

IMAGE PROCESSING METHOD AND DEVICE, ELECTRONIC EQUIPMENT, STORAGE MEDIUM AND PROGRAM

TECHNICAL FIELD

[0001] The disclosure relates to computer vision processing technologies, and particularly to an image processing method and device, electronic equipment, a computer storage medium and a computer program.

BACKGROUND

[0002] In the related technology, a Reference-based Super-Resolution (Ref-SR) technology attracts extensive attentions. Compared with single-image super-resolution taking a single low-resolution image as an input, the Ref-SR technology performs super-resolution processing on a specified low-resolution image using an additional high-resolution image as a reference image to obtain a super-resolution image. In the process of super-resolution processing, the reference image may provide more references in terms of image details for the process of super-resolution processing.

[0003] The key of the Ref-SR technology is searching for a correspondence between pixels of the low-resolution image and the reference image. However, a resolution difference between the low-resolution image and the reference image makes it difficult to find the correspondence between the pixels of the low-resolution image and the reference image accurately, bringing influence to the quality of the generated super-resolution image.

SUMMARY

[0004] Embodiments of the disclosure provide technical solutions to image processing.

[0005] The embodiments of the disclosure provide an image processing method, which may include the following operations.

[0006] An image to be processed and a reference image are acquired, a resolution of the reference image being higher than a resolution of the image to be processed.

[0007] The image to be processed and the reference image are processed using a first contrastive relation network to obtain a first correspondence between pixels in the image to be processed and the reference image, the first contrastive relation network being obtained by performing supervised training based on a first relation between a feature of a first sample image and a feature of a sample reference image and a second sample image, the feature of the first sample image and the feature of the sample reference image being data obtained based on a trained second contrastive relation network, and resolutions of both the first sample image and the sample reference image being higher than a resolution of the second sample image.

[0008] A feature of the reference image is fused into the image to be processed according to the first correspondence to obtain a super-resolution image of the image to be processed.

[0009] In some embodiments, a training process of the first contrastive relation network may include the following operations.

[0010] A second relation between a feature of the second sample image and the feature of the sample reference image is determined.

[0011] A network parameter value of the first contrastive relation network is regulated according to the first relation and the second relation.

[0012] In some embodiments, the first relation is configured to represent correlation data between the features in the first sample image and the sample reference image, and the second relation is configured to represent correlation data between the features in the second sample image and the sample reference image.

[0013] The operation that the network parameter value of the first contrastive relation

network is regulated according to the first relation and the second relation includes the following operations.

[0014] A loss of the first contrastive relation network is determined according to the first relation and the second relation, and the network parameter value of the first contrastive relation network is regulated according to the loss of the first contrastive relation network, the loss of the first relation network including a correlation loss, and the correlation loss representing a difference between the first relation and the second relation.

[0015] In some embodiments, the second relation is further configured to represent distance information between a feature of a pixel in the second sample image and a feature of a pixel in the sample reference image.

[0016] The operation that the loss of the first contrastive relation network is determined according to the first relation and the second relation includes the following operations.

[0017] A triple loss is determined according to the distance information represented by the second relation, and the loss of the first contrastive relation network is determined according to the first relation, the second relation and the triple loss, the loss of the first contrastive relation network further including the triple loss.

[0018] In some embodiments, resolutions of at least one group of sample images used for training the second contrastive relation network are higher than the resolution of the second sample image.

[0019] In some embodiments, the method may further include the following operations.

[0020] The first sample image is acquired. Homography transformation is performed on the first sample image to obtain the sample reference image. Down-sampling processing is performed on the first sample image to obtain the second sample image.

[0021] In some embodiments, the operation that the feature of the reference image is

fused into the image to be processed according to the first correspondence to obtain the super-resolution image of the image to be processed includes the following operations.

[0022] For each pixel in the image to be processed, the following operations are executed.

[0023] A target pixel corresponding to the pixel in the image to be processed in the reference image is determined according to the first correspondence.

[0024] A position offset of the target pixel in the reference image is determined according to a position difference between the pixel and the corresponding target pixel, a pixel after position offsetting of the target pixel is determined according to a position of the target pixel and the position offset, and a reference feature is obtained, the reference feature representing a feature of the pixel after position offsetting of the target pixel.

[0025] A feature of each pixel in the image to be processed is fused with the reference feature corresponding to each pixel to obtain a fused feature, and the super-resolution image of the image to be processed is obtained according to the fused features.

[0026] In some embodiments, the feature of the reference image includes a texture feature.

[0027] In some embodiments, training data for the second contrastive relation network includes at least part of images among the first sample image and the sample reference image.

[0028] The embodiments of the disclosure also provide an image processing device, which includes an acquisition module, a first processing module and a second processing module.

[0029] The acquisition module is configured to acquire an image to be processed and a reference image, a resolution of the reference image being higher than a resolution of the image to be processed.

[0030] The first processing module is configured to process the image to be processed and the reference image using a first contrastive relation network to obtain a first

correspondence between pixels in the image to be processed and the reference image, the first contrastive relation network being obtained by performing supervised training based on a first relation between a feature of a first sample image and a feature of a sample reference image and a second sample image, the feature of the first sample image and the feature of the sample reference image being data obtained based on a trained second contrastive relation network and resolutions of both the first sample image and the sample reference image being higher than a resolution of the second sample image.

[0031] The second processing module is configured to fuse a feature of the reference image into the image to be processed according to the first correspondence to obtain a super-resolution image of the image to be processed.

[0032] In some embodiments, the device further includes a training module, and the training module is used for training the first contrastive relation network through the following steps.

[0033] A second relation between a feature of the second sample image and the feature of the sample reference image is determined.

[0034] A network parameter value of the first contrastive relation network is regulated according to the first relation and the second relation.

[0035] In some embodiments, the first relation is configured to represent correlation data between the features in the first sample image and the sample reference image, and the second relation is configured to represent correlation data between the features in the second sample image and the sample reference image.

[0036] The training module is specifically configured to determine a loss of the first contrastive relation network according to the first relation and the second relation, and regulate the network parameter value of the first contrastive relation network according to the loss of the first contrastive relation network, the loss of the first relation network including a correlation loss, and the correlation loss representing a difference between the first relation and the second relation.

[0037] In some embodiments, the second relation is further configured to represent

distance information between a feature of a pixel in the second sample image and a feature of a pixel in the sample reference image.

[0038] The training module is specifically configured to determine a triple loss according to the distance information represented by the second relation and determine the loss of the first contrastive relation network according to the first relation, the second relation and the triple loss, the loss of the first contrastive relation network further including the triple loss.

[0039] In some embodiments, resolutions of at least one group of sample images used for training the second contrastive relation network are higher than the resolution of the second sample image.

[0040] In some embodiments, the acquisition module is further configured to acquire the first sample image, perform homography transformation on the first sample image to obtain the sample reference image, and perform down-sampling processing on the first sample image to obtain the second sample image.

[0041] In some embodiments, the second processing module is specifically configured to execute the following operations.

[0042] For each pixel in the image to be processed, the second processing module is configured to execute the following operations.

[0043] A target pixel corresponding to the pixel in the image to be processed in the reference image is determined according to the first correspondence.

[0044] A position offset of the target pixel in the reference image is determined according to a position difference between the pixel and the corresponding target pixel, a pixel after position offsetting of the target pixel is determined according to a position of the target pixel and the position offset, and a reference feature is obtained, the reference feature representing a feature of the pixel after position offsetting of the target pixel.

[0045] The second processing module is further configured to fuse a feature of each pixel in the image to be processed and the reference feature corresponding to each pixel

to obtain a fused feature, and obtain the super-resolution image of the image to be processed according to the fused features.

