to further improve the security of the image content indicated by the content protection information in the question answering attention region, the computer device may also perform content protection processing on the image content in the question answering attention region.

**[00121]** The computer device may obtain the first attention score corresponding to the question answering attention region. If the first attention score is greater than a content protection

processing threshold (a preset value, which may be for example 0.2), the computer device may determine a second to-be-processed region from the question answering attention region. If the first attention score is less than or equal to the content protection processing threshold, the computer device may directly determine the second to-be-processed region as the question answering attention region. The computer device may avoid, by determining the second to-be-processed region, performing content protection processing on key image content configured for recognizing the question text in the image. To be specific, the computer device can still ensure accuracy of question answering while further improving the security of the content protection information.

**[00122]** The process of determining the second to-be-processed region from the question answering attention region by the computer device may include: generating T pieces of key question information based on the question text, T being a positive integer, and the key question information being key information configured for performing recognition to obtain the question text; performing attention processing on the image content in the question answering attention region and the T pieces of key question information, to obtain a question attention result vector, and determining, from the question answering attention region based on the question attention result vector, a key question region associated with the T pieces of key question information; and determining the second to-be-processed region as a region in the question answering attention region other than the key question region.

**[00123]** Specifically, the computer device may generate T pieces of key question information based on the question text. The key question information may be key information configured for performing recognition to obtain the question text. The key question information may be a token in the question text, a synonym of the token, or a text configured for recognizing the token. The key question information may represent a key feature corresponding to the question text recognized in the image. For example, when the question text is "what type of car is shown in the image", the key question information may be "car" or a text for recognizing the "car", which may include "a wheel", "a vehicle door", "a vehicle logo", or the like.

**[00124]** The computer device may perform attention processing on the image content in the question answering attention region and the T pieces of key question information, to obtain a question attention result vector, and determine, from the question answering attention region based on the question attention result vector, a key question region associated with the T pieces of key question information. The image content in each key question region may be configured for representing the question text. The computer device may determine the second to-be-processed region as a region in the question answering attention region other than the key question region. The

computer device may obtain a question answering attention result vector corresponding to the question answering attention region, and generate a correlation degree score for the question answering attention region based on the first attention score in the question answering attention result vector that is associated with a target connected component. The correlation degree score may be a score ranging from 0 to 1, which may be for example 0.8. The computer device may perform recording by setting the correlation degree score as a grayscale value of the image mask of the question answering attention region.

**[00125]** The computer device may determine the protection level score of the question answering attention region based on the correlation degree score. The protection level score of the question answering attention region may be a difference between a value of 1 and the correlation degree score of the question answering attention region. For example, when the correlation degree score of the question answering attention region is 0.8, the protection level score of the question answering attention region may be 0.2. For the content protection processing mode corresponding to each protection level score, refer to Table 1 in the foregoing operation S205. Details are not described herein again.

**[00126]** The computer device may determine the target content protection processing mode for the question answering attention region based on the protection level score of the question answering attention region, determine the second image content as the image content in the second to-be-processed region, and perform content protection processing on the second image content based on the target content protection processing mode, to obtain second processed content. For detailed descriptions of the target content protection processing mode, refer to the specific description of the foregoing operation S205. Details are not described herein again. Redrawn pixel points in the second processed content may include pixel points of image content associated with the content protection information.

**[00127]** The computer device may replace, in the image, the first processed content with the first image content, and replace the second processed content with the second image content, to obtain a protected question-answering image including the first processed content and the second processed content.

