

IMAGE SEGMENTATION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND COMPUTER-READABLE STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The present application is based upon and claims priority to Chinese patent application No. 202110246708.5, filed on March 05, 2021 and entitled "IMAGE SEGMENTATION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND COMPUTER-READABLE STORAGE MEDIUM", the disclosure of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] Embodiments of the present disclosure relate to the technical field of artificial intelligence, and in particular, to a method for image segmentation and apparatus, an electronic device, and a computer-readable storage medium.

BACKGROUND

[0003] Semantic segmentation aims to predict the semantic category of each pixel of an input image, is one of the most basic problems in the field of computer vision, and is widely applied in various fields, including but not limited to autonomous driving, man-machine interaction, etc. At present, a semantic segmentation method mainly focuses on using contextual information of high-level features to perform network inference through deep Fully Convolutional Networks (FCN), so as to obtain a semantic segmentation result of the image. However, because the high-level features are extracted from a large receptive field of a neural network, using only deep advanced features to perform semantic inference will lead to a rough and inaccurate output, thereby reducing the accuracy of image segmentation.

SUMMARY

- [0004] The embodiments of the present disclosure provide a method and apparatus for image segmentation, an electronic device, and a computer-readable storage medium, which can improve the accuracy of semantic inference.
- [0005] Technical solutions of the embodiments of the present disclosure are implemented as follows.
- [0006] A first aspect of the embodiments of the present disclosure provides a method for image segmentation, which may include the following operations.
- [0007] Low-level image features are extracted from an image to be processed, and semantic segmentation is performed on the low-level image features to obtain high-level semantic features.
- [0008] At least one of texture enhancement and texture feature statistics is performed on the low-level image features to obtain low-level texture features, the low-level texture features representing at least details of the texture enhancement or statistical distribution of texture features of the image to be processed.
- [0009] A semantic segmentation image is obtained by combining the high-level semantic features and the low-level texture features.
- [0010] In the above method, the operation that at least one of texture enhancement and texture feature statistics is performed on the low-level image features to obtain the low-level texture features may include any one of the following.
- [0011] Texture enhancement is performed based on the low-level image features to obtain an enhanced texture feature; and the low-level texture features are determined based on the enhanced texture features.
- [0012] Texture feature statistics is performed based on the low-level image features to obtain a statistical texture feature; and the low-level texture features are determined

based on the statistical texture features.

[0013] Texture enhancement and texture feature statistics are performed based on the low-level image features to obtain the low-level texture features.

[0014] In the above method, the operation that the low-level texture features are determined based on the enhanced texture features may include the following operation.

[0015] The enhanced texture features are determined as the low-level texture features;
or

[0016] the enhanced texture features are combined with the low-level image features to obtain the low-level texture features.

[0017] In the above method, the operation that the low-level texture features are determined based on the statistical texture features may include the following operation.

[0018] The statistical texture features are determined as the low-level texture features;
or

[0019] the statistical texture features are combined with the low-level image features to obtain the low-level texture features.

[0020] In the above method, the operation that texture enhancement and texture feature statistics are performed based on the low-level image features to obtain the low-level texture features may include the following operations.

[0021] Texture enhancement is performed based on the low-level image features to obtain the enhanced texture features.

[0022] At least texture feature statistics is performed based on the enhanced texture features to obtain the low-level texture features.

[0023] In the above method, the operation that at least texture feature statistics is performed based on the enhanced texture features to obtain the low-level texture features may include the following operations.

- [0024] The enhanced texture features are combined with the low-level image features, and then texture feature statistics is performed to obtain an intermediate low-level texture features. The intermediate low-level texture features are determined as the low-level texture features.
- [0025] In the above method, after the enhanced texture features are combined with the low-level image features, and then texture feature statistics is performed to obtain the intermediate low-level texture features, the method may also include the following operation.
- [0026] At least one of the enhanced texture features and the low-level image features is combined with the intermediate low-level texture features to obtain the low-level texture features.
- [0027] In the above method, the operation that texture enhancement is performed based on the low-level image features to obtain the enhanced texture features may include the following operations.
- [0028] One-dimensional quantization is performed on the low-level image features to obtain first coding features and a first initial quantization feature matrix.
- [0029] The enhanced texture features are obtained based on the first coding features and the first initial quantization feature matrix.
- [0030] In the above method, the operation that one-dimensional quantization is performed on the low-level image features to obtain the first coding features and the first initial quantization feature matrix may include the following operations.
- [0031] The low-level image features are pooled to obtain a first average feature corresponding to the low-level image features.
- [0032] A first similarity between the low-level image features and the first average feature is calculated to obtain a first similarity matrix.
- [0033] One-dimensional quantization is performed on the first similarity matrix to obtain a first quantization progression vector and the first coding features.

- [0034] The first coding features are fused with the first quantization progression vector to obtain the first initial quantization feature matrix.
- [0035] In the above method, the operation that one-dimensional quantization is performed on the first similarity matrix to obtain the first quantization progression vector and the first coding features may include the following operations.
- [0036] Dimensions of length and width of the first similarity matrix are merged to obtain a one-dimensional similarity vector.
- [0037] N-level quantization is performed on the one-dimensional similarity vector to obtain a first quantization eigenvalue, the first quantization eigenvalue forming the first quantization progression vector.
- [0038] The first coding features are determined based on the first quantization eigenvalue and the one-dimensional similarity vector. N is a positive integer greater than or equal to 1.
- [0039] In the above method, the first quantization eigenvalue includes: N sub quantization eigenvalues. The operation that N-level quantization is performed on the one-dimensional similarity vector to obtain the first quantization eigenvalue may include the following operations.
- [0040] N-level quantization is performed according to the maximum value and the minimum value in the one-dimensional similarity vector, and an average quantization value is determined.
- [0041] The sub quantization eigenvalue of each level is obtained based on the average quantization value and the minimum value, so as to obtain the N sub quantization eigenvalues.
- [0042] In the above method, the operation that the first coding features are fused with the first quantization progression vector to obtain the first initial quantization feature matrix may include the following operations.
- [0043] Spatial size averaging is performed on the first coding features to obtain a first

average coding feature.

[0044] The first average coding feature is spliced with the first quantization progression vector to obtain a first quantization statistics vector.

[0045] Spatial domain transformation is performed on the first quantization statistics vector, and the first quantization statistics vector is fused with the first average feature to obtain the first initial quantization feature matrix.

[0046] In the above method, the operation that the enhanced texture features are obtained based on the first coding features and the first initial quantization feature matrix may include the following operations.

[0047] Graph reasoning is performed on the first initial quantization feature matrix to obtain a first quantization feature matrix.

[0048] The enhanced texture features are obtained based on the first quantization feature matrix and the first coding features.

[0049] In the above method, when the low-level image features are input image features, target statistical texture features are the statistical texture features; or

[0050] when the enhanced texture features are the input image features, the target statistical texture features are the low-level texture features; or

[0051] when the features obtained by combining the enhanced texture features with the low-level image features are the input image features, the target statistical texture features are the intermediate low-level texture features.

[0052] In the above method, the operation that texture feature statistics is performed based on the input image features to obtain the target statistical texture features may include the following operations.

[0053] Two-dimensional quantization is performed on the input image features to obtain the target statistical texture features, the target statistical texture features representing a feature distribution relationship between pixels; or

- [0054] size segmentation is performed on the input image features to obtain local image features corresponding to the input image features; and two-dimensional quantization is performed on the local image features to obtain the target statistical texture feature; or
- [0055] two-dimensional quantization is performed on the input image features and the local image features respectively to obtain global quantization features corresponding to the input image features and local quantization features corresponding to the local image features.
- [0056] Feature fusion is performed based on the global quantization features and the local quantization features to obtain the target statistical texture features.
- [0057] In the above method, the operation that two-dimensional quantization is performed on the input image features to obtain the target statistical texture features may include the following operations.
- [0058] The input image features are pooled to obtain a second average feature corresponding to the input image features.
- [0059] A second similarity between the input image features and the second average feature is calculated to obtain a second similarity matrix.
- [0060] Two-dimensional quantization is performed on the second similarity matrix to obtain a second quantization series matrix and second coding features.
- [0061] The second coding features are fused with the second quantization series matrix to obtain a second initial quantization feature matrix.
- [0062] In the above method, the operation that two-dimensional quantization is performed on the second similarity matrix to obtain the second quantization series matrix and the second coding features may include the following operations.
- [0063] M-level quantization is performed on two dimensions of the second similarity matrix to obtain a second quantization eigenvalue, the second quantization eigenvalue forming the second quantization series matrix. M is a positive integer greater than or

equal to 1.

[0064] An intermediate coding feature matrix is determined based on the second quantization eigenvalue and the similarity matrix.

[0065] Based on the intermediate coding feature matrix, the current intermediate coding feature is multiplied by the transpose of its adjacent intermediate coding feature, and a coding feature corresponding to the current intermediate coding feature is determined, so as to obtain the second coding features.

[0066] In the above method, the operation that the second coding features are fused with the second quantization series matrix to obtain the second initial quantization feature matrix may include the following operations.

[0067] Spatial size averaging is performed on the second coding features to obtain a second average coding feature.

[0068] The second average coding feature is spliced with the second quantization progression vector to obtain a second quantization statistical matrix.

[0069] Spatial domain transformation is performed on the second quantization statistical matrix, and the second quantization statistical matrix is fused with the second average feature to obtain the second initial quantization feature matrix.

[0070] In the above method, the operation that feature fusion is performed based on the global quantization features and the local quantization features to obtain the target statistical texture features may include the following operations.

[0071] A global average feature is obtained by performing spatial transformation and averaging on the global quantization features,.

[0072] A local average feature is obtained by performing spatial transformation and averaging on the local quantization features,.

[0073] Feature fusion is performed on the global average feature and the local average feature to obtain the target statistical texture features.

- [0074] A second aspect of the embodiments of the present disclosure provides an image segmentation apparatus, which may include: a feature extraction part, a semantic segmentation part, a texture feature processing part, and a feature fusion part.
- [0075] The feature extraction part is configured to extract the low-level image features from the image to be processed.
- [0076] The semantic segmentation part is configured to perform semantic segmentation on the low-level image features to obtain the high-level semantic features.
- [0077] The texture feature processing part is configured to perform at least one of texture enhancement and texture feature statistics on the low-level image features to obtain the low-level texture features, the low-level texture features representing at least one of the details of the texture enhancement and the statistical distribution of texture features of the image to be processed.
- [0078] The feature fusion part is configured to obtain the semantic segmentation image by combining the high-level semantic features and the low-level texture features.
- [0079] In the above apparatus, the texture feature processing part is further configured to perform any one of the following:
- [0080] performing texture enhancement based on the low-level image features to obtain the enhanced texture feature; and determining the low-level texture features based on the enhanced texture feature;
- [0081] performing texture feature statistics based on the low-level image features to obtain the statistical texture feature; and determining the low-level texture features based on the statistical texture feature;
- [0082] performing texture enhancement and texture feature statistics based on the low-level image features to obtain the low-level texture features.
- [0083] In the above device, the texture feature processing part is further configured to determine the enhanced texture features as the low-level texture features, or combine the enhanced texture features with the low-level image features to obtain the low-level

texture features.