[0046] In some embodiments, the feature of the reference image may include a texture feature.

[0047] In some embodiments, training data for the second contrastive relation network may include at least part of images among the first sample image and the sample reference image.

[0048] The embodiments of the disclosure also provide electronic equipment, which includes a processor and a memory configured to store a computer program capable of running in the processor.

[0049] The processor may be configured to run the computer program to execute any image processing method described above.

[0050] The embodiments of the disclosure also provide a computer storage medium having stored therein computer program which is executed by a processor to implement any image processing method described above.

[0051] The embodiments of the disclosure also provide a computer program including a computer-readable code which runs in electronic equipment to enable a processor of the electronic equipment to implement any image processing method described above.

[0052] In the image processing method and device, electronic equipment and computer storage medium disclosed in the embodiments of the disclosure, at first, the image to be processed and the reference image are acquired, the resolution of the reference image being higher than the resolution of the image to be processed; then, the image to be processed and the reference image are processed using the first contrastive relation network to obtain the first correspondence between the pixels in the image to be processed and the reference image, the first contrastive relation network being obtained by performing supervised training based on the first relation between the feature of the first sample image and the feature of the sample reference image and the second sample image, the feature of the first sample image and the feature of the sample reference image

being the data obtained based on the trained second contrastive relation network and the resolutions of both the first sample image and the sample reference image being higher than the resolution of the second sample image; finally, the feature of the reference image is fused into the image to be processed according to the first correspondence to obtain the super-resolution image of the image to be processed.

[0053] It can be seen that, in the embodiments of the disclosure, the first contrastive relation network may be trained based on the relation between the feature of the first sample image and the feature of the sample reference image, the resolutions of both the first sample image and the sample reference image being higher than the resolution of the second sample image. That is, in the embodiments of the disclosure, the first contrastive relation network may be trained based on a relation between two images with relatively high resolutions. Compared with a matching relation between a low-resolution image and a high-resolution image, a matching relation between two high-resolution images may provide more accurate and comprehensive matching information. Therefore, supervised training based on the first relation between the feature of the first sample image and the feature of the sample reference image and the second sample image is favorable for obtaining the correspondence between the pixels in the image to be processed and the reference image more accurately through the trained first contrastive relation network and further improving the quality of the super-resolution image of the image to be processed on the basis of the accurately obtained correspondence between the pixels in the image to be processed and the reference image.

[0054] It is to be understood that the above general description and the following detailed description are only exemplary and explanatory and not intended to limit the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0055] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments consistent with the disclosure and, together with the specification, serve to describe the technical solutions of the disclosure.

[0056] FIG. 1 is a flowchart of an image processing method according to an embodiment of the disclosure.

[0057] FIG. 2 is a schematic diagram of a training flow of a first contrastive relation network and a second contrastive relation network according to an embodiment of the disclosure.

[0058] FIG. 3 is a flow block diagram of performing super-resolution processing using a restoration network according to an embodiment of the disclosure.

[0059] FIG. 4A is a schematic diagram of correspondences between an image to be processed and a reference image according to an embodiment of the disclosure and the related art.

[0060] FIG. 4B is a super-resolution image of an image to be processed according to the related art.

[0061] FIG. 4C is a super-resolution image of an image to be processed according to an embodiment of the disclosure.

[0062] FIG. 5 is a composition structure diagram of an image processing device according to an embodiment of the disclosure.

[0063] FIG. 6 is a composition structure diagram of electronic equipment according to an embodiment of the disclosure.

DETAILED DESCRIPTION

[0064] The disclosure will further be described below in combination with the drawings and the embodiments in detail. It is to be understood that the embodiments provided herein are only adopted to explain the disclosure and not intended to limit the disclosure. In addition, the embodiments provided below are not all embodiments implementing the disclosure but part of embodiments implementing the disclosure, and the technical solutions recorded in the embodiments of the disclosure may be freely

combined for implementation without conflicts.

[0065] It is to be noted that, in the embodiments of the disclosure, terms "include" and "contain" or any other variant thereof is intended to cover nonexclusive inclusions herein, so that method or device including a series of elements not only includes those clearly recorded elements but also includes other elements which are not clearly listed or further includes intrinsic elements for implementing the method or the device. Under the condition of no more limitations, an element defined by a statement "including a/an....." does not exclude existence of another related element in a method or device including the element (for example, step in the method or a unit in the device, the unit may be, for example, part of a circuit, part of a processor and part of a program or software).

[0066] For example, an image processing method provided in the embodiments of the disclosure includes a series of steps, but the image processing method provided in the embodiments of the disclosure is not limited to the recorded steps. Similarly, an image processing device provided in the embodiments of the disclosure includes a series of modules, but the device provided in the embodiments of the disclosure is not limited to include the clearly recorded modules and may further include a module required to be arranged under the condition that related information is acquired or processing is performed based on information.

[0067] In the disclosure, term "and/or" is only an association relationship describing associated objects and represents that three relationships may exist. For example, A and/or B may represent three conditions: existence of A alone, existence of both A and B, and existence of B alone. In addition, term "at least one" in the disclosure represents any one of multiple or any combination of at least two of multiple. For example, including at least one of A, B and C may represent including any one or more elements selected from a set formed by A, B and C.

[0068] The embodiments of the disclosure may be applied to a computer system that consists of a terminal and/or a server and may be operated together with numerous other universal or dedicated computing system environments or configurations. Herein, the terminal may be a thin client, a thick client, a handheld or laptop device, a microprocessor-based system, a set-top box, a programmable consumer electronic

product, a network personal computer and the like. The server may be a microcomputer system, a large computer system, a distributed cloud computing technical environment including any abovementioned system, and the like.

[0069] Electronic equipment such as the terminal and the server may realize a corresponding function through execution of a program module. Under a normal condition, the program module may include a routine, a program, a target program, a component, a logic, a data structure, etc. The computer system/server may be implemented in a distributed cloud computing environment, and in the distributed cloud computing environment, tasks are executed by a remote processing device connected through a communication network. In the distributed cloud computing environment, the program module may be in a storage medium of a local or remote computer system including a storage device.

[0070] The key of a Ref-SR technology is searching for a correspondence between pixels of a low-resolution image and a reference image. In the related technology, the low-resolution image and a high-resolution image taken as the reference image may be matched implicitly to determine the correspondence between the pixels of the low-resolution image and the reference image, the correspondence being obtained according to a content similarity and appearance similarity of the images. After the correspondence is determined, super-resolution processing may be performed on the low-resolution image based on the correspondence.

[0071] However, a resolution difference between the low-resolution image and the high-resolution image taken as the reference image makes it difficult to obtain the correspondence between the pixels of the low-resolution image and the reference image accurately. Herein, due to the unbalanced amounts of information in the low-resolution image and the high-resolution image taken as the reference image, down-sampling processing is usually performed on the high-resolution image to obtain a down-sampled image, the down-sampled image being the same as the low-resolution image in resolution. However, down-sampling processing inevitably causes an information loss. For example, information of a fine-texture region in the high-resolution image may be lost. Consequently, it is difficult to obtain the correspondence between the pixels of the low-

resolution image and the reference image accurately.

[0072] For the foregoing technical problem, the disclosure provides an image processing method to perform super-resolution processing on various low-resolution images. Herein, the low-resolution image may include, but not limited to, a scene image (for example, an outdoor scene image, an indoor scene image), an object image (for example, a building image, an animal image, an plant image), etc.

[0073] FIG. 1 is a flowchart of an image processing method according to an embodiment of the disclosure. As shown in FIG. 1, the flow may include the following steps.

[0074] In 101, an image to be processed and a reference image are acquired, a resolution of the reference image being higher than a resolution of the image to be processed.