**[00128]** Because the question answering attention region and the content protection region may be both image masks, the first to-be-processed region and the second to-be-processed region may also be both image masks. The image mask is configured to hide or display a specific part of an image. The image mask may be a binary image of the same size as the original image (to be specific, each pixel point has only two values, 0 and 1), where 0 denotes transparent (or hidden),

and 1 denotes opaque (or visible). By combining the image mask with the original image, it is possible to control which portions of the image are visible and which portions are hidden. Because the image mask may include image vertex coordinates defining a region, the computer device may determine a position of specific image content (for example, the image content in the question answering attention region or the image content in the content protection region) by using the image vertex coordinates in the image mask, perform localized re-drawing on the image content in the second to-be-processed region based on the target content protection processing mode corresponding to the content protection region and the position of the second to-be-processed region, and perform localized re-drawing on the image content in the question answering attention region based on the target content protection processing mode corresponding to the question answering attention region and the position of the question answering attention region, to obtain a protected question-answering image. The protected question-answering image may be configured to generate an answer text corresponding to the question text. Localized re-drawing may refer to performing content protection processing on the image content in the first to-be-processed region by using an image mask, and directly covering or replacing the original image content with generated new image content, so as to obtain a protected question-answering image including the first processed content. Similarly, content protection processing is performed on the image content in the second to-be-processed region, and the original image content is directly covered or replaced with the generated new image content, to obtain a protected question-answering image including the first processed content and the second processed content. Localized re-drawing technology is an image processing method for modifying a specific region of an image without affecting other portions. The method includes selecting a region to be modified, generating new content to replace the original content, and seamlessly fusing the new content with the original image. This method is commonly applied to image inpainting, object replacement, and region enhancement, which is generating new content through copy-paste, texture synthesis, or deep learning models, and using a blending algorithm to ensure natural transition.

**[00129]** In embodiments of this application, a question answering attention region and a content protection region are determined, so that image content required by a model to process question answering can be distinguished, and then a first to-be-processed region is determined from the content protection region based on the question answering attention region, and content protection processing is performed on first image content in the first to-be-processed region, thereby reducing the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model. Meanwhile, since the first to-be-processed region and the question answering attention region do

not overlap, the model can still recognize, by using the question answering attention region, the image content associated with the question text. Accordingly, this application can improve the accuracy of question answering while ensuring the security of the content protection information. The protection level score is determined through the content protection region, and different content protection processing modes for the first to-be-processed region are determined through the protection level score, which can enhance diversity of content protection processing, and reduce the risk of content protection information leakage.

**[00130]** In another aspect, according embodiments of this application, the security of the image content in the question answering attention region may be further enhanced. In other words, content protection processing is also performed on the image content in the question answering attention region. The correlation degree score is determined through the question answering attention region. Different content protection processing modes for the second to-be-processed region are determined through the correlation degree score, and different desensitization strategies may be adopted for the image content in the first to-be-processed region. Following the principle of "minimum necessary information disclosure", this can minimize the impact on the accuracy of question answering as much as possible while protecting the content protection information.

**[00131]** FIG. 6 is a schematic diagram 3 of a data processing scenario according to an embodiment of this application. As shown in FIG. 6, an object (a user) may upload, in a terminal device 100, an image and a question text corresponding to the image by using an application client. The question text corresponding to the image may be a text obtained by performing speech recognition on a voice input by the object, or a text directly input by the object on the application client. This is not limited herein in embodiments of this application. For example, the question text may be "a type of a car in an image".

**[00132]** The terminal device 100 may obtain content protection information. The content protection information may be description information for private image content in the image The content protection information may be a type description or a specific description directly specified, may be configured by selecting from private content types in the application client, may be input by the object, or may be obtained through automatic recognition by a model. The content protection information may be text data, image data, or audio data. Types of the content protection information may include personal identity information, contact information, financial information, vehicle information, medical information, address location information, and the like. The content protection information may be for example a "person/portrait".

**[00133]** In some embodiments, when the content protection information is image data, the

image data may be an image including an image box that indicates private content. To be specific, when uploading an image, the user may frame and select, by using the image box, the private content subject to content protection processing, thereby obtaining the content protection information.

**[00134]** An on-device multimodal model may be deployed in the terminal device 100. The on-device multimodal model may refer to an artificial intelligence model that runs on the terminal device 100 and is capable of simultaneously processing data of a plurality of modalities (such as text, images, and audio), which may be for example a multimodal large language model such as a Phi-3-mini-128k-instruct model (an open-source lightweight model), a Chat-GLM model (a natural language generation model), or a Qwen-VLM model (a large-scale visual language model) operable on an edge device.