[0084] In the above device, the texture feature processing part is further configured to determine the statistical texture features as the low-level texture features, or combine the statistical texture features with the low-level image features to obtain the low-level texture features.

[0085] In the above device, the texture feature processing part is further configured to perform texture enhancement based on the low-level image features to obtain the enhanced texture features, and perform at least texture feature statistics based on the enhanced texture features to obtain the low-level texture features.

[0086] In the above device, the texture feature processing part is further configured to: combine the enhanced texture features with the low-level image features, and then perform texture feature statistics to obtain the intermediate low-level texture features; and determine the intermediate low-level texture features as the low-level texture features.

[0087] In the above device, the texture feature processing part is further configured to: combine the enhanced texture features with the low-level image features, and then perform texture feature statistics to obtain the intermediate low-level texture features; and after obtaining the intermediate low-level texture features, combine at least one of the enhanced texture features and the low-level image features with the intermediate low-level texture features to obtain the low-level texture features.

[0088] In the above device, the texture feature processing part is further configured to: perform one-dimensional quantization on the low-level image features to obtain the first coding features and the first initial quantization feature matrix; and obtain the enhanced texture features based on the first coding features and the first initial quantization feature matrix.

[0089] In the above device, the texture feature processing part is further configured to: pool the low-level image features to obtain the first average feature corresponding to the low-level image features; calculate the first similarity between the low-level image features and the first average feature to obtain the first similarity matrix; perform one-dimensional quantization on the first similarity matrix to obtain the first quantization

progression vector and the first coding feature; and fuse the first coding features with the first quantization progression vector to obtain the first initial quantization feature matrix.

[0090] In the above device, the texture feature processing part is further configured to: merge the dimensions of length and width of the first similarity matrix to obtain the one-dimensional similarity vector; perform N-level quantization on the one-dimensional similarity vector to obtain the first quantization eigenvalue, the first quantization eigenvalue forming the first quantization progression vector; and determine the first coding features based on the first quantization eigenvalue and the one-dimensional similarity vector. N is a positive integer greater than or equal to 1.

[0091] In the above device, the first quantization eigenvalue includes: N sub quantization eigenvalues. The texture feature processing part is further configured to: perform N-level quantization according to the maximum value and the minimum value in the one-dimensional similarity vector, and determine the average quantization value; and obtain the sub quantization eigenvalue of each level based on the average quantization value and the minimum value, so as to obtain the N sub quantization eigenvalues.

[0092] In the above device, the texture feature processing part is further configured to: perform spatial size averaging on the first coding features to obtain the first average coding feature; splice the first average coding feature with the first quantization progression vector to obtain the first quantization statistics vector; perform spatial domain transformation on the first quantization statistics vector, and fuse the first quantization statistics vector with the first average feature to obtain the first initial quantization feature matrix.

[0093] In the above device, the texture feature processing part is further configured to: perform graph reasoning on the first initial quantization feature matrix to obtain the first quantization feature matrix; and obtain the enhanced texture features based on the first quantization feature matrix and the first coding features.

[0094] In the above device, the texture feature processing part is further configured as: when the low-level image features are the input image features, the target statistical texture features are the statistical texture features; or when the enhanced texture features

are the input image features, the target statistical texture features are the low-level texture features; or when the features obtained by combining the enhanced texture features with the low-level image features are the input image features, the target statistical texture features are the intermediate low-level texture features.

[0095] In the above device, the texture feature processing part is further configured to: perform two-dimensional quantization on the input image features to obtain the target statistical texture features, the target statistical texture features representing the feature distribution relationship between pixels; or perform size segmentation on the input image features to obtain the local image features corresponding to the input image features, and perform two-dimensional quantization on the local image features to obtain the target statistical texture feature; or perform two-dimensional quantization on the input image features and the local image features respectively to obtain the global quantization features corresponding to the input image features and the local quantization features corresponding to the local image feature; and perform feature fusion based on the global quantization features and the local quantization features to obtain the target statistical texture features.

[0096] In the above device, the texture feature processing part is further configured to: pool the input image features to obtain the second average feature corresponding to the input image features; calculate the second similarity between the input image features and the second average feature to obtain the second similarity matrix; perform two-dimensional quantization on the second similarity matrix to obtain the second quantization series matrix and the second coding feature; and fuse the second coding features with the second quantization series matrix to obtain the second initial quantization feature matrix.

[0097] In the above device, the texture feature processing part is further configured to: perform M-level quantization on two dimensions of the second similarity matrix to obtain the second quantization eigenvalue, the second quantization eigenvalue forming the second quantization series matrix, and M being a positive integer greater than or equal to 1; determine the intermediate coding feature matrix based on the second quantization eigenvalue and the similarity matrix; and based on the intermediate coding

feature matrix, multiply the current intermediate coding feature by the transpose of its adjacent intermediate coding feature, and determine a coding feature corresponding to the current intermediate coding feature, so as to obtain the second coding features.

[0098] In the above device, the texture feature processing part is further configured to: perform spatial size averaging on the second coding features to obtain the second average coding feature; splice the second average coding feature with the second quantization progression vector to obtain the second quantization statistical matrix; and perform spatial domain transformation on the second quantization statistical matrix, and fuse the second quantization statistical matrix with the second average feature to obtain the second initial quantization feature matrix.

[0099] In the above device, the texture feature processing part is further configured to: after performing spatial transformation and averaging on the global quantization features, obtain the global average feature; after performing spatial transformation and averaging on the local quantization features, obtaining the local average feature; and perform feature fusion on the global average feature and the local average feature to obtain the target statistical texture features.

[00100] A third aspect of the embodiments of the present disclosure provides an electronic device, which may include: a memory and a processor.

[00101] The memory is configured to store executable instructions.

[00102] The processor is configured to implement, when executing the executable instructions stored in the memory, the image segmentation method provided by the embodiments of the present disclosure.

[00103] A fourth aspect of the embodiments of the present disclosure provides a computer-readable storage medium, in which an executable instruction is stored. The executable instruction is configured to implement, when executed by the processor, the image segmentation method provided by the embodiments of the present disclosure.

[00104] A fifth aspect of the embodiments of the present disclosure provides a computer program, which includes a computer-readable code. When the computer-

readable code runs in the electronic device, the processor in the electronic device implements the steps in the first aspect.

[00105] In the above solutions, by obtaining the low-level image features in the image and performing at least one of texture enhancement and texture feature statistics on the low-level image features, the obtained low-level texture features can include multi-scale deep low-level texture information in the image to be processed, representing at least one of the details of the texture enhancement and the statistical distribution of texture features of the image to be processed. In this way, by combining the low-level texture features with the high-level semantic features in image segmentation, texture detail information of the image to be processed may fully participate in a network inference process of semantic segmentation, thereby improving the accuracy of segmentation result information such as the boundary, texture and structure of a segmented region, and then improving the accuracy of image segmentation.

BRIEF DESCRIPTION OF THE DRAWINGS

[00106] FIG. 1 is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure.

[00107] FIG. 2A is an effect diagram of local structure features of image texture in street view image 1 provided by an embodiment of the present disclosure.

[00108] FIG. 2B is an effect diagram of performing equalization enhancement on street view image 1 provided by an embodiment of the present disclosure.

[00109] FIG. 3 is a structural schematic diagram of a Statistical Texture Learning Network (STLNet) provided by an embodiment of the present disclosure.

[00110] FIG. 4 is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure.

[00111] FIG. 5 is an optional structural schematic diagram of a STLNet provided by an embodiment of the present disclosure.

- [00112] FIG. 6 is an optional structural schematic diagram of a STLNet provided by an embodiment of the present disclosure.
- [00113] FIG. 7 is an optional structural schematic diagram of a STLNet provided by an embodiment of the present disclosure.
- [00114] FIG. 8 is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure.
- [00115] FIG. 9 is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure.
- [00116] FIG. 10 is an optional structural schematic diagram of a 1d-Quantization and Counting Operator (1d-QCO) provided by an embodiment of the present disclosure.
- [00117] FIG. 11A is an optional structural schematic diagram of a Texture Enhancement Module (TEM) provided by an embodiment of the present disclosure.
- [00118] FIG. 11B is an optional structural schematic diagram of a TEM provided by an embodiment of the present disclosure.
- [00119] FIG. 12 is an optional structural schematic diagram of a Pyramid Texture Feature Extraction Module (PTFEM) provided by an embodiment of the present disclosure.
- [00120] FIG. 13 is a schematic diagram of a comparison between the effects before and after texture feature statistics is performed on an original image provided by an embodiment of the present disclosure.
- [00121] FIG. 14 is a schematic diagram of a comparison between the effect of image segmentation of the image segmentation method provided by an embodiment of the present disclosure and the effect of image segmentation based on the neural network DeepLabV3.
- [00122] FIG. 15 is a structural schematic diagram of an image segmentation apparatus provided by an embodiment of the present disclosure.

[00123] FIG. 16 is a structural schematic diagram of an electronic device provided by an embodiment of the present disclosure.

DETAILED DESCRIPTION

[00124] For making the objectives, technical solutions and advantages of the present disclosure clearer, the present disclosure will further be described below in combination with the drawings in detail. The described embodiments should not be considered as limits to the present disclosure. All other embodiments obtained by those of ordinary skill in the art without creative work shall fall within the scope of protection of the present disclosure.

[00125] "Some embodiments" involved in the following descriptions describes a subset of all possible embodiments. However, it can be understood that "some embodiments" may be the same subset or different subsets of all the possible embodiments, and may be combined without conflicts.

[00126] Term "first/second/third" involved in the following descriptions is only for distinguishing similar objects and does not represent a specific sequence of the objects. It can be understood that "first/second/third" may be interchanged to specific sequences or orders if allowed to implement the embodiments of the present disclosure described herein in sequences except the illustrated or described ones.

[00127] Unless otherwise defined, all technical and scientific terms in the specification have the same meaning as those skilled in the art, belonging to the present disclosure, usually understand. Terms used in the specification are only used for describing the purpose of the embodiments of the present disclosure, but not intended to limit the present disclosure.