[0075] Herein, the image to be processed may be an image requiring super-resolution processing. Exemplarily, the image to be processed may be acquired from a local storage region or through a network. For example, the image to be processed may be an image shot through a mobile terminal, a camera or the like. The format of the image to be processed may be Joint Photographic Experts Group (JPEG), Bitmap (BMP), Portable Network Graphics (PNG) or other format. It is to be noted that the format and source of the image to be processed are only described herein with examples, and the format and source of the image to be processed are not limited in the embodiment of the disclosure.

[0076] In the embodiment of the disclosure, the reference image is a pre-acquired image. In a first example, a reference image of which the similarity with the image to be processed is greater than a similarity threshold may be selected from a pre-stored image set. The similarity threshold may be an empirically set value. For example, the similarity threshold is 80%, 85% or 90%. In a second example, a reference image relatively similar to the image to be processed may be selected manually from the pre-stored image set. In a third example, multiple images may be found based on the image to be processed by using a search engine in an image-based searching manner, and a reference image of which the similarity with the image to be processed is greater than the similarity

threshold may be selected from the multiple images that are found, or, a reference image relatively similar to the image to be processed may be selected manually. It can be understood that searching for the image through the search engine is favorable for obtaining the reference image with a relatively high similarity with the image to be processed and favorable for obtaining diversified reference images.

[0077] In 102, the image to be processed and the reference image are processed using a first contrastive relation network to obtain a first correspondence between pixels in the image to be processed and the reference image, the first contrastive relation network being obtained by performing supervised training based on a first relation between a feature of a first sample image and a feature of a sample reference image and a second sample image, the feature of the first sample image and the feature of the sample reference image being data obtained based on a trained second contrastive relation network, resolutions of both the first sample image and the sample reference image being higher than a resolution of the second sample image.

[0078] In the embodiment of the disclosure, the first contrastive relation network may be a neural network adopting a deep learning technology. In some embodiments, the first contrastive relation network includes a feature extraction subnetwork, which is configured to extract an image feature. Exemplarily, the feature extraction subnetwork includes a first feature extraction unit and a second feature extraction unit. The first feature extraction unit is configured to extract a feature of the image to be processed, and the second feature extraction unit is configured to extract a feature of the reference image. For example, each feature extraction unit may include a convolutional layer and a pooling layer.

[0079] Considering that the image to be processed and the reference image are different in resolution, the first feature extraction unit and the second feature extraction unit may be two different network units, for example, network units with the same network structure but different network parameters. Herein, the network parameter may be a weight or other parameter. Of course, network structures of the two network units may also be different, and no limits are made herein.

[0080] The first contrastive relation network may further include a feature processing

subnetwork. After the feature of the image to be processed and the feature of the reference image are extracted, the feature of the image to be processed and the feature of the reference image may be processed using the feature processing subnetwork to obtain the first correspondence between the pixels in the image to be processed and the reference image. Exemplarily, the feature processing subnetwork may include a convolutional layer, an activation layer and an output layer.

[0081] It can be seen that, for training the first contrastive relation network, it is necessary to pre-train the second contrastive relation network and then process the input first sample image and sample reference image using the trained second contrastive relation network to obtain the first relation between the feature of the first sample image and the feature of the sample reference image. That is, input data in the second contrastive relation network is different from input data in the first contrastive relation network. The input data in the second contrastive relation network includes the first sample image and the sample reference image, the first sample image and the sample reference image being two images with relatively high resolutions (the resolutions are both higher than the second sample image). The input data of the first contrastive relation network includes the image to be processed and the reference image.

[0082] In the embodiment of the disclosure, since the first relation between the feature of the first sample image and the feature of the sample reference image is required to be determined, the number of sample reference images may be the same as the number of first sample images. Of course, different first sample images may also correspond to the same sample reference image, namely the number of first sample images may also be larger than the number of sample reference images.

[0083] The second contrastive relation network may include a feature extraction subnetwork. The feature extraction subnetwork in the second contrastive relation network is configured to extract the feature of the first sample image and the feature of the sample reference image. A structure of the feature extraction subnetwork in the second contrastive relation network may be the same as a structure of the feature extraction subnetwork in the first contrastive relation network, or may be different from the structure of the feature extraction subnetwork in the first contrastive relation network.

[0084] An acquisition manner for the first sample image, the second sample image, the sample reference image and training data for the second contrastive relation network will be described below.

[0085] In a first example, the sample reference image may be acquired according to the manner for acquiring the reference image after the first sample image or the second sample image is acquired.

[0086] In a second example, the first sample image, the second sample image, the sample reference image or the training data for the second contrastive relation network may be acquired from a common dataset, or the first sample image, the second sample image, the sample reference image or the training data for the second contrastive relation network may also be acquired from the local storage region. For example, the common dataset may be a CUFED5 dataset, a SUN dataset, a Manga 109 dataset, etc.

[0087] In a third example, at least part of images among the first sample image and the sample reference image may be directly determined as the training data for the second contrastive relation network after the first sample image and the sample reference image are acquired, or an image completely different from the first sample image and the sample reference image may also be determined as the training data for the second contrastive relation network.

[0088] Considering that features of two images need to be obtained from the second contrastive relation network, the training data for the second contrastive relation network may include at least one group of sample images, and each group of sample images in the at least one group of sample images may include two sample images, i.e., the first sample image and the sample reference image corresponding to the first sample image.

[0089] In combination with the contents recorded above, it can be seen that a relation between features of two images with relatively high resolutions needs to be output from the second contrastive relation network, namely the first relation between the feature of the first sample image and the feature of the sample reference image needs to be output from the second contrastive relation network. Therefore, as an example, resolutions of the at least one group of sample images used for training the second contrastive relation

network are higher than the resolution of the second sample image.

[0090] It can be understood that the second contrastive relation network is obtained by training based on two images with relatively high resolutions and, moreover, the resolutions of both the first sample image and the sample reference image are higher than the resolution of the second sample image, namely both the first sample image and the sample reference image are images with relatively high resolutions, so that using the trained second contrastive relation network is favorable for extracting the corresponding features from the first sample image and the sample reference image accurately and further obtaining the first relation between the feature of the first sample image and the feature of the sample reference image accurately.

[0091] A training process of the second contrastive relation network will be exemplarily described below in combination with the contents recorded above.

[0092] The training process of the second contrastive relation network may include that: the training data for the second contrastive relation network is input to the second contrastive relation network, and each group of sample images in the training data are processed using the second contrastive relation network to obtain a feature of each sample image in each group of sample images; and then a network parameter value of the second contrastive relation network is regulated according to a correspondence between pixels of the two sample images in each group of sample images and the feature of each sample image in each group of sample images.

[0093] It can be seen that, for training the second contrastive relation network, it is necessary to pre-acquire the correspondence between the pixels of the two sample images in each group of sample images. Exemplarily, homography transformation may be performed on one sample image in each group of sample image to obtain another sample image in each group of sample image. Herein, homography transformation may include at least one of rotation and scaling. Homography transformation may reflect a correspondence between pixels of images, so that the correspondence between the pixels of the two sample images in each group of sample images may be determined.

[0094] An implementation mode of regulating the network parameter value of the

second contrastive relation network may be as follows: a loss of the second contrastive relation network is obtained according to the correspondence between the pixels of the two sample images in each group of sample images and the feature of each sample image in each group of sample images; and the network parameter value of the second contrastive relation network is regulated according to the loss of the second contrastive relation network.

[0095] In the embodiment of the disclosure, under the condition that the second contrastive relation network after regulation of the network parameter value does not meet a training ending condition, the steps that the feature of each sample image in each group of sample images is obtained, the loss of the second contrastive relation network is obtained and the network parameter value of the second contrastive relation network is regulated according to the loss of the second contrastive relation network are re-executed. Under the condition that the second contrastive relation network after regulation of the network parameter value meets the training ending condition, the second contrastive relation network after regulation of the network parameter value is determined as the trained second contrastive relation network.