**[00135]** The terminal device 100 may perform attention processing on the image and the question text by using the on-device multimodal model, and determine, from the image by using an obtained question answering attention result vector, a question answering attention region associated with the question text. Similarly, the terminal device 100 may perform attention processing on the image and the content protection information by using the on-device multimodal model, and determine, from the image by using an obtained content protection result vector, a content protection region associated with the content protection information. Both the question answering attention region and the content protection region may be image masks defined by polygon vertices.

**[00136]** If an overlapping region exists between the question answering attention region and the content protection region, the terminal device 100 may determine a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determine a first to-be-processed region as a region in the content protection region other than the service overlap region. The terminal device 100 further protects image content associated with the content protection information in the question answering attention region, and the terminal device 100 may determine the question answering attention region as a second to-be-processed region.

**[00137]** The terminal device 100 may determine a protection level score by using the content protection result vector corresponding to the content protection region, and determine a target content protection processing mode for content protection processing of the first to-be-processed region by using the protection level score of the content protection region. The terminal device 100 may perform content protection processing on the first to-be-processed region by using the target

content protection processing mode of the content protection region, to obtain first processed content.

**[00138]** The terminal device 100 may determine a correlation degree score by using the question answering attention result vector corresponding to the question answering attention region, and determine a content protection processing mode for content protection processing of the second to-be-processed region by using the correlation degree score of the question answering attention region. The terminal device 100 may perform content protection processing on the second to-be-processed region by using the target content protection processing mode of the question answering attention region, to obtain second processed content. The content protection processing mode may include image concealment re-drawing, image blurring re-drawing, same-category replacement re-drawing, feature blurring re-drawing, and the like. This is not limited herein in embodiments of this application.

**[00139]** The terminal device 100 may determine a protected question-answering image as an image in which the first image content has been replaced with the first processed content and the second image content has been replaced with the second processed content. The protected question-answering image and the question text are transmitted to a service server 200. The service server recognizes the protected question-answering image by using a cloud-based multimodal model, to generate an answer text corresponding to the question text, and returns the answer text to the terminal device 100.

**[00140]** In embodiments of this application, a question answering attention region and a content protection region are determined, so that image content required by a model to process question answering can be distinguished, and then a first to-be-processed region is determined from the content protection region based on the question answering attention region, and content protection processing is performed on first image content in the first to-be-processed region, thereby reducing the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model. Meanwhile, since the first to-be-processed region and the question answering attention region do not overlap, the model can still recognize, by using the question answering attention region, the image content associated with the question text. Accordingly, this application can appropriately improve security of the content protection information while ensuring accuracy of question answering. The protection level score is determined through the content protection region, and different content protection processing modes for the first to-be-processed region are determined through the protection level score, which can enhance diversity of content protection processing,

and reduce the risk of content protection information leakage.

**[00141]** In another aspect, according embodiments of this application, the security of the image content in the question answering attention region may be further enhanced. In other words, content protection processing is also performed on the image content in the question answering attention region. The correlation degree score is determined through the question answering attention region. Different content protection processing modes for the second to-be-processed region are determined through the correlation degree score, and different desensitization strategies may be adopted for the image content in the first to-be-processed region. Following the principle of "minimum necessary information disclosure", this can minimize the impact on the accuracy of question answering as much as possible while protecting the content protection information.

**[00142]** FIG. 7 is a schematic structural diagram of a data processing apparatus according to an embodiment of this application. As shown in FIG. 7, a data processing apparatus 1 includes a question answering region processing module 710, a protection region processing module 720, and a re-drawing region processing module 730.

**[00143]** The question answering region processing module 710 is configured to determine a question answering attention region from an image based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text.

**[00144]** The protection region processing module 720 is configured to obtain content protection information, and determine a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information.

**[00145]** The re-drawing region processing module 730 is configured to determine a first to-be-processed region from the content protection region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping, and perform content protection processing on first image content in the first to-be-processed region in the image.