[00128] In recent years, with the rapid development of deep learning, semantic segmentation of images using a neural network has achieved great success. Most modern semantic segmentation models use, based on an FCN, a convolutional layer instead of a fully connected layer in a common classification network to obtain a pixel-level

prediction result. Moreover, on this basis, a variety of methods for improving the basic FCN have evolved. Some methods use a pyramid structure to obtain information from receptive fields of different sizes. Exemplarily, a Pyramid Scene Parsing Network (PSPNet) inputs features in pyramid merge layers with different merge ratios. The DeeplabV3 series of semantic segmentation models propose an atrous spatial pyramid merge layer consisting of multiple dilated convolutions with different dilation rates. However, a convolution operator in the FCN is sensitive to a local change, and the convolution operator may be used to mine local features such as the boundary and other image information, but cannot be used to describe or make statistics on an image texture. Accordingly, many methods have been proposed to extract and utilize low-level features, for example, the low-level features are transferred to a high-level network layer through a skip connection technology. However, a simple multi-level feature adding or splicing operation may lead to a problem of feature dislocation, thus reducing the effectiveness of the low-level features. To sum up, texture information of an image cannot be used effectively in the existing image segmentation technology, resulting in the inaccuracy of boundary and texture information of various objects in an image segmentation result, and thus reducing the accuracy of image segmentation.

[00129] A method for image segmentation provided by an embodiment of the present disclosure will be described in combination with the exemplary application and implementation of an electronic device provided by an embodiment of the present disclosure.

[00130] Referring to FIG. 1 which is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure, the image segmentation method will be described in combination with the steps illustrated in FIG. 1.

[00131] At S101, low-level image features are extracted from an image to be processed, and semantic segmentation is performed on the low-level image features to obtain high-level semantic features.

[00132] In the embodiments of the present disclosure, the electronic device may use a feature extraction method to perform feature extraction on pixels included in the image to be processed, so as to obtain the low-level image features.

[00133] In some embodiments, the electronic device may use a multi-level feature extraction layer in ResNet101, namely the backbone network of a neural network, to perform multi-scale image feature extraction on the image to be processed, and take low-level outputs, such as the features output by the first layer or the second layer, in the multi-level feature extraction layer as the low-level image features. Other types of feature extraction networks or feature extraction methods may also be used to extract the low-level image features from the image to be processed, which depends on the actual situation, and is not limited in the embodiments of the present disclosure.

[00134] Here, the low-level image features represent appearance features of the image to be processed, such as color, boundary division, contrast and brightness, which is not limited in the embodiments of the present disclosure.

[00135] In the embodiments of the present disclosure, the electronic device may further extract high-level features based on the low-level image features to obtain high-level semantic information of the image to be processed. Exemplarily, the low-level image features are input in a high-level extraction layer in the feature extraction network to extract the high-level semantic information, such as whether the image to be processed includes vehicle, pedestrian, road and other information. The electronic device performs semantic prediction on a semantic category of each pixel according to the extracted high-level semantic information, and performs semantic segmentation on the image to be processed according to a result of semantic prediction to obtain the high-level semantic features.

[00136] In some embodiments, the electronic device may perform semantic segmentation on the low-level image features through a semantic segmentation network. Exemplarily, the semantic segmentation network may be a convolutional pooling layer of the backbone network ResNet101, an Atrous Spatial Pyramid Pooling (ASPP) network, or other types of semantic segmentation networks, which depends on the actual situation, and is not limited in the embodiments of the present disclosure.

[00137] At S102, at least one of texture enhancement and texture feature statistics is performed on the low-level image features to obtain low-level texture features, the low-level texture features representing details of the texture enhancement and/or statistical

distribution of texture features of the image to be processed.

[00138] In the embodiments of the present disclosure, the image texture in digital image processing has not only local structural features, but also global statistical features. Here, the local structural features may be local pattern information, such as boundary, smoothness and roughness, and may be extracted by a low-level filter in Convolutional Neural Networks (CNN). FIG. 2(a) is an effect diagram of extracting street view image 1 by low-level filter 1 to obtain local structure features of the image texture in the street view image 1. It can be seen that a local structure feature map of FIG. 2(a) reflects the texture, boundary, object structure and other information in the street view image 1. The global statistical features may be obtained by performing statistics of low-level image information (such as a pixel value or a local region attribute) on the image. In some embodiments, the global statistical features may be calculated through a gray histogram for representing the distribution of local structural features. FIG. 2(b) is an effect diagram of calculating the global statistical features of the image texture in the street view image 1 through the gray histogram, and performing equalization enhancement on the street view image 1 according to the global statistical features. It can be seen that the original street view image 1 taken in a dark environment often has poor visual quality. After the step of performing equalization enhancement through the histogram, the image segmentation is more detailed and the segmentation effect is better.

[00139] Based on the effect diagrams illustrated in FIG. 2(a) and FIG. 2(b), it can be seen that adding texture feature information to an image segmentation task may further optimize the boundary and texture information of a segmentation object region and improve the accuracy of image segmentation.

[00140] Generally, because the low-level image features extracted from the image to be processed through a low-level feature extraction layer of the backbone network are often of low quality, especially the low contrast, texture details may be blurred, affecting the extraction and utilization of low-level information. Therefore, the embodiments of the present disclosure may analyze and mine the low-level image features and enhance the texture details of the low-level image features, so as to achieve texture enhancement of the low-level image features and obtain the low-level texture features representing the

details of the texture enhancement of the image to be processed, so that in the next steps of the method, it is easier to capture the image information related to the texture, and the accuracy of image segmentation is improved using the image information related to the texture.

[00141] In the embodiments of the present disclosure, because the image texture is highly correlated with the statistical information of a spatial relationship between pixels, in order to describe the spatial distribution of the image texture between the pixels, the electronic device may use the texture information in the low-level image features to extract the statistical features of the texture information on multiple scales, so as to achieve texture feature statistics of the low-level image features and obtain the low-level texture features representing the statistical distribution of texture features of the image to be processed.

[00142] In the embodiments of the present disclosure, the electronic device may obtain the low-level texture features used for representing the details of the texture enhancement and/or the statistical distribution of texture features of the image to be processed by performing at least one of texture enhancement and texture feature statistics on the low-level image features.

[00143] At S103, a segmented image is obtained by combining the high-level semantic features and the low-level texture features.

[00144] In the embodiments of the present disclosure, the electronic device may fuse the high-level semantic features obtained by performing semantic segmentation on the image to be processed with the low-level texture features, and analyze synthetically each pixel in the image to be processed according to the fused image features. During the analysis, the electronic device determines a corresponding segmentation result of the pixel in a segmented image by combining a semantic segmentation category to which each pixel belongs and its corresponding low-level texture features, and obtains the segmentation result corresponding to each pixel. Then the electronic device may obtain the segmented image corresponding to the image to be processed based on the segmentation result corresponding to each pixel.

[00145] In some embodiments, when the image to be processed passes through the feature extraction layer, the image size is reduced to a preset output size required by the feature extraction layer. The electronic device performs at least one of semantic segmentation, texture enhancement and texture feature statistics based on the reduced image to be processed, and then obtains the segmentation result corresponding to each pixel in the reduced image. In this case, before applying the segmentation result corresponding to each pixel to the final segmented image, the electronic device performs up-sampling on the reduced image, restores it to the size of the original image to be processed, and finally obtains the segmented image with the same size as the image to be processed.

[00146] Understandably, because the electronic device takes into account both the high-level semantic features and the low-level texture features in the final image segmentation, the boundaries and textures of various image objects in the segmented image are nearly consistent with that in the image to be processed, thereby making the result of image segmentation more accurate.

[00147] In some embodiments, a STLNet provided by the embodiments of the present disclosure is used to implement the image segmentation method provided by the embodiments of the present disclosure. As illustrated in FIG. 3, a network structure of the STLNet may be divided into two parts: backbone network branch 30 and texture feature extraction module 31. The backbone network branch may be network ResNet-101 or any neural network with a semantic segmentation function, which depends on the actual situation, and is not limited in the embodiments of the present disclosure.

[00148] As illustrated in FIG. 3, after inputting the image to be processed in the STLNet, the electronic device performs multi-scale image extraction on the image to be processed through a feature extraction layer 301 of the backbone network branch 30. The electronic device merges the low-level features output by the first two layers of the feature extraction layer 301 as the low-level image features, and inputs the high-level features finally output by the feature extraction layer 301 in the semantic segmentation network 302. The semantic segmentation network 302 extracts and analyzes high-level contextual features based on the high-level features to achieve semantic segmentation

and obtain high-level image features. Here, the semantic segmentation network may be the ASPP network. The electronic device inputs the low-level image features in the texture feature extraction module 31, and the texture feature extraction module 31 performs at least one of texture enhancement and texture feature statistics on the low-level image features to obtain the low-level texture features. Then the electronic device combines the high-level semantic features output by the semantic segmentation network 302 with the low-level texture features output by the texture feature extraction module 31, and restores the image size through an up-sampling process to obtain the final segmented image.

[00149] It is understandable that, in the embodiments of the present disclosure, the electronic device performs at least one of texture enhancement and texture feature statistics on the low-level image features of the image to be processed, which can enhance the texture details of the image to be processed, mine multi-scale deep texture information in the image to be processed, and obtain the low-level texture features that can represent the details of the texture enhancement and/or the statistical distribution of texture features. The electronic device combines the low-level texture features with the high-level semantic features to obtain the final segmented image, which can better restore the low-level texture information of the original image in the segmented image, thus improving the accuracy of image segmentation.

[00150] In the embodiments of the present disclosure, referring to FIG. 4 which is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure. S102 in FIG. 1 may be implemented through any process in S201 to S202, or S301 to S302, or S401, and will be described in combination with the steps.

[00151] At S201, texture enhancement is performed based on the low-level image features to obtain the enhanced texture features.

[00152] At S202, the low-level texture features are determined based on the enhanced texture features.

[00153] In the embodiments of the present disclosure, the electronic device may perform texture enhancement on the low-level image features and determine the low-

level texture features that can be combined with the high-level semantic features based on the obtained enhanced texture features.

[00154] In some embodiments, S202 in FIG. 4 may be implemented by executing S2021 or S2022, and will be described in combination with the steps.

[00155] At S2021, the enhanced texture features are determined as the low-level texture features.

[00156] In the embodiments of the present disclosure, the electronic device may directly take the enhanced texture features as the low-level texture features combined with the high-level semantic features.

[00157] At S2022, the enhanced texture features are combined with the low-level image features to obtain the low-level texture features.

[00158] In the embodiments of the present disclosure, the electronic device may perform feature fusion on the enhanced texture features obtained by performing texture enhancement on the low-level image features and the original low-level image features obtained by feature extraction to obtain the low-level texture features.

[00159] Here, by combining the enhanced texture features with the low-level image features, a texture enhancement effect of the low-level texture features can be further improved.

[00160] In some embodiments, as illustrated in FIG. 5, the texture feature processing module 31 in FIG. 3 may be a TEM 31_1. The electronic device may directly take the enhanced texture features output by the TEM as the low-level texture features, and may also combine the enhanced texture features output by the TEM with the low-level image features as the low-level texture features, as shown by the dotted arrow in FIG. 5.