[0096] The training ending condition may be that an iteration number of the training process of the second contrastive relation network reaches a set number or image processing of the second contrastive relation network after regulation of the network parameter value meets a set accuracy requirement. Herein, the set number may be an empirically set value. For example, the set number may be set with reference to a feature extraction network similar to the second contrastive relation network. The set accuracy requirement is correlated with the loss of the second contrastive relation network. For example, the set accuracy requirement may be that the loss of the second contrastive relation network is less than a set loss, and the set loss may be set with reference to the feature extraction network similar to the second contrastive relation network.

[0097] In 103, a feature of the reference image is fused into the image to be processed according to the first correspondence to obtain a super-resolution image of the image to be processed.

[0098] The feature of the reference image may include a texture feature. In such case,

the super-resolution image of the image to be processed may include the texture feature of the reference image.

[0099] 103 may be implemented based on a restoration network. Specifically, the feature of the image to be processed and the feature of the reference image may be input to the trained restoration network. In the restoration network, the feature of the reference image is fused into the image to be processed according to the first correspondence to obtain the super-resolution image of the image to be processed. The restoration network may output the super-resolution image of the image to be processed.

[00100] In practical applications, 101 to 103 may be implemented using a processor in electronic equipment. The processor may be at least one of an Application Specific Integrated Circuit (ASIC), a Digital Signal Processor (DSP), a Digital Signal Processing Device (DSPD), a Programmable Logic Device (PLD), a Field-Programmable Gate Array (FPGA), a Central Processing unit (CPU), a controller, a microcontroller and a microprocessor.

[00101] It can be seen that, in the embodiment of the disclosure, the first contrastive relation network may be trained based on the relation between the feature of the first sample image and the feature of the sample reference image, the resolutions of both the first sample image and the sample reference image being higher than the resolution of the second sample image. That is, in the embodiment of the disclosure, the first contrastive relation network may be trained based on a relation between two images with relatively high resolutions. Compared with a matching relation between a low-resolution image and a high-resolution image, a matching relation between two high-resolution images may provide more accurate matching information. Therefore, performing supervised training based on the first relation between the feature of the first sample image and the feature of the sample reference image and the second sample image is favorable for obtaining the correspondence between the pixels in the image to be processed and the reference image (i.e., a correspondence between pixels in a low-resolution image and a high-resolution image) more accurately through the trained first contrastive relation network and further improving the quality of the super-resolution image of the image to be processed on the basis of the accurately obtained correspondence between the pixels in the image to be

processed and the reference image.

[00102] With adoption of the technical solution of the embodiment of the disclosure, under the condition that a low-resolution image is required to be restored, super-resolution processing may be performed using a reference image similar to a scene of the low-resolution image, and moreover, relational matching may be performed on the low-resolution image and the high-resolution reference image accurately to restore a texture (for example, a text, a tree) in the low-resolution image.

[00103] A training process of the first contrastive relation network will be exemplarily described below.

[00104] The training process of the first contrastive relation network may include that: a second relation between a feature of the second sample image and the feature of the sample reference image is determined; and a network parameter value of the first contrastive relation network is regulated according to the first relation and the second relation.

[00105] In the embodiment of the disclosure, the second relation may be distance information, similarity information or other information between the feature of the second sample image and the feature of the sample reference image. Since the second relation between the feature of the second sample image and the feature of the sample reference image is required to be determined, the number of sample reference images may be the same as the number of second sample images.

[00106] It can be seen that, in the embodiment of the disclosure, the first contrastive relation network may be trained in combination with the first relation and the second relation. The first relation represents a relation between the feature of the first sample image and the feature of the sample reference image, the resolutions of both the first sample image and the sample reference image are higher than the resolution of the second sample image, and the second relation represents a relation between the second sample image with a relatively low resolution and the sample reference image with a relatively high resolution. Therefore, in the embodiment of the disclosure, training the first contrastive relation network based on the relation between the two images with the

relatively high resolutions and the relation between the second sample image with the relatively low resolution and the sample image with the relatively high resolution is favorable for fusing a matching relation between the two high-resolution images into the second relation and further obtaining the correspondence between the pixels in the image to be processed and the reference image more accurately through the trained first contrastive relation network. That is, under the condition that the second contrastive relation network is determined as a teacher network and the first contrastive relation network is determined as a student network, a feature extracted by the teacher network may be transferred to the student network by knowledge distillation, and supervised training may be performed on the student network through the teacher network, so that the correspondence between the pixels in the image to be processed and the reference image may be helped to be obtained accurately through the trained first contrastive relation network.

[00107] The first relation may be configured to represent correlation data between the features in the first sample image and the sample reference image, and the second relation may be configured to represent correlation data between the features in the second sample image and the sample reference image.

[00108] Correspondingly, the operation that the network parameter value of the first contrastive relation network is regulated according to the first relation and the second relation may include that: a loss of the first contrastive relation network is determined according to the first relation and the second relation; and the network parameter value of the first contrastive relation network is regulated according to the loss of the first contrastive relation network.

[00109] Herein, the loss of the first contrastive relation network may include a correlation loss, and the correlation loss represents a difference between the first relation and the second relation.

[00110] In the embodiment of the disclosure, for each feature of the second sample image, correlation data with each feature of the sample reference image may be determined. Exemplarily, there are N features in each of the first sample image and the second sample image and M features in the sample reference image, and the correlation

data between each feature of the second sample image and each feature of the sample reference image may be calculated to obtain the correlation data between the feature of the second sample image and the feature of the sample reference image. The correlation data between the feature of the second sample image and the feature of the sample reference image may be represented by an $N \times M$ matrix. Herein, both N and M are integers greater than 1. Exemplarily, each feature in the first sample image, the second sample image or the reference image may be a feature in a vector form.

[00111] Exemplarily, correlation data between a feature of a pixel p in the second sample image and a feature of a pixel q in the sample reference image may be calculated according to a formula (1):

$$cor_{pq} = \frac{e^{f_p \cdot f_q / \tau}}{\sum_{k \in I'} e^{f_p \cdot f_k / \tau}} \quad (1);$$

where cor_{pq} represents the correlation data between the feature of the pixel p in the second sample image and the feature of the pixel q in the sample reference image, f_p represents the feature of the pixel p in the second sample image, f_q represents the feature of the pixel q in the sample reference image, f_k represents a feature of a pixel k in the sample reference image, I' represents a set of pixels in the sample reference image, and τ is a preset parameter. For example, τ is 0.14, 0.15 or 0.16.

[00112] In the embodiment of the disclosure, correlation data between a feature of each pixel in the first sample image and a feature of each pixel in the sample reference image may be calculated with reference to the calculation method of the formula (1), thereby determining the correlation data between the features in the first sample image and the sample reference image. Herein, the correlation data between the feature of the first sample image and the feature of the sample reference image may be recorded as COR_p^T , and the correlation data between the feature of the second sample image and the feature of the sample reference image may be recorded as COR_p^S .

[00113] The correlation loss may represent any one of a Kullback-Leibler (KL) divergence, a Jensen-Shannon (JS) divergence, a cross entropy and a Wasserstein distance.