**[00146]** In one embodiment, when determining a first to-be-processed region from the content protection region based on the question answering attention region, the re-drawing region processing module 730 is specifically configured to perform the following operations:

determining a service overlap region as an overlapping region between the question

answering attention region and the content protection region, and determining the first to-be-processed region as a region in the content protection region other than the service overlap region.

**[00147]** In one embodiment, the re-drawing region processing module 730 is further configured to perform the following operations:

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content, and replacing the first image content with the first processed content in the image, to obtain a protected question-answering image including the first processed content, the protected question-answering image being configured for generating an answer text corresponding to the question text.

**[00148]** In one embodiment, the re-drawing region processing module 730 is further configured to perform the following operations:

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content;

determining a second to-be-processed region from the question answering attention region, determining second image content as image content in the second to-be-processed region, and performing content protection processing on the second image content to obtain second processed content; and

replacing, in the image, the first processed content with the first image content, and replacing the second processed content with the second image content, to obtain a protected question-answering image including the first processed content and the second processed content.

**[00149]** In one embodiment, when determining a second to-be-processed region from the question answering attention region, the re-drawing region processing module 730 is specifically configured to perform the following operations:

generating T pieces of key question information based on the question text, T being a positive integer, and the key question information being key information configured for performing recognition to obtain the question text;

performing attention processing on the image content in the question answering attention region and the T pieces of key question information, to obtain a question attention result vector, and determining, from the question answering attention region based on the question attention result vector, a key question region associated with the T pieces of key question information; and

determining the second to-be-processed region as a region in the question answering attention region other than the key question region.

**[00150]** In one embodiment, when determining the question answering attention region from the image based on the image and the question text corresponding to the image, the question answering region processing module 710 is specifically configured to perform the following operations:

performing connected component labeling on each pixel point in the image to obtain M connected components, M being a positive integer, and each of the connected components being configured for representing an image region in which pixel values of pixel points in the image meet a connected component labeling condition and positions of the pixel points are adjacent to each other;

respectively performing feature extraction on the M connected components to obtain an image feature vector, and performing feature extraction on the question text corresponding to the image to obtain a question text feature vector; and

performing attention processing on the image feature vector and the question text feature vector, to obtain the question answering attention result vector, and determining the question answering attention region from the image based on the question answering attention result vector.

**[00151]** In one embodiment, when determining the question answering attention region from the image based on the question answering attention result vector, the question answering region processing module 710 is specifically configured to perform the following operations:

generating attention scores respectively corresponding to the M connected components based on the question answering attention result vector; and

determining a target connected component as a connected component having a maximum attention score, generating boundary contour points based on position information of pixel points in the target connected component, and determining the question answering attention region as a region defined by the boundary contour points in the image.

**[00152]** In one embodiment, when performing content protection processing on the image content in the first to-be-processed region to obtain the first processed content, the re-drawing region processing module 730 is specifically configured to perform the following operations:

obtaining a protection level score associated with the content protection region, the protection level score being generated based on a content protection result vector corresponding to the content protection information, the content protection result vector being an attention vector

obtained by performing attention processing on the content protection information and the image;

determining a target re-drawing level based on the protection level score, and determining, from N content protection processing modes, a target content protection processing mode indicated by the target re-drawing level, N being a positive integer; and

performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content.

**[00153]** In one embodiment, the target content protection processing mode is image concealment re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module 730 is specifically configured to perform the following operations:

determining re-drawing source content as image content in the image other than the first to-be-processed region;

obtaining a background image in the re-drawing source content, and performing feature extraction on the background image, to obtain a background feature vector; and

performing re-drawing processing on the image content in the first to-be-processed region based on the background feature vector, to obtain the first processed content, a content feature vector of the first processed content having a similarity relation with the background feature vector.

**[00154]** In one embodiment, the target content protection processing mode is image blurring re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module 730 is specifically configured to perform the following operations:

obtaining a blurring unit value P, and dividing the image content in the first to-be-processed region into P to-be-blurred image units based on the blurring unit value P, P being a positive integer, and the P to-be-blurred image units including a target to-be-blurred image unit;

generating average color channel information of the target to-be-blurred image unit based on color channel information corresponding to each pixel point in the target to-be-blurred image unit, and performing blurring processing on the pixel point in the target to-be-blurred image unit based on the average color channel information, to obtain a blurred image unit; and

determining the first processed content as the P blurred image units when blurred image units respectively corresponding to the P to-be-blurred image units are obtained.