[00161] At S301, texture feature statistics is performed based on the low-level image features to obtain statistical texture features.

[00162] At S302, the low-level texture features are determined based on the statistical texture features.

- [00163] In the embodiments of the present disclosure, the electronic device may perform texture feature statistics on the low-level image features, and take the obtained statistical texture features as the low-level texture features to be further combined with the high-level semantic features, so as to obtain the segmented image.
- [00164] In some embodiments, S302 in FIG. 4 may be implemented by executing S3021 or S3022, and will be described in combination with the steps.
- [00165] At S3021, the statistical texture features are determined as the low-level texture features.
- [00166] In the embodiments of the present disclosure, the electronic device may directly take the statistical texture features as the low-level texture features combined with the high-level semantic features.
- [00167] At S3022, the statistical texture features are combined with the low-level image features to obtain the low-level texture features.
- [00168] In the embodiments of the present disclosure, the electronic device may perform feature fusion on the statistical texture features obtained by performing texture feature statistics on the low-level image features and the original low-level image features obtained by feature extraction to obtain the low-level texture features.
- [00169] Here, by combining the statistical texture features with the low-level image features, the accuracy of the statistical distribution of texture features represented by the low-level texture features can be further improved.
- [00170] In some embodiments, as illustrated in FIG. 6, the texture feature processing module 31 in FIG. 3 may be a PTFEM 31_2. The electronic device may directly take the statistical texture features output by the PTFEM as the low-level texture features, and may also combine the statistical texture features output by the PTFEM with the low-level image features as the low-level texture features, as shown by the dotted arrow in FIG. 6.
- [00171] At S401, texture enhancement and texture feature statistics are performed based on the low-level image features to obtain the low-level texture features.

- [00172] In some embodiments, S401 in FIG. 4 may be implemented by executing S4011 or S4012, and will be described in combination with the steps.
- [00173] At S4011, texture enhancement is performed based on the low-level image features to obtain the enhanced texture features.
- [00174] In the embodiments of the present disclosure, the electronic device first performs texture enhancement based on the low-level image features to obtain the enhanced texture features.
- [00175] At S4012, texture feature statistics is at least performed based on the enhanced texture features to obtain the low-level texture features.
- [00176] In the embodiments of the present disclosure, S4012 may be implemented by executing S4012A or through any process in S4012B to S4012C, and will be described in combination with the steps.
- [00177] At S4012A, texture feature statistics is performed based on the enhanced texture features to obtain the low-level texture features.
- [00178] In S4012A, the electronic device further performs texture feature statistics on the enhanced texture features obtained in the process of texture enhancement to obtain the low-level texture features.
- [00179] At S4012B, the enhanced texture features and the low-level image features are combined, and then texture feature statistics is performed to obtain intermediate low-level texture features.
- [00180] At S4012C, the intermediate low-level texture features are determined as the low-level texture features.
- [00181] In S4012B to S4012C, the electronic device first performs feature fusion on the enhanced texture features and the low-level image features, and performs the texture feature statistics on the combined features to obtain the intermediate low-level texture features. The electronic device may take the intermediate low-level texture features as the low-level texture features.

- [00182] In some embodiments, after S4012A, S4012D is included as follows.
- [00183] At S4012D, at least one of the enhanced texture features and the low-level image features is combined with the intermediate low-level texture features to obtain the low-level texture features.
- [00184] In S4012D, the electronic device performs feature fusion on intermediate low-level features and at least one of the enhanced texture features and the low-level image features to obtain the low-level texture features.
- [00185] In some embodiments, as illustrated in FIG. 7, the output of the TEM 31_1 is the enhanced texture features. The electronic device may directly input the enhanced texture features in the texture feature statistics module 31_2 through the method of S4012A, and then output the low-level texture features through the texture feature statistics module 31_2. Alternatively, the electronic device may also first combine the enhanced texture features with the low-level image features output by the first two layers of the feature extraction layer 301 through the method of S4012B to S4012C, and then input the combined features in the texture feature statistics module 31_2; then, the texture feature statistics module 31_2 performs texture feature statistics on the combined features, outputs the intermediate low-level texture features, and takes the intermediate low-level texture features as the low-level texture features. Alternatively, after the texture feature statistics module 31_2 outputs the intermediate low-level texture features, the electronic device fuses, through the method of S4012D, the enhanced texture features output by the TEM 31_1 and at least one of the low-level image features output by the first two layers of the feature extraction layer 301 with the intermediate low-level texture features to obtain the low-level image features.
- [00186] It is understandable that, in the embodiments of the present disclosure, when the electronic device performs texture enhancement on the low-level image features, the contrast, brightness and other information of the low-level image features can be enhanced to make the texture information clearer, and then by combining the low-level texture features obtained by performing texture enhancement with the high-level semantic features for semantic segmentation, the boundary, structure, texture and other information of each segmented region can be accurately reflected in the segmentation

result, thereby improving the accuracy of semantic segmentation. When the electronic device performs texture feature statistics on the low-level image features, the low-level texture features for representing the distribution of texture information can be obtained by mining the statistical features of the texture information, and then by combining the low-level texture features obtained by performing texture feature statistics with the high-level semantic features for semantic segmentation, the correlation of the texture information among the pixels in the image to be processed can be restored more accurately, thereby improving the accuracy of image segmentation. The electronic device may also perform both texture enhancement and texture feature statistics on the low-level image features. Combining the advantages of two processing modes, the accuracy of image segmentation is further improved.

[00187] In some embodiments, referring to FIG. 8 which is an optional flowchart of a method for image segmentation provided by an embodiment of the present disclosure, S201 in FIG. 4 may be implemented by executing S2011 and S2012, and corresponding to the processing flow of the TEM 31_1 in FIG. 5 or FIG. 7, will be described in combination with the steps.

[00188] At S2011, one-dimensional quantization is performed on the low-level image features to obtain a first coding features and a first initial quantization feature matrix.

[00189] In the embodiments of the present disclosure, the distribution of the low-level image features of the image to be processed in a spectral domain is usually continuous and varies widely, so it is difficult to extract and optimize the low-level image features in deep neural networks. Therefore, the electronic device first performs once quantization on the low-level image features to quantize the continuous low-level image features into scattered eigenvalues and obtain the first coding features and the first initial quantization feature matrix respectively.

[00190] The first coding features represent a quantization coding image corresponding to the low-level image features, and the first initial quantization feature matrix represents the statistical features after the low-level image features are quantized and scattered.

[00191] In some embodiments, referring to FIG. 9 which is an optional flowchart of a

method for image segmentation provided by an embodiment of the present disclosure, S2011 may be implemented by executing the process of S2011A to S2011D, and will be described in combination with the steps.

[00192] At S2011A, the low-level image features are pooled to obtain a first average feature corresponding to the low-level image features.

[00193] In S2011A, the electronic device calculates the average of the low-level image features by pooling and obtains the first average feature corresponding to the low-level image features.

[00194] In the embodiments of the present disclosure, the low-level image features are a feature map obtained by performing multi-scale feature extraction on the whole image to be processed. Exemplarily, the low-level image features may be represented by a three-dimensional feature matrix of $C \times H \times W$, where H and W represent the height and width of the feature matrix respectively, and C represents the feature dimension of the low-level image features.

[00195] In some embodiments, the electronic device may pool the low-level image features through a global pooling layer, and average the three-dimensional feature matrix into a one-dimensional average feature vector as the first average feature. The first average feature may be a matrix of $C \times 1 \times 1$.

[00196] At S2011B, a first similarity between the low-level image features and the first average feature is calculated to obtain a first similarity matrix.

[00197] In S2011B, for each position in the feature matrix of the low-level image features, the electronic device may calculate the first similarity between the low-level feature at each position and the first average feature to obtain a similarity value at each position, and then obtain the first similarity matrix.

[00198] In some embodiments, the electronic device may calculate a cosine similarity between the low-level image features and the first average feature as the first similarity.

[00199] Because the similarity value is a one-dimensional value, the first similarity

matrix may be a two-dimensional matrix of $H \times W$. In some embodiments, the first similarity between the low-level image features and the first average feature may be calculated by formula (1) to obtain the first similarity matrix, as follows:

$$[00200] \quad S_{ij} = \frac{\mathbf{g} \cdot \mathbf{A}_{i,j}}{\|\mathbf{g}\|_2 \cdot \|\mathbf{A}_{i,j}\|_2} \quad (1),$$

$$[00201] \quad \text{where } \mathbf{A} \in \mathbb{R}^{C \times H \times W}, \mathbf{g} \in \mathbb{R}^{C \times 1 \times 1}, \mathbf{A}_{i,j} (i \in [1, W], j \in [1, H]), \mathbf{S} \in \mathbb{R}^{1 \times H \times W}.$$

[00202] In formula (1), $\mathbf{A}$ is the feature matrix of $C \times H \times W$ corresponding to the low-level image features, $\mathbf{A}_{i,j}$ is the low-level feature at each position in the feature matrix. The value of i ranges from 1 to W , where W is the width of the feature matrix; the value of j ranges from 1 to H , where H is the height of the feature matrix. $\mathbf{g}$ is the first average feature. The electronic device may perform inner product computation on $\mathbf{A}_{i,j}$ at each position in the feature matrix and the first average feature $\mathbf{g}$ through the formula (1) to obtain S_{ij} corresponding to each $\mathbf{A}_{i,j}$. S_{ij} represents the first similarity corresponding to position (i, j) . Then the electronic device may construct the first similarity matrix $\mathbf{S}$ through S_{ij} , where $\mathbf{S}$ is the two-dimensional matrix of $H \times W$.

[00203] At S2011C, one-dimensional quantization is performed on the first similarity matrix to obtain a first quantization progression vector and the first coding features.

[00204] In S2011C, the electronic device may perform one-dimensional quantization on the first similarity matrix based on the dimension of values of the first similarity in the first similarity matrix, at the same time, determine the first quantization progression vector used by the one-dimensional quantization, and obtain the first coding features after the one-dimensional quantization is completed.

[00205] The first quantization progression vector is the vector formed by the quantization values corresponding to each of preset first quantization series corresponding to the one-dimensional quantization.