[00114] Under the condition that the correlation loss represents the KL divergence, the correlation loss may be calculated according to a formula (2) and a formula (3):

$$\text{Div}_p = \text{KL}(\text{COR}_p^T \square \text{COR}_p^S) = \sum_{k \in I'} \text{cor}_{pk}^T \log\left(\frac{\text{cor}_{pk}^T}{\text{cor}_{pk}^S}\right) \quad (2);$$

$$L_{kl} = \frac{1}{N} \sum_{p \in I} \text{Div}_p \quad (3);$$

where, cor_{pk}^T represents correlation data between the feature of the pixel p in the first sample image and the feature of the pixel k in the sample reference image in the correlation data COR_p^T , cor_{pk}^S represents correlation data between the feature of the pixel p in the second sample image and the feature of the pixel k in the sample reference image in the correlation data COR_p^S , I represents a set of pixels in the second sample image, and L_{kl} represents the correlation loss.

[00115] It can be understood that the correlation loss represents the difference between the first relation and the second relation and the first relation is configured to represent the correlation data between the features in the first sample image and the sample reference image, so that training the first contrastive relation network through the correlation loss is favorable for transferring the matching relation between the first sample image and the sample reference image to the matching relation between the second sample image and the sample reference image by knowledge distillation to help to obtain the correspondence between the pixels in the image to be processed with the relatively low resolution and the reference image with the relatively high resolution more accurately through the trained first contrastive relation network.

[00116] In the embodiment of the disclosure, data in the second relation may not be

limited to the correlation data between the features in the second sample image and the sample reference image. For example, the second relation may further be configured to represent distance information between a feature of a pixel in the second sample image and a feature of a pixel in the sample reference image.

[00117] Correspondingly, the operation that the loss of the first contrastive relation network is determined according to the first relation and the second relation may include that: a triple loss is determined according to the distance information represented by the second relation; and the loss of the first contrastive relation network is determined according to the first relation, the second relation and the triple loss, the loss of the first contrastive relation network further including the triple loss.

[00118] In the embodiment of the disclosure, the triple loss is determined according to a distance between features of a positive sample pair and a distance between features of a negative sample pair. The positive sample pair and the negative sample pair will be described below with the pixel p of the second sample image as an example. The positive sample pair may include the pixel p of the second sample image and a pixel p' in the sample reference image, and in the positive sample pair, the pixel p' in the sample reference image is a corresponding pixel of the pixel p of the second sample image. The negative sample pair includes the pixel p of the second sample image and the pixel k in the sample reference image, and in the negative sample pair, a distance between the pixel k in the sample reference image and the pixel p' in the sample reference image is greater than a set threshold T . For example, the set threshold T may be 3.5, 4 or 4.5. Herein, the pixel p of the second sample image, the pixel p' in the sample reference image and the pixel k in the sample reference image form a triple.

[00119] The triple loss may be calculated according to a formula (4):

$$L_{\text{margin}} = \frac{1}{N} \sum_{p \in I} \max(0, m + \text{Pos}(p) - \text{Neg}(p)) \quad (4);$$

where, L_{margin} represents the triple loss, and m represents a preset margin value. For example, m may be 0.9, 1 or 1.1. $\text{Pos}(p)$ may be calculated according to the feature

distance of the positive sample pair, and $\text{Neg}(p)$ may be calculated according to the feature distance of the negative sample pair.

[00120] In some embodiments, $\text{Pos}(p)$ may be calculated according to a formula (5):

$$\text{Pos}(p) = \|f_p - f_{p'}\|_2^2 \quad (5);$$

where, $f_{p'}$ represents a feature of the pixel p' in the sample reference image, and $\|\cdot\|_2$ represents a 2-norm.

[00121] In some embodiments, $\text{Neg}(p)$ may be calculated according to a formula (6):

$$\text{Neg}(p) = \min\left(\min_{k \in I', \|k-p'\|_\infty > T} \|f_p - f_k\|_2^2, \min_{k \in I, \|k-p\|_\infty > T} \|f_{p'} - f_k\|_2^2\right) \quad (6);$$

where, $\|\cdot\|_\infty$ represents a ∞ -norm, $k - p'$ represents a position difference between the pixel k in the sample reference image and the pixel p' in the sample reference image, and $k - p$ represents a position difference between the pixel k in the second sample image and the pixel p in the second sample image.

[00122] In the related technology, a low-resolution image and a high-resolution image taken as a reference image usually have the same object or similar texture patterns. However, a transformation relation (for example, scaling transformation and rotation transformation) between the low-resolution image and the high-resolution image taken as the reference image may cause relatively great appearance and content differences between the low-resolution image and the high-resolution image taken as the reference image. In such case, a correspondence, which is calculated between pixels of the images according to an appearance similarity and a content similarity only is usually inaccurate, and furthermore, a high-resolution texture in the high-resolution image cannot be transferred accurately. In the embodiment of the disclosure, the triple loss is positively correlated with the feature distance of the positive sample pair and negatively correlated with the feature distance of the negative sample pair, so that training the first contrastive relation network based on the triple loss is favorable for finding a pixel closest to a

feature of each pixel in the low-resolution image to be processed from the high-resolution image taken as the reference image through the trained first contrastive relation network, the correspondence between the pixels of the low-resolution image to be processed and the high-resolution image taken as the reference image may be obtained accurately and reliably in the embodiment of the disclosure for the transformation relation between the low-resolution image to be processed and the reference image, and moreover, reduction of a pixel matching error caused by the transformation relation between the low-resolution image and the high-resolution image taken as the reference image is facilitated.

[00123] In combination with the contents recorded above, it can be seen that the triple loss is determined according to the distance between the features of the positive sample pair and the distance between the features of the negative sample pair, and both the positive sample pair and the negative sample pair are required to be determined according to a correspondence between the pixels in the second sample image and the sample reference image, namely a real correspondence between the pixel in the sample reference image and the pixel in the second sample image is required to be predetermined.

[00124] In order to conveniently determine the real correspondence between the pixel in the sample reference image and the pixel in the second sample image, after the first sample image is acquired, homography transformation may be performed on the first sample image to obtain the sample reference image, and down-sampling processing is performed on the first sample image to obtain the second sample image.

[00125] It can be understood that, since the sample reference image is an image obtained by performing homography transformation on the first sample image, a real correspondence between the pixels in the first sample image and the sample reference image may be determined directly, and since the second sample image is an image obtained by performing down-sampling processing on the first sample image, a real correspondence between the pixels in the first sample image and the second sample image may be determined directly and the real correspondence between the pixels in the second sample image and the sample reference image may further be determined according to the real correspondence between the pixels in the first sample image and the sample reference image and the real correspondence between the pixels in the first

sample image and the second sample image.

[00126] Under the condition that the loss L of the first contrastive relation network includes both the correlation loss L_{kl} and the triple loss L_{margin} , the loss L of the first contrastive relation network may be calculated according to a formula (7):

$$L = L_{\text{margin}} + \alpha_{kl} \cdot L_{kl} \quad (7);$$

where, α_{kl} represents a weight of the correlation loss L_{kl} , and α_{kl} may be preset according to a practical requirement. In an example, α_{kl} may be 14, 15 or 16. In another example, α_{kl} may be set to be a relatively large value in a scenario that a resolution difference between the image to be processed and the reference image is relatively great, and α_{kl} may be set to be a relatively small value in a scenario that the resolution difference between the image to be processed and the reference image is relatively small. A learning rate of the first contrastive relation network may be determined by trying values, namely the learning rate of the first contrastive relation network may be set to be multiple values, the first contrastive relation network may be trained for each value of the learning rate of the first contrastive relation network, then the first contrastive relation network is tested through test data to obtain a test result of the first contrastive relation network, and the value of the learning rate of the first contrastive relation network may finally be determined according to the test result corresponding to each value of the learning rate of the first contrastive relation network. For example, the learning rate of the first contrastive relation network may be 10^{-3} .