**[00155]** In one embodiment, the target content protection processing mode is same-category replacement re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module 730 is specifically configured to perform the following operations:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information;

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, and determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information;

obtaining a content prompt text associated with the content protection information, determining, from the content prompt text, a target prompt text associated with the key object regions, and performing feature extraction on the target prompt text to obtain a text prompt feature vector;

performing corruption processing respectively on image content corresponding to the key object regions to obtain corrupted content, determining re-drawing source content as image content in the image other than the first to-be-processed region, and performing feature extraction on the re-drawing source content to obtain a re-drawing source feature vector; and

performing image inpainting on the corrupted content based on the text prompt feature vector and the re-drawing source feature vector to obtain the first processed content.

**[00156]** In one embodiment, the target content protection processing mode is feature blurring re-drawing, and when performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content, the re-drawing region processing module 730 is specifically configured to perform the following operations:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information; and

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information, and performing image blurring re-drawing on image content corresponding to the key object regions, to obtain the first processed content.

**[00157]** In embodiments of this application, a question answering attention region and a content protection region are determined, so that image content required by a model to process question answering can be distinguished, and then a first to-be-processed region is determined from the content protection region based on the question answering attention region, and content protection processing is performed on first image content in the first to-be-processed region, thereby reducing the occurrence of leakage of image content associated with the content protection information that may arise during image learning and question text answering of the model. Meanwhile, since the first to-be-processed region and the question answering attention region do not overlap, the model can still recognize, by using the question answering attention region, the image content associated with the question text. Accordingly, this application can appropriately improve security of the content protection information while ensuring accuracy of question answering. The protection level score is determined through the content protection region, and different content protection processing modes for the first to-be-processed region are determined through the protection level score, which can enhance diversity of content protection processing, and reduce the risk of content protection information leakage.

**[00158]** In another aspect, according embodiments of this application, the security of the image content in the question answering attention region may be further enhanced. In other words, content protection processing is also performed on the image content in the question answering attention region. The correlation degree score is determined through the question answering attention region. Different content protection processing modes for the second to-be-processed region are determined through the correlation degree score, and different desensitization strategies may be adopted for the image content in the first to-be-processed region. Following the principle of "minimum necessary information disclosure", this can minimize the impact on the accuracy of question answering as much as possible while protecting the content protection information.

**[00159]** In embodiments of this application, the term "module" or "unit" refers to a computer program or a part of the computer program that has a predetermined function, and operates together with another relevant part to achieve a predetermined objective, and may be entirely or partially

implemented by using software, hardware (such as a processing circuit or a memory), or a combination thereof. Similarly, one processor (or a plurality of processors or memories) may be configured to implement one or more modules or units. In addition, each module or unit may be a part of an overall module or unit including a function of the module or unit.

**[00160]** FIG. 8 is a schematic structural diagram of a computer device according to an embodiment of this application. As shown in FIG. 8, a computer device 1000 may include a processor 1001, a network interface 1004, and a memory 1005. In addition, the computer device 1000 may further include a user interface 1003 and at least one communication bus 1002. The communication bus 1002 is configured to implement connection and communication between the components. The user interface 1003 may include a display and a keyboard. In some embodiments, the user interface 1003 may further include a standard wired interface and a standard wireless interface. In some embodiments, the network interface 1004 may include a standard wired interface and a standard wireless interface (such as a Wi-Fi interface). The memory 1005 may be a high-speed random access memory (RAM) memory, or may be a non-volatile memory, for example, at least one magnetic disk memory. In some embodiments, the memory 1005 may further be at least one storage apparatus away from the foregoing processor 1001. As shown in FIG. 8, the memory 1005, as a computer-readable storage medium, may include an operating system, a network communication module, a user interface module, and a device control application program.