- [00206] In some embodiments, S2011C may be implemented by executing S2011C_1 to S2011C_3, and will be described in combination with the steps.
- [00207] At S2011C_1, dimensions of length and width of the first similarity matrix are merged to obtain a one-dimensional similarity vector.
- [00208] In S2011C_1, the electronic device deforms the first similarity matrix, and merges the length dimension and the width dimension of the first similarity matrix to obtain the one-dimensional similarity vector.
- [00209] In some embodiments, the one-dimensional similarity vector may be a matrix of $1 \times HW$, $\square^{HW}$.
- [00210] At S2011C_2, N-level quantization is performed on the one-dimensional similarity vector to obtain a first quantization eigenvalue, the first quantization eigenvalue forming the first quantization progression vector. N is a positive integer greater than or equal to 1.
- [00211] In S2011C_2, N represents the preset quantization series of quantization, and N is a positive integer greater than or equal to 1. When performing N-level quantization on the one-dimensional similarity vector, the electronic device determines the quantization value corresponding to each level of quantization simultaneously as the first quantization eigenvalue. The first quantization eigenvalue forms the first quantization progression vector.
- [00212] In some embodiments, for the N-level quantization, the first quantization eigenvalue may include N sub quantization eigenvalues. Each sub quantization eigenvalue represents the quantization value corresponding to each level of the N-level quantization. The process of determining the quantization value in S2011C_2 may be implemented by formula (2), that is, the N-level quantization is performed according to the maximum value and the minimum value in the one-dimensional similarity vector to determine the average quantization value; and the sub quantization eigenvalue of each level is obtained based on the average quantization value and the minimum value, and then N sub quantization eigenvalues are obtained; as follows:

[00213]
$$L_n = \frac{\max(S) - \min(S)}{N} \cdot n + \min(S) \quad (2).$$

[00214] In formula (2), $\max(S)$ is the maximum value in the one-dimensional similarity vector, $\min(S)$ is the minimum value in the one-dimensional similarity vector, the calculated result of $\frac{\max(S) - \min(S)}{N}$ is the average quantization value, n indicates each quantization series, ranging from 1 to N . L_n is the sub quantization eigenvalue corresponding to each quantization series. By formula (2), the electronic device may obtain N sub quantization eigenvalues L_n , that is, the first quantization eigenvalue L_n . Then the electronic device may construct the first quantization progression vector L through the first quantization eigenvalue L_n , where $L = [L_1, L_2, \dots, L_n, \dots, L_N]$. Then the electronic device performs the quantization of each level of the N -level quantization on the one-dimensional similarity vector according to each sub quantization eigenvalue L_n in L .

[00215] At S2011C_3, the first coding features are determined based on the first quantization eigenvalue and the one-dimensional similarity vector.

[00216] In S2011C_3, by formula (3), the electronic device may determine the first coding features based on the first quantization eigenvalue and the one-dimensional similarity vector, as follows:

[00217]
$$E_{i',n} = \begin{cases} 1 - |L_n - S'_{i'}| & \text{if } -\frac{0.5}{N} \leq L_n - S'_{i'} < \frac{0.5}{N} \\ 0 & \text{else} \end{cases} \quad (3).$$

[00218] In formula (3), $S'_{i'}$ is a one-dimensional similarity value obtained by merging the length dimension and the width dimension of the first similarity matrix S , and performing matrix dimensional transformation on each first similarity S_{ij} in S . Here, $i' \in [1, HW]$, $S'_{i'}$ may form the one-dimensional similarity vector S' , $S' \in \mathbb{R}^{HW}$. By

formula (3), the electronic device may use the N sub quantization eigenvalues L_n in the first quantization progression vector to perform N-level quantization on each one-dimensional similarity value S'_i in the one-dimensional similarity vector, and obtain the first coding features $E_{i,n}$ of the dimension N×HW.

[00219] At S2011D, the first coding features are fused with the first quantization progression vector to obtain the first initial quantization feature matrix.

[00220] In S2011D, the electronic device fuses the first coding features with the first quantization progression vector to obtain the first initial quantization feature matrix for representing the low-level image features after quantization.

[00221] In some embodiments, by formula (4), the electronic device may perform spatial size averaging on the first coding features to obtain a first average coding feature, and splice the first average coding feature with the first quantization progression vector to obtain a first quantization statistics vector, as follows:

[00222]
$$C = Cat\left(L, \frac{\sum_{i=1}^{HW} E_{i,n}}{\sum_{n=1}^N \sum_{i=1}^{HW} E_{i,n}}\right) \quad (4).$$

[00223] In formula (4), $\frac{\sum_{i=1}^{HW} E_{i,n}}{\sum_{n=1}^N \sum_{i=1}^{HW} E_{i,n}}$ is the process of performing spatial size

averaging on the first coding features $E_{i,n}$, and the calculated result of $\frac{\sum_{i=1}^{HW} E_{i,n}}{\sum_{n=1}^N \sum_{i=1}^{HW} E_{i,n}}$

is the first average coding feature. $Cat\left(L, \frac{\sum_{i=1}^{HW} E_{i,n}}{\sum_{n=1}^N \sum_{i=1}^{HW} E_{i,n}}\right)$ represents that the first

average coding feature is spliced with the first quantization progression vector L, and C represents the first quantization statistics vector.

[00224] Then the electronic device performs spatial domain transformation on the first quantization statistics vector C by formula (5), and fuses the first quantization statistics vector with the first average feature to obtain the first initial quantization feature matrix.

Formula (5) is as follows:

[00225]
$$D = Cat(MLP(C), g) \quad (5).$$

[00226] In formula (5), MLP represents the process of spatial domain transformation. The electronic device may transform the first quantization statistics vector C to the same feature space of the first average feature g by formula (5) for feature fusion, and obtain the first initial quantization feature matrix D .

[00227] In some embodiments, before fusing the first average feature g with the first quantization statistics vector C after spatial domain transformation, the electronic device may also perform up-sampling on the first average feature g ; or the electronic device may also directly take the first quantization statistics vector C obtained by formula (4) as the first initial quantization feature matrix, which depends on the actual situation, and is not limited in the embodiments of the present disclosure.

[00228] In some embodiments, referring to FIG. 10 which is an optional structural schematic diagram of a 1d-QCO provided by an embodiment of the present disclosure. The electronic device may perform one-dimensional quantization on the low-level image features through the 1d-QCO to obtain the first coding features and the first initial quantization feature matrix. As illustrated in FIG. 10, the low-level image features A may be pooled through global pooling in the 1d-QCO to obtain the first average feature corresponding to A . The 1d-QCO calculates the cosine similarity between A and the first average feature to obtain the first similarity matrix S . S is a two-dimensional matrix of $1 \times H \times W$. The 1d-QCO merges the length and the dimension of the first similarity matrix S , and transforms S into a one-dimensional similarity vector of $1 \times HW$. The 1d-QCO performs the N -level quantization on the one-dimensional similarity vector according to N first quantization eigenvalues L_1 to L_N corresponding to the N -level quantization to obtain the first coding feature E of the dimension $N \times HW$. The 1d-QCO performs spatial size averaging on each line in the first coding feature E to obtain N first average coding features. The 1d-QCO splices the N first average coding features with N first quantization eigenvalues to obtain the first quantization statistics vector C of the dimension $N \times 2$. The 1d-QCO performs spatial domain transformation on the first

quantization statistics vector C , and fuses the first quantization statistics vector with the first average feature after up-sampling to obtain the first initial quantization feature matrix D .

[00229] At S2012, the enhanced texture features are obtained based on the first coding features and the first initial quantization feature matrix.

[00230] In the embodiments of the present disclosure, when performing the one-dimensional quantization on the low-level image features to obtain the first coding features and the first initial quantization feature matrix, the electronic device may fuse the first coding features with the first initial quantization feature matrix, for example, performing matrix multiplication or matrix splicing, and take a fusion result as the enhanced texture features.

[00231] In some embodiments, FIG. 11(a) is an optional structural schematic diagram of a TEM provided by an embodiment of the present disclosure. As illustrated in FIG. 11(a), the TEM may include a 1d-QCO and a fusion module. Exemplarily, the structure of the 1d-QCO may be illustrated in FIG. 10. The TEM may perform, through the fusion module, matrix multiplication on the first coding feature E , namely the matrix of $N \times HW$ in FIG. 11(a), output by the 1d-QCO and the first initial quantization feature matrix D , namely the matrix of $C_1 \times N$ in FIG. 11(a), output by the 1d-QCO to obtain the output of the TEM, that is, the enhanced texture features and the matrix of $C_1 \times HW$.

[00232] In some embodiments of the present disclosure, in order to improve the balance of the distribution of quantization results, the electronic device may also perform quadratic quantization optimization on the first initial quantization feature matrix D obtained from the one-dimensional quantization to obtain the first quantization feature matrix D' , and then fuse the first quantization feature matrix D' with the first coding features to obtain the enhanced texture features.

[00233] In some embodiments, quadratic quantization optimization may be implemented by graph reasoning. S2012 may be implemented by executing the process of S2012A to S2012B, and will be described in combination with the steps.

[00234] At S2012A, graph reasoning is performed on the first initial quantization feature matrix to obtain the first quantization feature matrix.

[00235] In S2012A, the process of graph reasoning needs to construct nodes in a graph model and edges between the nodes, and then a graph reasoning algorithm is implemented based on the graph model. The electronic device may take an initial quantization feature value at each quantization level in the first initial quantization feature matrix as the vertex to construct the graph model. Exemplarily, when the first initial quantization feature matrix is $\square^{C_1 \times N}$, the vertex of the graph model may be the values of C_1 corresponding to each of the N quantization levels. The adjacent relation between the vertexes may be obtained by formula (6), as follows:

$$[00236] \quad X = \text{Softmax}(\phi_1(D)^T \cdot \phi_2(D)) \quad (6).$$

[00237] In formula (6), ϕ_1 and ϕ_2 represent two convolutional layers of 1×1 , and D is the first initial quantization feature matrix. In some embodiments, Softmax is a nonlinear normalized function. X is an adjacency matrix, which is used for representing the adjacent relation between the vertexes of the graph model.

[00238] Then the electronic device updates each initial node by fusing the features of all other vertexes by formula (7) to obtain the reconstructed first quantization feature matrix $D' \in \square^{C_2 \times N}$, as follows:

$$[00239] \quad D' = \phi_3(D) \cdot X \quad (7).$$

[00240] In formula (7), ϕ_3 represents another convolutional layer of 1×1 except ϕ_1 and ϕ_2 .

[00241] At S2012B, the enhanced texture features are obtained based on the first quantization feature matrix and the first coding features.

[00242] In S2012B, the electronic device multiplies the first quantization feature matrix D' obtained from quadratic quantization optimization by E to obtain the enhanced

texture features R , as shown in formula (8):

[00243]
$$R = D' \cdot E^T \quad (8).$$

[00244] In formula (8), E^T is the transpose of the first coding feature $E'_{i,n}$, and R is the matrix of $C_2 \times H \times W$.

[00245] In some embodiments, FIG. 11(b) is an optional structural schematic diagram of a TEM 112 provided by an embodiment of the present disclosure. As illustrated in FIG. 11(b), the TEM may also include a graph reasoning module. By executing the method in S2012A to S2012B, the graph reasoning module may perform quadratic quantization optimization on the first initial quantization feature matrix D output by the 1d-QCO to obtain the first quantization feature matrix D' . D' is the matrix of $C_2 \times HW$. The TEM 112 may obtain the enhanced texture features R by fusing D' with the first coding features $HW \times N$.