[00127] Of course, as an implementation mode, the network loss of the second contrastive relation network may include a triple loss, and the triple loss of the second contrastive relation network is configured to measure a distance between pixel features of the two sample images in each group of sample images. A calculation manner for the triple loss of the second contrastive relation network may refer to the calculation manner for the triple loss of the first contrastive relation network and will not be elaborated herein.

[00128] A training flow of the first contrastive relation network in the embodiment of

the disclosure will be described below with reference to the drawings.

[00129] FIG. 2 is a schematic diagram of a training flow of a first contrastive relation network and a second contrastive relation network according to an embodiment of the disclosure. Referring to FIG. 2, after the trained second contrastive relation network is obtained, the feature of the first sample image and the feature of the sample reference image may be extracted using the second contrastive relation network to further determine the correlation data COR_p^T between the feature of the first sample image and the feature of the sample reference image.

[00130] In the embodiment of the disclosure, the triple loss L_{margin} in the first contrastive relation network may be calculated with reference to the formula (4) to the formula (6), and the correlation data COR_p^S between the feature of the second sample image and the feature of the sample reference image may be determined according to the feature of the second sample image and the feature of the sample reference image.

[00131] After the correlation data COR_p^T and the correlation data COR_p^S are obtained, the correlation loss L_{kl} may be calculated according to the formula (2) and the formula (3), and then the loss L of the first contrastive relation network may be calculated according to the formula (7).

[00132] The first contrastive relation network may be trained according to the loss L of the first contrastive relation network to obtain the trained first contrastive relation network.

[00133] In some embodiments, an implementation mode of 103 may be as follows.

[00134] For each pixel in the image to be processed, the following steps are executed: a target pixel in the reference image corresponding to the pixel in the image to be processed is determined according to the first correspondence; a position offset of the target pixel in the reference image is determined according to a position difference between the pixel and the corresponding target pixel; a pixel after position offsetting of the target pixel is determined according to a position of the target pixel and the position

offset; and a reference feature is obtained, the reference feature representing a feature of the pixel after position offsetting of the target pixel.

[00135] After the reference feature corresponding to each pixel is obtained, a feature of each pixel in the image to be processed may be fused with the reference feature corresponding to each pixel to obtain a fused feature, and the super-resolution image of the image to be processed is obtained according to the fused features.

[00136] In some embodiments, there may be one or more position offsets for the target pixel. Therefore, for each pixel, under the condition that there is one position offset, there is one pixel after position offsetting of the target pixel, and furthermore, one reference feature may be obtained. For each pixel, under the condition that the number of position offsets is K , there are K pixels after position offsetting of the target pixel, and furthermore, K reference features may be obtained, K being an integer greater than 1, namely the feature of each pixel may be fused with multiple reference features in the embodiment of the disclosure.

[00137] In some embodiments, a position difference between each pixel and a corresponding target pixel may be represented by p_0 , where $p_0 = p - p'$, $p - p'$ representing a position difference between a pixel p in the image to be processed and a corresponding target pixel p' .

[00138] In some embodiments, for the pixel p in the image to be processed, a reference feature $y(p)$ may be calculated according to a formula (8):

$$y(p) = \sum_{k=1}^K w_k \cdot x(p + p_0 + p_k + \Delta p_k) \cdot \Delta m_k \quad (8);$$

Herein, w_k represents a weight of a convolution kernel, Δp_k represents a learnable offset, Δm_k represents a learnable modulation scalar, $p + p_0 + p_k + \Delta p_k$ represents a position after the pixel p is offset by $p_0 + p_k + \Delta p_k$, $x(p + p_0 + p_k + \Delta p_k)$ represents an original feature at a position $p + p_0 + p_k + \Delta p_k$ in the reference image (i.e., a feature extracted using the first contrastive relation network), and p_k is a preset position offset. For

example, p_k may be valued according to a formula (9):

$$p_k \in \{(-1, -1), (-1, 0), (-1, 1), (0, -1), (0, 0), (0, 1), (1, -1), (1, 0), (1, 1)\} \quad (9).$$

[00139] In some embodiments, referring to FIG. 3, the image to be processed and the reference image may be input to the trained first contrastive relation network, and the image to be processed and the reference image are processed using the first contrastive relation network to obtain the first correspondence between the pixels in the image to be processed and the reference image.

[00140] After the first correspondence is obtained, the first correspondence may be input to the restoration network. In the restoration network, the following steps are executed: the pixel after position offsetting of the target pixel is determined to obtain the reference feature; and the feature of each pixel in the image to be processed is fused with the reference feature corresponding to each pixel to obtain the fused feature, and the super-resolution image of the image to be processed is obtained according to the fused features.

[00141] In some embodiments, the feature of each pixel in the image to be processed is fused with the reference feature corresponding to each pixel, and the fused feature may be processed through a residual block to obtain the super-resolution image of the image to be processed.

[00142] In the embodiment of the disclosure, the restoration network may be a Deep&Cross Network (DCN) or a neural network of another type. In some embodiments, the restoration network may be trained according to a loss of the restoration network. The loss of the restoration network may include at least one of a reconstruction loss L_{rec} , a perceptual loss L_{per} and an adversarial loss L_{adv} . Exemplarily, a norm in the reconstruction loss L_{rec} may be a 1-norm, and the perceptual loss L_{per} may be calculated according to a Visual Geometry Group (VGG) feature.

[00143] In some embodiments, a weight of each loss in the reconstruction loss L_{rec} , the perceptual loss L_{per} and the adversarial loss L_{adv} may be determined by trying values,

namely the weight of each loss may be set to be multiple values, the restoration network may be trained for each value of each loss, then the restoration network is tested through test data to obtain a test result of the restoration network and the value of each loss may finally be determined according to the test result corresponding to each value of each loss. For example, the weights of the reconstruction loss L_{rec} , the perceptual loss L_{per} and the adversarial loss L_{adv} are 1.0, 10^{-4} and 10^{-6} respectively. Similarly, a learning rate for training the restoration network may also be determined by trying values. For example, the learning rate for training the restoration network may be 10^{-4} .

[00144] In some embodiments, Δp_k and Δm_k may be network parameters in the restoration network. In a process of training the restoration network, the parameters Δp_k and Δm_k may be regulated. The use of the learnable offset Δp_k enables to dynamically utilize information of pixels around the target pixel; the use of the learnable modulation scalar Δm_k enables to summarize the features in the image with reasonable weights.

[00145] It can be seen that the position difference between each pixel and the corresponding target pixel is favorable for determining the reference feature accurately, the reference feature being a feature of a pixel around the target pixel, so that fusing the feature of each pixel in the image to be processed with the corresponding reference feature is favorable for transferring the feature of the pixel around the target pixel to the image to be processed and further improving the quality of the super-resolution image of the image to be processed.

[00146] The effects of the super-resolution processing method in the embodiment of the disclosure and the related art will be comparatively described below with reference to the drawings.

[00147] FIG. 4A is a schematic diagram of correspondences between an image to be processed and a reference image according to an embodiment of the disclosure and the related art. Referring to FIG. 4A, the first connecting line 401 represents the correspondence between the pixels of the image to be processed and the reference image in the embodiment of the disclosure, and the second connecting line 402 represents the

correspondence between the pixels of the image to be processed and the reference image in the related art. It can be seen that the correspondence represented by the second connecting line 402 is inaccurate and the correspondence represented by the first connecting line 401 is more accurate.

[00148] FIG. 4B is a super-resolution image of an image to be processed according to the related art. FIG. 4C is a super-resolution image of an image to be processed according to an embodiment of the disclosure. It can be seen by comparison between FIG. 4B and FIG. 4C that, compared with the scheme of the related art, the embodiment of the disclosure has the advantages that a super-resolution image with more texture features may be obtained.