**[00161]** In the computer device 1000 shown in FIG. 8, the network interface 1004 may provide a network communication function. The user interface 1003 is mainly configured to provide an input interface for a user. The processor 1001 may be configured to call a device control application stored in the memory 1005 to:

determine a question answering attention region from an image based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text;

obtain content protection information, and determine a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information; and

determine a first to-be-processed region from the content protection region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping, and

perform the content protection processing on first image content in the first to-be-processed region, the first image content being image content in the first to-be-processed region in the image.

**[00162]** The computer device 1000 described in embodiments of this application may perform the description of the data processing method in the foregoing embodiment corresponding to either of FIG. 3 and FIG. 4. Details are not described herein again. In addition, beneficial effects of using the same method are not described in detail herein.

**[00163]** In addition, an embodiment of this application further provides a computer-readable storage medium. The foregoing computer-readable storage medium stores a computer program. When executing the foregoing computer program, the foregoing processor can perform the foregoing description of the data processing method in the embodiment corresponding to either of FIG. 3 and FIG. 4. Therefore, details are not described herein again. In addition, beneficial effects of using the same method are not described in detail herein. For technical details not disclosed in the embodiment of the computer-readable storage medium involved in this application, reference is made to the description of the method embodiment of this application.

**[00164]** The foregoing computer-readable storage medium may be an internal storage unit in the data processing apparatus in any one of the foregoing embodiments or in the foregoing computer device, for example, a hard disk or an internal memory of the computer device. The computer-readable storage medium may alternatively be an external storage device in the computer device, for example, a plug-in hard disk, a smart media card (SMC), a secure digital (SD) card, and a flash card equipped on the computer device. Further, the computer-readable storage medium may further include both the internal storage unit and the external storage device of the computer device. The computer-readable storage medium is configured to store the computer program and other programs and data required by the computer device. The computer-readable storage medium may be further configured to temporarily store data that has been displayed or that is to be displayed.

**[00165]** In addition, an embodiment of this application further provides a computer program product, the computer program product including a computer program. The computer program is stored in a computer-readable storage medium. A processor of a computer device reads the computer program from the computer-readable storage medium. The processor executes the computer program, causing the computer device to perform the foregoing method provided in the embodiment corresponding to either of FIG. 3 and FIG. 4.

**[00166]** The terms "first", "second", and the like in the specification, the claims, and the accompanying drawings of embodiments of this application are configured for distinguishing

different objects, rather than being configured for describing a specific order. Moreover, the term "include" and any variation thereof are intended to cover non-exclusive inclusions. For example, processes, methods, apparatuses, products, or devices including a series of operations or units are not limited to the listed operations or modules, but instead, include operations or modules not listed in some embodiments, or include other operations or units inherent in these processes, methods, apparatuses, products, or devices in some embodiments.

**[00167]** A person of ordinary skill in the art may realize that operations of units and algorithms of various examples described with reference to embodiments disclosed in this specification can be implemented in electronic hardware, computer software, or a combination of the electronic hardware and the computer software. To clearly describe interchangeability of hardware and software, the compositions and operations of the various examples are generally described in terms of the network element in the foregoing description. Whether the network elements are performed by hardware or software depends on specific applications and design constraints of the technical solutions. A person skilled in the art may use different methods to implement the described network element for each particular application, but it is not considered that the implementation goes beyond the scope of this application.

**[00168]** The method and the related apparatus provided in the embodiments of this application are described with reference to the method flowcharts and/or schematic structural diagrams provided in the embodiments of this application. Specifically, each process and/or block in the method flowcharts and/or schematic structural diagrams and a combination of processes and/or blocks in the flowcharts and/or block diagrams may be implemented through computer program instructions. These computer program instructions may be provided to a processor of a general-purpose computer, a special-purpose computer, an embedded processing machine, or another programmable device to generate a machine, so that instructions executed by the processor of the computer or the another programmable device generate an apparatus for implementing functions specified in one or more processes of the flowcharts and/or one or more blocks of the schematic structural diagrams. These computer program instructions may also be stored in a computer-readable memory that can direct a computer or another programmable device to operate in a specific manner, so that the instructions stored in the computer-readable memory produce a product including an instruction apparatus. The instruction apparatus implements the functions specified in one or more processes of the flowchart and/or one or more blocks of the schematic structural diagram. These computer program instructions may also be loaded onto a computer or another programmable device, so that a series of operations are performed on the computer or the another programmable device to generate a computer-implemented process, and the instructions

executed on the computer or the another programmable device provide operations for implementing the functions specified in one or more processes of the flowchart and/or one or more blocks of the schematic structural diagram.