[00246] It is understandable that, in the embodiments of the present disclosure, the electronic device may extract quantized low-level texture information from the low-level image features through the process of one-dimensional quantization, and further enhance the low-level texture information through texture enhancement to obtain the low-level texture features, so that the texture details in the low-level texture features are enhanced. In this way, when the details of the texture enhancement are applied to image segmentation, it is helpful to improve the accuracy of image segmentation.

[00247] In some embodiments, a method for performing texture feature statistics based on input image features to obtain target statistical texture features is provided. Corresponding to the process of the texture feature statistics module 31_2 in FIG. 6 or FIG. 7, the method may include any process branch in S501, or S502 to S503, or S504 to S505, and will be described in combination with the steps.

[00248] At S501, two-dimensional quantization is performed on the input image features to obtain the target statistical texture features. The target statistical texture features represent a feature distribution relationship between the pixels.

[00249] In the embodiments of the present disclosure, the electronic device performs two-dimensional quantization on the input image features input in the texture feature statistics module, quantizes the input image features respectively from the dimensions of length and width of the feature space to obtain the target statistical texture features, and represents the feature distribution relationship between the pixels through the target statistical texture features.

[00250] It is to be noted that, here, when the low-level image features are the input image features, the target statistical texture features are the statistical texture features; or

[00251] when the enhanced texture features are the input image features, the target statistical texture features are the low-level texture features; or

[00252] when the features obtained by combining the enhanced texture features with the low-level image features are the input image features, the target statistical texture features are the intermediate low-level texture features.

[00253] At S502, size segmentation is performed on the input image features to obtain local image features corresponding to the input image features.

[00254] At S503, two-dimensional quantization is performed on the local image features to obtain the target statistical texture features.

[00255] In S502 to S503, the electronic device may also first perform size segmentation on the input image features to segment the input image features into several local image features. The electronic device may perform two-dimensional quantization on each of the several local image features in the same process to obtain the target statistical texture features.

[00256] At S504, two-dimensional quantization is performed on the input image features and the local image features respectively to obtain global quantization features corresponding to the input image features and local quantization features corresponding to the local image features.

[00257] At S505, feature fusion is performed based on the global quantization features

and the local quantization features to obtain the target statistical texture features.

[00258] In S504 to S505, the electronic device may perform two-dimensional quantization respectively on the input image features representing the global features and the local image features obtained from size segmentation to obtain the global quantization features and the local quantization features, and then perform feature fusion on the global quantization features and the local quantization features to obtain the target statistical texture features.

[00259] It is to be noted that, in the embodiments of the present disclosure, the electronic device may use the same method to perform two-dimensional quantization on the global image features and the local quantization features.

[00260] In some embodiments, S505 may be implemented by executing the process of S5051 to S5053, and will be described in combination with the steps.

[00261] At S5051, after spatial transformation and averaging is performed on the global quantization features, a global average feature is obtained.

[00262] At S5052, after spatial transformation and averaging is performed on the local quantization features, a local average feature is obtained.

[00263] In S5051 and S5052, by formula (9) and formula (10), the electronic device may perform spatial transformation and averaging on the global quantization features to obtain the global average feature, and obtain the local average feature after performing spatial transformation and averaging on the local quantization features, as follows:

[00264]
$$F' = MLP(F) \quad (9),$$

[00265]
$$T = \frac{\sum_{m=1}^N \sum_{n=1}^N F'_{:,m,n}}{N \cdot N} \quad (10).$$

[00266] In formula (9), for S5051, F represents the global quantization feature; for S5052, F represents the local quantization features. The electronic device may perform MLP spatial domain transformation on the global quantization features or the local

quantization features by formula (9), and accordingly obtain the global quantization features or the local quantization features F' after spatial transformation, where $F' \in \mathbb{R}^{C' \times N \times N}$.

[00267] In formula (10), for S5051, $F'_{:,m,n}$ is the global quantization feature after spatial transformation. The electronic device averages $F'_{:,m,n}$ to obtain the global average feature T . By formula (10), the electronic device may use the same method to average the local quantization features after spatial transformation to obtain the local average feature.

[00268] At S5053, feature fusion is performed on the global average feature and the local average feature to obtain the target statistical texture features.

[00269] At S5053, the electronic device may perform feature fusion on the global average feature and the local average feature to obtain the target statistical texture features.

[00270] In some embodiments, FIG. 12(a) is an optional structural schematic diagram of a PTFEM provided by an embodiment of the present disclosure. As illustrated in FIG. 12, the PTFEM may take the whole low-level image features as the global features and perform pyramid step-by-step size segmentation on the low-level image features to obtain the local features corresponding to each step of size segmentation. Two-dimensional quantization is performed on the global features and each local feature respectively through a 2d-Quantization and Counting Operator (2d-QCO) in the PTFEM to obtain the global quantization features and the local quantization features. Spatial domain transformation and averaging is performed on the global quantization features and the local quantization features respectively to obtain the global average feature and the local average feature, and finally the global average feature and the local average feature are fused after up-sampling to obtain the target statistical texture features corresponding to the low-level image features.

[00271] In some embodiments, S501 may be implemented by executing the process of S5011 to S5014, and will be described in combination with the steps.

- [00272] At S5011, the input image features are pooled to obtain a second average feature corresponding to the input image features.
- [00273] In the embodiments of the present disclosure, the process of performing two-dimensional quantization by the electronic device is similar to the process of performing one-dimensional quantization. First, the input image features are pooled to obtain a second average feature corresponding to the input image features.
- [00274] At S5012, a second similarity between the input image features and the second average feature is calculated to obtain a second similarity matrix.
- [00275] In the embodiments of the present disclosure, the electronic device calculates the second similarity between the input image features and the second average feature to obtain the second similarity matrix.
- [00276] At S5013, two-dimensional quantization is performed on the second similarity matrix to obtain a second quantization series matrix and second coding features.
- [00277] In the embodiments of the present disclosure, for the second similarity matrix, the electronic device may perform two-dimensional quantization on the second similarity matrix in the dimensions of length and width to obtain the second quantization series matrix and the second coding features.
- [00278] In some embodiments, S5013 may be implemented by executing the process of S5013A to S5013C, and will be described in combination with the steps.
- [00279] At S5013A, M-level quantization is performed on two dimensions of the second similarity matrix to obtain a second quantization eigenvalue, the second quantization eigenvalue forming the second quantization series matrix. M is a positive integer greater than or equal to 1.
- [00280] In the embodiments of the present disclosure, the electronic device may perform M-level quantization on two dimensions, such as length and width, in the second similarity matrix. Here, the specific value of M may be different on length or dimension. M is a positive integer greater than or equal to 1.

[00281] In the embodiments of the present disclosure, by formula (11), the electronic device may determine the level $\hat{L}_{m,n}$ of M-level quantization performed on two dimensions of the second similarity matrix, as follows:

$$[00282] \quad \hat{L}_{m,n} = [L_m, L_n] \quad (11),$$

[00283] where, $\hat{L} \in \mathbb{R}^{2 \times N \times N}$.

[00284] At S5013B, an intermediate coding feature matrix is determined based on the second quantization eigenvalue and the similarity matrix.

[00285] At S5013, the electronic device may perform quantization on each of the two dimensions of the second similarity matrix based on the level of quantization determined by formula (11) to obtain the intermediate coding feature matrix.

[00286] At S5013C, based on the intermediate coding feature matrix, the current intermediate coding feature is multiplied by the transpose of its adjacent intermediate coding feature, and a coding feature corresponding to the current intermediate coding feature is determined, so as to obtain the second coding features.

$$[00287] \quad \hat{E} = E_{i,j} \cdot E_{i,j+1}^T \quad (12).$$

[00288] In S5013C, by formula (12), the electronic device may multiply the current intermediate coding feature $E_{i,j}$ in the intermediate coding feature matrix by the transpose $E_{i,j+1}^T$ of its adjacent intermediate coding feature $E_{i,j+1}$, and determine a coding feature corresponding to the current intermediate coding feature, so as to obtain the second coding features $\hat{E}$.

[00289] At S5014, the second coding features are fused with the second quantization series matrix to obtain a second initial quantization feature matrix.

[00290] In S5014, the electronic device fuses the second coding features with the

second quantization series matrix to obtain the second initial quantization feature matrix.

[00291] In some embodiments, S5014 may be implemented by executing the process of S5014A to S5014C, and will be described in combination with the steps.

[00292] At S5014A, spatial size averaging is performed on the second coding features to obtain a second average coding feature.

[00293] At S5014B, the second average coding feature is spliced with the second quantization progression vector to obtain a second quantization statistical matrix.

[00294] In the embodiments of the present disclosure, the process of S5014A and S5014B may be implemented by formula (13) whose description is similar to formula (4) and will not be repeated here.

[00295]
$$C_t = Cat \left(\hat{L}, \frac{\sum_{i=1}^H \sum_{j=1}^W \hat{E}_{m,n,i,j}}{\sum_{m=1}^N \sum_{n=1}^N \sum_{i=1}^H \sum_{j=1}^W \hat{E}_{m,n,i,j}} \right) \quad (13).$$

[00296] In formula (13), $\frac{\sum_{i=1}^H \sum_{j=1}^W \hat{E}_{m,n,i,j}}{\sum_{m=1}^N \sum_{n=1}^N \sum_{i=1}^H \sum_{j=1}^W \hat{E}_{m,n,i,j}}$ is the second average coding

feature, $\hat{L}$ is the second quantization progression vector, and C_t is the calculated second quantization statistical matrix.

[00297] At S5014C, spatial domain transformation is performed on the second quantization statistical matrix, and the second quantization statistical matrix is fused with the second average feature to obtain the second initial quantization feature matrix.

[00298] In S5014C, the electronic device may perform spatial domain transformation on the second quantization statistical matrix in a similar way to the process of one-dimensional quantization, and fuse the second quantization statistical matrix with the second average feature to obtain the second initial quantization feature matrix.

[00299] In some embodiments, the process of executing S5014C may be implemented

by formula (14), as follows:

[00300]
$$D_2 = \text{Cat}(\text{MLP}(C_t), g_2) \quad (14).$$

[00301] In formula (14), C_t is the second quantization statistical matrix, g_2 is the second average feature, and D_2 is the second initial quantization feature matrix.

[00302] It is to be noted that, for the STLNet illustrated in FIG. 7, in the process of training the neural network, formula (15) may be taken as a loss function of network training, and the total loss corresponding to each round of network training is obtained by formula (15). Formula (15) is as follows:

[00303]
$$Lo = Lo_f + \alpha \cdot Lo_a \quad (15).$$

[00304] In formula (15), Lo_f is a prediction loss which represents a global loss value corresponding to the image segmentation result output by the STLNet, Lo_a is an auxiliary prediction loss which represents an auxiliary loss of the low-level image features output by the low-level feature extraction layer, and α is a preset value for adjusting a weight ratio between the prediction loss and the auxiliary prediction loss. In some embodiments, α may be 0.4.