[00149] The super-resolution processing method of the embodiment of the disclosure may at least include a method 1 and a method 2. The loss function adopted in a training process of the method 1 includes the reconstruction loss L_{rec} , the perceptual loss L_{per} and the adversarial loss L_{adv} . The loss function adopted in a training process of the method 2 includes the reconstruction loss L_{rec} only. No matter which loss function is adopted, more texture features in the high-resolution image may be fused into the low-resolution image effectively to obtain the super-resolution image.

[00150] It can be understood by those skilled in the art that, in the method of the specific implementation modes, the writing sequence of each step does not mean a strict execution sequence and is not intended to form any limit to the implementation process and a specific execution sequence of each step should be determined by functions and probable internal logic thereof.

[00151] Based on the image processing method disclosed in the abovementioned embodiments, the embodiments of the disclosure disclose an image processing device.

[00152] FIG. 5 is a composition structure diagram of an image processing device according to an embodiment of the disclosure. As shown in FIG. 5, the device may include an acquisition module 701, a first processing module 702 and a second processing module 703.

- [00153] The acquisition module 701 is configured to acquire an image to be processed and a reference image, a resolution of the reference image being higher than a resolution of the image to be processed.
- [00154] The first processing module 702 is configured to process the image to be processed and the reference image using a first contrastive relation network to obtain a first correspondence between pixels in the image to be processed and the reference image, the first contrastive relation network being obtained by performing supervised training based on a first relation between a feature of a first sample image and a feature of a sample reference image and a second sample image, the feature of the first sample image and the feature of the sample reference image being data obtained based on a trained second contrastive relation network and resolutions of both the first sample image and the sample reference image being higher than a resolution of the second sample image.
- [00155] The second processing module 703 is configured to fuse a feature of the reference image into the image to be processed according to the first correspondence to obtain a super-resolution image of the image to be processed.
- [00156] In some embodiments, the device further includes a training module, and the training module is used for training the first contrastive relation network through the following steps.
- [00157] A second relation between a feature of the second sample image and the feature of the sample reference image is determined.
- [00158] A network parameter value of the first contrastive relation network is regulated according to the first relation and the second relation.
- [00159] In some embodiments, the first relation is configured to represent correlation data between the features in the first sample image and the sample reference image, and the second relation is configured to represent correlation data between the features in the second sample image and the sample reference image.
- [00160] The training module is specifically configured to determine a loss of the first contrastive relation network according to the first relation and the second relation and

regulate the network parameter value of the first contrastive relation network according to the loss of the first contrastive relation network, the loss of the first relation network including a correlation loss, and the correlation loss representing a difference between the first relation and the second relation.

[00161] In some embodiments, the second relation is further configured to represent distance information between a feature of a pixel in the second sample image and a feature of a pixel in the sample reference image.

[00162] The training module is specifically configured to determine a triple loss according to the distance information represented by the second relation and determine the loss of the first contrastive relation network according to the first relation, the second relation and the triple loss, the loss of the first contrastive relation network further including the triple loss.

[00163] In some embodiments, resolutions of at least one group of sample images used for training the second contrastive relation network are higher than the resolution of the second sample image.

[00164] In some embodiments, the acquisition module 701 is further configured to acquire the first sample image, perform homography transformation on the first sample image to obtain the sample reference image, and perform down-sampling processing on the first sample image to obtain the second sample image.

[00165] In some embodiments, the second processing module 703 is specifically configured to execute the following operations.

[00166] For each pixel in the image to be processed, the second processing module is configured to execute the following operations.

[00167] A target pixel corresponding to the pixel in the image to be processed in the reference image is determined according to the first correspondence.

[00168] A position offset of the target pixel in the reference image is determined according to a position difference between the pixel and the corresponding target pixel, a

pixel after position offsetting of the target pixel is determined according to a position of the target pixel and the position offset, and a reference feature is obtained, the reference feature representing a feature of the pixel after position offsetting of the target pixel.

[00169] The second processing module is further configured to fuse a feature of each pixel in the image to be processed and the reference feature corresponding to each pixel to obtain a fused feature, and obtain the super-resolution image of the image to be processed according to the fused features.

[00170] In some embodiments, the feature of the reference image includes a texture feature.

[00171] In some embodiments, training data for the second contrastive relation network includes at least part of images among the first sample image and the sample reference image.

[00172] In practical applications, all the acquisition module 701, the first processing module 702 and the second processing module 703 may be implemented using a processor in a computer device. The processor may be at least one of an ASIC, a DSP, a DSPD, a PLD, an FPGA, a CPU, a controller, a microcontroller and a microprocessor.

[00173] In addition, each functional module in the embodiment may be integrated into a processing unit, each unit may also exist independently, and two or more than two units may also be integrated into a unit. The integrated unit may be implemented in a hardware form and may also be implemented in form of software function module.

[00174] When implemented in form of a software function module and sold or used not as an independent product, the integrated unit may be stored in a computer-readable storage medium. Based on such an understanding, the technical solution of the embodiment substantially or parts making contributions to the conventional art or all or part of the technical solution may be embodied in form of software product, and the computer software product is stored in a storage medium, including a plurality of instructions configured to enable a computer device (which may be a personal computer, a server, a network device or the like) or a processor to execute all or part of the steps of the method in the embodiment. The storage medium includes various media capable of

storing program codes such as a U disk, a mobile Hard Disk Drive (HDD), a Read-Only Memory (ROM), a Random Access Memory (RAM), a magnetic disk or an optical disk.

[00175] Specifically, computer program instructions corresponding to the image processing method in the embodiments may be stored in a storage medium such as an optical disk, a hard disk and a U disk. When the computer program instructions corresponding to the image processing method in the storage medium are read or executed by electronic equipment, any image processing method of the abovementioned embodiments is implemented.

[00176] Based on the same technical concept of the abovementioned embodiments, referring to FIG. 6, electronic equipment 80 provided in the embodiments of the disclosure is shown, which may include a memory 801 and a processor 802.

[00177] The memory 801 is configured to store a computer program and data.

[00178] The processor 802 is configured to execute the computer program stored in the memory to implement any image processing method of the abovementioned embodiments.

[00179] During the practical application, the memory 801 may be a volatile memory such as a RAM, or a non-volatile memory such as a ROM, a flash memory, an HDD or a Solid-State Drive (SSD), or a combination of the memories, and provides instructions and data for the processor 802.

[00180] The processor 802 may be at least one of an ASIC, a DSP, a DSPD, a PLD, an FPGA, a CPU, a controller, a microcontroller and a microprocessor. It can be understood that, for different devices, other electronic components may be configured to realize functions of the processor, and no specific limits are made in the embodiment of the disclosure.

[00181] The embodiments of the disclosure also provide a computer program including a computer-readable code which runs in electronic equipment to enable a processor of the electronic equipment to implement any image processing method.

[00182] In some embodiments, functions or modules of the device provided in the embodiment of the disclosure may be configured to execute the method described in the method embodiment and specific implementation thereof may refer to the descriptions about the method embodiment and, for simplicity, will not be elaborated herein.

[00183] The above descriptions about the embodiments focus on differences between each embodiment and the same or similar parts may refer to each other and will not be elaborated herein for simplicity.

[00184] The methods disclosed in each method embodiment provided in the application may be freely combined without conflicts to obtain new method embodiments.

[00185] The characteristics disclosed in each product embodiment provided in the application may be freely combined without conflicts to obtain new product embodiments.

[00186] The characteristics disclosed in each method or device embodiment provided in the application may be freely combined without conflicts to obtain new method embodiments or device embodiments.