**[00169]** The operations of the method of embodiments of this application may be reordered, merged, and deleted based on actual requirements.

**[00170]** The modules of the apparatus of embodiments of this application may be merged, divided, and deleted based on an actual requirement.

**[00171]** What is disclosed above is merely preferred embodiments of this application, and certainly is not intended to limit the scope of the claims of this application. Therefore, equivalent variations made in accordance with the claims of this application still fall within the scope of this application.

CLAIMS

What is claimed is:

1. A data processing method, comprising:

determining, from an image, a question answering attention region based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text;

obtaining content protection information, and determining a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information, and the content protection information being a preset type description or a directly specified specific description;

determining, from the content protection region, a first to-be-processed region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping; and

performing content protection processing on first image content in the first to-be-processed region in the image.

2. The method according to claim 1, wherein the determining a first to-be-processed region from the content protection region based on the question answering attention region comprises:

determining a service overlap region as an overlapping region between the question answering attention region and the content protection region, and determining the first to-be-processed region as a region in the content protection region other than the service overlap region.

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

performing content protection processing on the image content in the first to-be-processed region, to obtain first processed content, and replacing the first image content with the first processed content in the image, to obtain a protected question-answering image comprising the first processed content, the protected question-answering image being configured for generating an answer text corresponding to the question text.

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

performing content protection processing on the image content in the first to-be-processed

region, to obtain first processed content;

determining a second to-be-processed region from the question answering attention region, determining second image content as image content in the second to-be-processed region, and performing content protection processing on the second image content to obtain second processed content; and

replacing, in the image, the first processed content with the first image content, and replacing the second processed content with the second image content, to obtain a protected question-answering image.

5. The method according to claim 4, wherein the determining a second to-be-processed region from the question answering attention region comprises:

generating T pieces of key question information based on the question text, T being a positive integer, and the key question information being key information configured for performing recognition to obtain the question text;

performing attention processing on the image content in the question answering attention region and the T pieces of key question information, to obtain a question attention result vector, and determining, from the question answering attention region based on the question attention result vector, a key question region associated with the T pieces of key question information; and

determining the second to-be-processed region as a region in the question answering attention region other than the key question region.

6. The method according to claim 1, wherein the determining a question answering attention region from an image based on the image and a question text corresponding to the image comprises:

performing connected component labeling on each pixel point in the image to obtain M connected components, M being a positive integer, and each of the connected components being configured for representing an image region in which pixel values of pixel points in the image meet a connected component labeling condition and positions of the pixel points are adjacent to each other;

respectively performing feature extraction on the M connected components to obtain an image feature vector, and performing feature extraction on the question text corresponding to the image to obtain a question text feature vector; and

performing attention processing on the image feature vector and the question text feature vector, to obtain the question answering attention result vector, and determining the question

answering attention region from the image based on the question answering attention result vector.

7. The method according to claim 6, wherein the determining the question answering attention region from the image based on the question answering attention result vector comprises:

generating attention scores respectively corresponding to the M connected components based on the question answering attention result vector; and

determining a target connected component as a connected component having a maximum attention score, generating boundary contour points based on position information of pixel points in the target connected component, and determining the question answering attention region as a region defined by the boundary contour points in the image.

8. The method according to claim 3, wherein the performing content protection processing on the image content in the first to-be-processed region to obtain first processed content comprises:

obtaining a protection level score associated with the content protection region, the protection level score being generated based on a content protection result vector corresponding to the content protection information, the content protection result vector being an attention vector obtained by performing attention processing on the content protection information and the image;

determining a target re-drawing level based on the protection level score, and determining, from N content protection processing modes, a target content protection processing mode indicated by the target re-drawing level, N being a positive integer; and

performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content.