[00305] It can be seen that the STLNet adjusted by using the loss value of formula (15) may improve the accuracy of semantic segmentation when applied to an image segmentation task.

[00306] It is understandable that, in the embodiments of the present disclosure, the electronic device may fully mine and extract, through the process of two-dimensional quantization, the distribution and proportion of the low-level texture information in the image to be processed from the low-level image features as the statistical texture features, so as to optimize the boundary, texture and other structure information of the segmented region in the image segmentation result through the statistical texture features, and then the accuracy of image segmentation is improved.

[00307] The following tables 1 to 4 show a comparison between the effect of image segmentation of the image segmentation method provided by the embodiments of the present disclosure and the effect of image segmentation based on the existing neural networks. It can be seen that the method provided by the embodiments of the present disclosure can effectively improve Intersection Over Union (IoU) after semantic segmentation, which greatly improves the accuracy of image segmentation.

Image segmentation method	mIoU(%)
ResNet-101	76.4
ResNet-101 + SLF	77.0
ResNet-101 + SLF + TEM	80.3
ResNet-101 + SLF + PTFEM	80.0
ResNet-101 + SLF + TEM + PTFEM	80.9
ResNet-101 + ASPP	78.8
ResNet-101 + ASPP + SLF	79.3
ResNet-101 + ASPP + SLF + TEM	80.9
ResNet-101 + ASPP + SLF + PTFEM	80.5
Image segmentation method of the embodiments of the present disclosure ResNet-101 + ASPP + SLF + TEM + PTFEM	81.5

Table 1

Image segmentation method	Backbone network	mIoU(%)
PSANet	ResNet-101	80.1
ANN	ResNet-101	81.3
CPNet	ResNet-101	81.3
CCNet	ResNet-101	81.4
DANet	ResNet-101	81.5
SpyGR	ResNet-101	81.6
ACFNet	ResNet-101	81.8
OCR	ResNet-101	81.8
SFNet	ResNet-101	81.8
Image segmentation method of the embodiments of the present disclosure	ResNet-101	82.3

Table 2

Image segmentation method	Backbone network	mIoU(%)
FCN-8s		37.8
VeryDeep		44.5
Deeplab-v2	ResNet-101	45.7
RefineNet	ResNet-152	47.3
PSPNet	ResNet-101	47.8
MSCI	ResNet-152	50.3
EncNet	ResNet-101	51.7
DANet	ResNet-101	52.6
SpyGR	ResNet-101	52.8
SVCNet	ResNet-101	53.2

CPNet	CResNet-101	53.9
CFNet	ResNet-101	54.0
DMNet	ResNet-101	54.4
RecoNet	esNet-101	54.8
CaC-Net	ResNet-101	55.4
Image segmentation method of the embodiments of the present disclosure	ResNet-101	56.2

Table 3

Image segmentation method	Backbone network	mIoU(%)
RefineNet	ResNet-152	40.70
PSPNet	ResNet-101	43.29
SAC	ResNet-101	44.30
EncNet	ResNet-101	44.65
CFNet	ResNet-101	44.89
CCNet	ResNet-101	45.22
RecoNet	ResNet-101	45.54
CaC-Net	ResNet-101	46.12
CPNet	ResNet-101	46.27
Image segmentation method of the embodiments of the present disclosure	ResNet-101	46.88

Table 4

[00308] In the embodiments of the present disclosure, FIG. 13 is a schematic diagram of a comparison between the effects before and after texture feature statistics is performed on an original image provided by an embodiment of the present disclosure. It can be seen from FIG. 13 that after texture feature statistics is performed on the original image, the texture features of the image are more obvious, which is helpful to improve the accuracy of image segmentation.

[00309] In the embodiments of the present disclosure, FIG. 14 is a schematic diagram of a comparison between the effect of image segmentation of the image segmentation method provided by an embodiment of the present disclosure and the effect of image segmentation based on the neural network DeepLabV3. It can be seen that for the same original image, the boundary of each semantic region of the segmented image in the present disclosure is more consistent with the original image, and the effect of image segmentation is more accurate.

[00310] The present disclosure also provides an image segmentation apparatus. FIG. 15 is a structural schematic diagram of an image segmentation apparatus provided by an embodiment of the present disclosure. As illustrated in FIG. 15, the image segmentation

apparatus 500 may include: a feature extraction part 501, a semantic segmentation part 502, a texture feature processing part 503 and a feature fusion part 504.

[00311] The feature extraction part 501 is configured to extract the low-level image features from the image to be processed.

[00312] The semantic segmentation part 502 is configured to perform semantic segmentation on the low-level image features to obtain the high-level semantic features.

[00313] The texture feature processing part 503 is configured to perform at least one of texture enhancement and texture feature statistics on the low-level image features to obtain the low-level texture features, the low-level texture features representing at least one of the details of the texture enhancement and the statistical distribution of texture features of the image to be processed.

[00314] The feature fusion part 504 is configured to obtain the semantic segmentation image by combining the high-level semantic features and the low-level texture features.

[00315] In some embodiments, the texture feature processing part 502 is further configured to perform any one of the following:

[00316] performing texture enhancement based on the low-level image features to obtain the enhanced texture features; and determining the low-level texture features based on the enhanced texture features;

[00317] performing texture feature statistics based on the low-level image features to obtain the statistical texture features; and determining the low-level texture features based on the statistical texture features;

[00318] performing texture enhancement and texture feature statistics based on the low-level image features to obtain the low-level texture features.

[00319] In some embodiments, the texture feature processing part 502 is further configured to determine the enhanced texture features as the low-level texture features, or combine the enhanced texture features with the low-level image features to obtain the low-level texture features.

- [00320] In some embodiments, the texture feature processing part 502 is further configured to determine the statistical texture features as the low-level texture features, or combine the statistical texture features with the low-level image features to obtain the low-level texture features.
- [00321] In some embodiments, the texture feature processing part 502 is further configured to perform texture enhancement based on the low-level image features to obtain the enhanced texture features, and perform at least texture feature statistics based on the enhanced texture features to obtain the low-level texture features.
- [00322] The texture feature processing part 502 is further configured to: combine the enhanced texture features with the low-level image features, and then perform texture feature statistics to obtain the intermediate low-level texture features; and determine the intermediate low-level texture features as the low-level texture features.
- [00323] In some embodiments, the texture feature processing part 502 is further configured to: combine the enhanced texture features with the low-level image features, and then perform texture feature statistics to obtain the intermediate low-level texture features; and after obtaining the intermediate low-level texture features, combine at least one of the enhanced texture features and the low-level image features with the intermediate low-level texture features to obtain the low-level texture features.
- [00324] The texture feature processing part 502 is further configured to: perform one-dimensional quantization on the low-level image features to obtain the first coding features and the first initial quantization feature matrix; and obtain the enhanced texture features based on the first coding features and the first initial quantization feature matrix.
- [00325] In some embodiments, the texture feature processing part 502 is further configured to: pool the low-level image features to obtain the first average feature corresponding to the low-level image features; calculate the first similarity between the low-level image features and the first average feature to obtain the first similarity matrix; perform one-dimensional quantization on the first similarity matrix to obtain the first quantization progression vector and the first coding features; and fuse the first coding features with the first quantization progression vector to obtain the first initial

quantization feature matrix.

[00326] In some embodiments, the texture feature processing part 502 is further configured to: merge the dimensions of length and width of the first similarity matrix to obtain the one-dimensional similarity vector; perform N-level quantization on the one-dimensional similarity vector to obtain the first quantization eigenvalue, the first quantization eigenvalue forming the first quantization progression vector; and determine the first coding features based on the first quantization eigenvalue and the one-dimensional similarity vector. N is a positive integer greater than or equal to 1.

[00327] In some embodiments, the first quantization eigenvalue includes: N sub quantization eigenvalues. The texture feature processing part 502 is further configured to: perform N-level quantization according to the maximum value and the minimum value in the one-dimensional similarity vector, and determine the average quantization value; and obtain the sub quantization eigenvalue of each level based on the average quantization value and the minimum value, so as to obtain the N sub quantization eigenvalues.

[00328] In some embodiments, the texture feature processing part 502 is further configured to: perform spatial size averaging on the first coding features to obtain the first average coding feature; splice the first average coding feature with the first quantization progression vector to obtain the first quantization statistics vector; perform spatial domain transformation on the first quantization statistics vector, and fuse the first quantization statistics vector with the first average feature to obtain the first initial quantization feature matrix.

[00329] In some embodiments, the texture feature processing part 502 is further configured to: perform graph reasoning on the first initial quantization feature matrix to obtain the first quantization feature matrix; and obtain the enhanced texture features based on the first quantization feature matrix and the first coding features.

[00330] In some embodiments, the texture feature processing part 502 is further configured as: when the low-level image features are the input image features, the target statistical texture features are the statistical texture feature; or when the enhanced texture features are the input image features, the target statistical texture features are the low-

level texture features; or when the features obtained by combining the enhanced texture features with the low-level image features are the input image features, the target statistical texture features are the intermediate low-level texture features.

[00331] In some embodiments, the texture feature processing part 502 is further configured to: perform two-dimensional quantization on the input image features to obtain the target statistical texture features, the target statistical texture features representing the feature distribution relationship between pixels; or perform size segmentation on the input image features to obtain the local image features corresponding to the input image features, and perform two-dimensional quantization on the local image features to obtain the target statistical texture feature; or perform two-dimensional quantization on the input image features and the local image features respectively to obtain the global quantization features corresponding to the input image features and the local quantization features corresponding to the local image feature; and perform feature fusion based on the global quantization features and the local quantization features to obtain the target statistical texture features.

[00332] In some embodiments, the texture feature processing part 502 is further configured to: pool the input image features to obtain the second average feature corresponding to the input image features; calculate the second similarity between the input image features and the second average feature to obtain the second similarity matrix; perform two-dimensional quantization on the second similarity matrix to obtain the second quantization series matrix and the second coding feature; and fuse the second coding features with the second quantization series matrix to obtain the second initial quantization feature matrix.

[00333] In some embodiments, the texture feature processing part 502 is further configured to: perform M-level quantization on two dimensions of the second similarity matrix to obtain the second quantization eigenvalue, the second quantization eigenvalue forming the second quantization series matrix, and M being a positive integer greater than and equal to 1; determine the intermediate coding feature matrix based on the second quantization eigenvalue and the similarity matrix; and based on the intermediate coding feature matrix, multiply the current intermediate coding feature by the transpose

of its adjacent intermediate coding feature, and determine a coding feature corresponding to the current intermediate coding feature, so as to obtain the second coding features.

[00334] In some embodiments, the texture feature processing part 502 is further configured to: perform spatial size averaging on the second coding features to obtain the second average coding feature; splice the second average coding feature with the second quantization progression vector to obtain the second quantization statistical matrix; and perform spatial domain transformation on the second quantization statistical matrix, and fuse the second quantization statistical matrix with the second average feature to obtain the second initial quantization feature matrix.