[00187] From the above descriptions about the implementation modes, those skilled in the art may clearly know that the method of the abovementioned embodiments may be implemented in a manner of combining software and a necessary universal hardware platform, and of course, may also be implemented through hardware, but the former is a preferred implementation mode under many circumstances. Based on such an understanding, the technical solutions of the disclosure substantially or parts making contributions to the conventional art may be embodied in form of software product, and the computer software product is stored in a storage medium (for example, a ROM/RAM, a magnetic disk and an optical disk), including a plurality of instructions configured to enable a terminal (which may be a personal computer, a server, an air conditioner, a network device or the like) to execute the method in each embodiment of the disclosure.

[00188] The embodiments of the disclosure are described above in combination with the drawings, but the disclosure is not limited to the abovementioned specific implementation modes. The abovementioned specific implementation modes are not

restrictive but only schematic, those of ordinary skill in the art may be inspired by the disclosure to implement many forms without departing from the purpose of the disclosure and the scope of protection of the claims, and all these shall fall within the scope of protection of the disclosure.

CLAIMS

1. An image processing method, comprising:

acquiring an image to be processed and a reference image, wherein a resolution of the reference image is higher than a resolution of the image to be processed;

processing the image to be processed and the reference image using a first contrastive relation network to obtain a first correspondence between pixels in the image to be processed and the reference image, wherein the first contrastive relation network is obtained by performing supervised training based on a first relation between a feature of a first sample image and a feature of a sample reference image and a second sample image, the feature of the first sample image and the feature of the sample reference image are data obtained based on a trained second contrastive relation network, resolutions of both the first sample image and the sample reference image are higher than a resolution of the second sample image; and

fusing a feature of the reference image into the image to be processed according to the first correspondence to obtain a super-resolution image of the image to be processed.

2. The method of claim 1, wherein a training process of the first contrastive relation network comprises:

determining a second relation between a feature of the second sample image and the feature of the sample reference image; and

regulating a network parameter value of the first contrastive relation network according to the first relation and the second relation.

3. The method of claim 2, wherein the first relation is configured to represent correlation data between the features in the first sample image and the sample reference image, and the second relation is configured to represent correlation data between the features in the second sample

image and the sample reference image;

wherein regulating the network parameter value of the first contrastive relation network according to the first relation and the second relation comprises:

determining a loss of the first contrastive relation network according to the first relation and the second relation, and regulating the network parameter value of the first contrastive relation network according to the loss of the first contrastive relation network, wherein the loss of the first relation network comprises a correlation loss, and the correlation loss represents a difference between the first relation and the second relation.

4. The method of claim 3, wherein the second relation is further configured to represent distance information between a feature of a pixel in the second sample image and a feature of a pixel in the sample reference image;

wherein determining the loss of the first contrastive relation network according to the first relation and the second relation comprises:

determining a triple loss according to the distance information represented by the second relation, and determining the loss of the first contrastive relation network according to the first relation, the second relation and the triple loss, wherein the loss of the first contrastive relation network further comprises the triple loss.

5. The method of any one of claims 1 to 4, wherein resolutions of at least one group of sample images used for training the second contrastive relation network are higher than the resolution of the second sample image.

6. The method of any one of claims 1 to 5, further comprising:

acquiring the first sample image, performing homography transformation on the first sample image to obtain the sample reference image, and performing down-sampling

processing on the first sample image to obtain the second sample image.

7. The method of any one of claims 1 to 6, wherein fusing the feature of the reference image into the image to be processed according to the first correspondence to obtain the super-resolution image of the image to be processed comprises:

for each pixel in the image to be processed, executing the following operations:

determining a target pixel in the reference image corresponding to the pixel in the image to be processed according to the first correspondence, and

determining a position offset of the target pixel in the reference image according to a position difference between the pixel and the corresponding target pixel, determining a pixel after position offsetting of the target pixel according to a position of the target pixel and the position offset, and obtaining a reference feature, wherein the reference feature represents a feature of the pixel after position offsetting of the target pixel; and

fusing a feature of each pixel in the image to be processed and the reference feature corresponding to each pixel to obtain a fused feature, and obtaining the super-resolution image of the image to be processed according to the fused features.

8. The method of any one of claims 1 to 7, wherein the feature of the reference image comprises a texture feature.

9. The method of any one of claims 1 to 8, wherein training data for the second contrastive relation network comprises at least part of images among the first sample image and the sample reference image.

10. An image processing device, comprising:

an acquisition module, configured to acquire an image to be processed and a reference image, wherein a resolution of the reference image is higher than a resolution of the image to be processed;

a first processing module, configured to process the image to be processed and the reference image using a first contrastive relation network to obtain a first correspondence between pixels in the image to be processed and the reference image, wherein the first contrastive relation network is obtained by performing supervised training based on a first relation between a feature of a first sample image and a feature of a sample reference image and a second sample image, the feature of the first sample image and the feature of the sample reference image are data obtained based on a trained second contrastive relation network and resolutions of both the first sample image and the sample reference image are higher than a resolution of the second sample image; and

a second processing module, configured to fuse a feature of the reference image into the image to be processed according to the first correspondence to obtain a super-resolution image of the image to be processed.

11. The device of claim 10, further comprising a training module, wherein the training module is used for training the first contrastive relation network through following steps:

determining a second relation between a feature of the second sample image and the feature of the sample reference image and

regulating a network parameter value of the first contrastive relation network according to the first relation and the second relation.

12. The device of claim 11, wherein the first relation is configured to represent correlation data between the features in the first sample image and the sample reference image, and the second relation is configured to represent correlation data between the features in the second

sample image and the sample reference image;

wherein the training module is specifically configured to:

determine a loss of the first contrastive relation network according to the first relation and the second relation, and regulate the network parameter value of the first contrastive relation network according to the loss of the first contrastive relation network, wherein the loss of the first relation network comprises a correlation loss, and the correlation loss represents a difference between the first relation and the second relation.

13. The device of claim 12, wherein the second relation is further configured to represent distance information between a feature of a pixel in the second sample image and a feature of a pixel in the sample reference image;

wherein the training module is specifically configured to:

determine a triple loss according to the distance information represented by the second relation and determine the loss of the first contrastive relation network according to the first relation, the second relation and the triple loss, wherein the loss of the first contrastive relation network further comprises the triple loss.

14. The device of any one of claims 10 to 13, wherein resolutions of at least one group of sample images used for training the second contrastive relation network are higher than the resolution of the second sample image.

15. The device of any one of claims 10 to 14, wherein the acquisition module is further configured to:

acquire the first sample image, perform homography transformation on the first sample image to obtain the sample reference image, and perform down-sampling processing on the first sample image to obtain the second sample image.

16. The device of any one of claims 10 to 15, wherein the second processing module is specifically configured to:

for each pixel in the image to be processed, execute following operations:

determining a target pixel in the reference image corresponding to the pixel in the image to be processed according to the first correspondence, and

determining a position offset of the target pixel in the reference image according to a position difference between the pixel and the corresponding target pixel, determining a pixel after position offsetting of the target pixel according to a position of the target pixel and the position offset, and obtaining a reference feature, wherein the reference feature represents a feature of the pixel after position offsetting of the target pixel; and

fuse a feature of each pixel in the image to be processed and the reference feature corresponding to each pixel to obtain a fused feature, and obtain the super-resolution image of the image to be processed according to the fused features.

17. The device of any one of claims 10 to 16, wherein the feature of the reference image comprises a texture feature.

18. Electronic equipment, comprising a processor and a memory configured to store a computer program capable of running in the processor, wherein

the processor is configured to run the computer program to execute the method of any one of claims 1 to 8.

19. A computer storage medium having stored therein a computer program which is executed

by a processor to implement the method of any one of claims 1 to 8.

20. A computer program, comprising computer-readable codes which run in electronic equipment to enable a processor in the electronic equipment to implement the method of any one of claims 1 to 8.