9. The method according to claim 8, wherein the target content protection processing mode is image concealment re-drawing, and the performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content comprises:

determining re-drawing source content as image content in the image other than the first to-be-processed region;

obtaining a background image in the re-drawing source content, and performing feature extraction on the background image, to obtain a background feature vector; and

performing re-drawing processing on the image content in the first to-be-processed region based on the background feature vector, to obtain the first processed content, a content feature vector of the first processed content having a similarity relation with the background feature vector.

10. The method according to claim 8, wherein the target content protection processing mode is image blurring re-drawing, and the performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content comprises:

obtaining a blurring unit value P, and dividing the image content in the first to-be-processed region into P to-be-blurred image units based on the blurring unit value P, P being a positive integer, and the P to-be-blurred image units comprising a target to-be-blurred image unit;

generating average color channel information of the target to-be-blurred image unit based on color channel information corresponding to each pixel point in the target to-be-blurred image unit, and performing blurring processing on the pixel point in the target to-be-blurred image unit based on the average color channel information, to obtain a blurred image unit; and

determining the first processed content as the P blurred image units when blurred image units respectively corresponding to the P to-be-blurred image units are obtained.

11. The method according to claim 8, wherein the target content protection processing mode is same-category replacement re-drawing, and the performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content comprises:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information;

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, and determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information;

obtaining a content prompt text associated with the content protection information, determining, from the content prompt text, a target prompt text associated with the key object regions, and performing feature extraction on the target prompt text to obtain a text prompt feature vector;

performing corruption processing respectively on image content corresponding to the key object regions to obtain corrupted content, determining re-drawing source content as image content in the image other than the first to-be-processed region, and performing feature extraction on the re-drawing source content to obtain a re-drawing source feature vector; and

performing image inpainting on the corrupted content based on the text prompt feature vector and the re-drawing source feature vector to obtain the first processed content.

12. The method according to claim 8, wherein the target content protection processing mode is feature blurring re-drawing, and the performing content protection processing on the image content in the first to-be-processed region based on the target content protection processing mode, to obtain the first processed content comprises:

generating K pieces of key object information based on the content protection information, K being a positive integer, and the key object information being key information configured for performing recognition to obtain the content protection information;

performing attention processing on the image content in the first to-be-processed region and the K pieces of key object information, to obtain an object attention result vector, determining, from the first to-be-processed region based on the object attention result vector, key object regions associated with the K pieces of key object information, and performing image blurring re-drawing on image content corresponding to the key object regions, to obtain the first processed content.

13. A data processing apparatus, comprising:

a question answering region processing module, configured to determine a question answering attention region from an image based on the image and a question text corresponding to the image, image content in the question answering attention region being image content in the image associated with the question text;

a protection region processing module, configured to obtain content protection information, and determine a content protection region from the image based on the image and the content protection information, image content in the content protection region being image content in the image associated with the content protection information, and the content protection information being a preset type description or a directly specified specific description;

a re-drawing region processing module, configured to determine a first to-be-processed region from the content protection region based on the question answering attention region when an overlapping region exists between the question answering attention region and the content protection region, the first to-be-processed region and the question answering attention region not overlapping, and perform content protection processing on first image content in the first to-be-processed region in the image.

14. A computer device, comprising a processor, a memory, and a network interface,

the processor being connected to the memory and the network interface, the network interface

being configured to provide a data communication function, the memory being configured to store a computer program, and the processor being configured to invoke the computer program, to enable the computer device to perform the method according to any one of claims 1 to 12.

15. A computer-readable storage medium, having a computer program stored therein, the computer program being adapted to be loaded and executed by a processor, to enable a computer device having the processor to perform the method according to any one of claims 1 to 12.

16. A computer program product, comprising a computer program, the computer program being stored in a computer-readable storage medium and adapted to be read and executed by a processor, to enable a computer device having the processor to perform the method according to any one of claims 1 to 12.