[00335] In some embodiments, the texture feature processing part 502 is further configured to: after performing spatial transformation and averaging on the global quantization features, obtain the global average feature; after performing spatial transformation and averaging on the local quantization features, obtaining the local average feature; and perform feature fusion on the global average feature and the local average feature to obtain the target statistical texture features.

[00336] It is to be noted that the above descriptions about the apparatus embodiments are similar to descriptions about the method embodiments and beneficial effects similar to those of the method embodiments are achieved. Technical details undisclosed in the apparatus embodiments of the present disclosure may be understood with reference to the descriptions about the method embodiments of the present disclosure.

[00337] The embodiments of the present disclosure also provide an electronic device. FIG. 16 is a structural schematic diagram of an electronic device provided by an embodiment of the present disclosure. As illustrated in FIG. 16, the electronic device 2 may include: a memory 21 and a processor 22. The memory 21 and the processor 22 are connected through a communication bus 23. The memory 21 is configured to store an executable computer program. The processor 22 is configured to implement, when executing the executable computer program stored in the memory 21, the image segmentation method provided by the embodiments of the present disclosure.

[00338] The embodiments of the present disclosure provide a computer-readable

storage medium, in which a computer program is stored. When executed by the processor 22, the computer program implements the image segmentation method provided by the embodiments of the present disclosure.

[00339] The embodiments of the present disclosure also provide a computer program, which may include a computer-readable code. When the computer-readable code runs in an electronic device, the processor in the electronic device implements the above method.

[00340] In some embodiments of the present disclosure, the storage medium may be an FRAM, an ROM, a PROM, an EPROM, an EEPROM, a flash memory, a magnetic surface storage, a compact disc, or a CD-ROM, and may also be a variety of devices including one or any combination of the above memories.

[00341] In some embodiments of the present disclosure, the executable instruction may be compiled according to a programming language of any form (including a compiling or interpretive language, or a declarative or procedural language) in form of a program, software, a software module, a script, or a code, and may be deployed according to any form, including deployed as an independent program or deployed as a module, a component, a subroutine or another unit suitable to be used in a computing environment.

[00342] As an example, the executable instruction may but not always correspond to a file in a file system, and may be stored in a part of a file that stores another program or data, for example, stored in one or more scripts in a Hyper Text Markup Language (HTML) document, stored in a single file dedicated to a discussed program, or stored in multiple collaborative files (for example, files storing one or more modules, subprograms or code parts).

[00343] As an example, the executable instruction may be deployed in a computing device for execution, or executed in multiple computing devices at the same place, or executed in multiple computing devices that are interconnected through a communication network at multiple places.

[00344] In some embodiments, the computer-readable storage medium may be a physical device that can hold and store instructions used by an instruction execution device, and may be a volatile storage medium or a non-volatile storage medium. The

computer-readable storage medium may be, but is not limited to, an FRAM, an ROM, a PROM, an EPROM, an EEPROM, a flash memory, a magnetic surface storage, a compact disc, or a CD-ROM, and may also be a variety of devices including one or any combination of the above memories.

[00345] In some embodiments, the executable instruction may be compiled according to a programming language of any form (including a compiling or interpretive language, or a declarative or procedural language) in form of a program, software, a software module, a script, or a code, and may be deployed according to any form, including deployed as an independent program or deployed as a module, a component, a subroutine or another unit suitable to be used in a computing environment.

[00346] As an example, the executable instruction may but not always correspond to a file in a file system, and may be stored in a part of a file that stores another program or data, for example, stored in one or more scripts in a Hyper Text Markup Language (HTML) document, stored in a single file dedicated to a discussed program, or stored in multiple collaborative files (for example, files storing one or more modules, subprograms or code parts).

[00347] As an example, the executable instruction may be deployed in a computing device for execution, or executed in multiple computing devices at the same place, or executed in multiple computing devices that are interconnected through a communication network at multiple places.

[00348] To sum up, in the embodiments of the present disclosure, the electronic device performs at least one of texture enhancement and texture feature statistics on the low-level image features of the image to be processed, which can enhance the texture details of the image to be processed, mine multi-scale deep texture information in the image to be processed, and obtain the low-level texture features that can represent the details of the texture enhancement and/or the statistical distribution of texture features. The electronic device combines the low-level texture features with the high-level semantic features to obtain the final segmented image, which can better restore the low-level texture information of the original image in the segmented image, thus improving the accuracy of image segmentation. Moreover, the electronic device may extract quantized

low-level texture information from the low-level image features through the process of one-dimensional quantization, and further enhance the low-level texture information through texture enhancement to obtain the low-level texture features, so that the texture details in the low-level texture features are enhanced. In this way, when the details of the texture enhancement are applied to image segmentation, it is helpful to improve the accuracy of image segmentation.

[00349] The above is only the embodiments of the present disclosure and not intended to limit the protection scope of the present disclosure. Any modifications, equivalent replacements, improvements and the like within the spirit and scope of the present disclosure should fall within the protection scope of the claims of the present disclosure.

INDUSTRIAL APPLICABILITY

[00350] In the embodiments of the present application, the low-level image features are extracted from the image to be processed, and semantic segmentation is performed on the low-level image features to obtain the high-level semantic features; at least one of texture enhancement and texture feature statistics is performed on the low-level image features to obtain the low-level texture features, the low-level texture features representing at least one of the details of the texture enhancement and the statistical distribution of texture features of the image to be processed; and the semantic segmentation image is obtained by combining the high-level semantic features and the low-level texture features. The accuracy of image segmentation is improved by the above method.

CLAIMS

1. A method for image segmentation, comprising:

extracting low-level image features from an image to be processed, and obtaining high-level semantic features by performing semantic segmentation on the low-level image features;

obtaining low-level texture features by performing at least one of texture enhancement and texture feature statistics on the low-level image features, the low-level texture features representing at least details of the texture enhancement or statistical distribution of the texture features of the image to be processed; and

obtaining a semantic segmentation image by combining the high-level semantic features and the low-level texture features.

2. The method of claim 1, wherein obtaining the low-level texture features by performing at least one of texture enhancement and texture feature statistics on the low-level image features comprises any one of:

obtaining an enhanced texture feature by performing texture enhancement based on the low-level image features; and determining the low-level texture features based on the enhanced texture feature;

obtaining a statistical texture feature by performing texture feature statistics based on the low-level image features; and determining the low-level texture features based on the statistical texture feature; and

obtaining the low-level texture features by performing texture enhancement and texture feature statistics based on the low-level image features.

3. The method of claim 2, wherein determining the low-level texture features based on the enhanced texture features comprises:

determining the enhanced texture features as the low-level texture features; or

obtaining the low-level texture features by combining the enhanced texture features with the low-level image features.

4. The method of claim 2, wherein determining the low-level texture features based on the statistical texture features comprises:

determining the statistical texture features as the low-level texture features; or

combining the statistical texture features with the low-level image features to obtain the low-level texture features.

5. The method of claim 2, wherein obtaining the low-level texture features by performing texture enhancement and texture feature statistics based on the low-level image features comprises:

obtaining the enhanced texture feature by performing texture enhancement based on the low-level image features; and

obtaining the low-level texture features by performing texture feature statistics based on at least the enhanced texture features.

6. The method of claim 5, wherein obtaining the low-level texture features by performing texture feature statistics based on at least the enhanced texture features comprises:

combining the enhanced texture features with the low-level image features, and obtaining an intermediate low-level texture feature by performing texture feature statistics;

and determining the intermediate low-level texture features as the low-level texture features.

7. The method of claim 6, wherein after combining the enhanced texture features with the low-level image features and obtaining the intermediate low-level texture features by performing texture feature statistics, the method further comprises:

obtaining the low-level texture features by combining at least one of the enhanced texture features and the low-level image features with the intermediate low-level texture features.

8. The method of any one of claims 2, 3, and 5 to 7, wherein obtaining the enhanced texture features by performing texture enhancement based on the low-level image features comprises:

obtaining a first coding features and a first initial quantization feature matrix by performing one-dimensional quantization on the low-level image features; and

obtaining the enhanced texture features based on the first coding features and the first initial quantization feature matrix.

9. The method of claim 8, wherein obtaining the first coding features and the first initial quantization feature matrix by performing one-dimensional quantization on the low-level image features comprises:

obtaining a first average feature corresponding to the low-level image features by pooling the low-level image features;

obtaining a first similarity matrix by calculating a first similarity between the low-level image features and the first average feature;

obtaining a first quantization progression vector and the first coding feature by performing one-dimensional quantization on the first similarity matrix; and

obtaining the first initial quantization feature matrix by fusing the first coding features with the first quantization progression vector.

10. The method of claim 9, wherein obtaining the first quantization progression vector and the first coding features by performing one-dimensional quantization on the first similarity matrix comprises:

obtaining a one-dimensional similarity vector by merging dimensions of length and width of the first similarity matrix;

obtaining a first quantization eigenvalue by performing N-level quantization on the one-dimensional similarity vector, the first quantization eigenvalue forming the first quantization progression vector; and

determining the first coding features based on the first quantization eigenvalue and the one-dimensional similarity vector; wherein N is a positive integer greater than or equal to 1.

11. The method of claim 10, wherein the first quantization eigenvalue comprises: N sub quantization eigenvalues; obtaining the first quantization eigenvalue by performing N-level quantization on the one-dimensional similarity vector comprises:

performing N-level quantization according to a maximum value and minimum value in the one-dimensional similarity vector, and determining an average quantization value; and

obtaining the N sub quantization eigenvalues by obtaining a sub quantization eigenvalue of each level based on the average quantization value and the minimum value.

12. The method of any one of claims 9 to 11, wherein obtaining the first initial quantization feature matrix by fusing the first coding features with the first quantization progression vector comprises:

obtaining a first average coding feature by performing spatial size averaging on the first coding features;

obtaining a first quantization statistics vector by splicing the first average coding feature with the first quantization progression vector;

performing spatial domain transformation on the first quantization statistics vector, and fusing the first quantization statistics vector with the first average feature to obtain the first initial quantization feature matrix.

13. The method of any one of claims 8 to 12, wherein obtaining the enhanced texture features based on the first coding features and the first initial quantization feature matrix comprises:

obtaining a first quantization feature matrix by performing graph reasoning on the first initial quantization feature matrix; and

obtaining the enhanced texture features based on the first quantization feature matrix and the first coding features.

14. The method of claim 6, wherein

when the low-level image features are input image features, target statistical texture features are the statistical texture features; or

when the enhanced texture features are the input image features, the target statistical texture features are the low-level texture features; or

when features obtained by combining the enhanced texture features with the low-level image features are the input image features, the target statistical texture features are the intermediate low-level texture features.

15. The method of claim 14, wherein obtaining the target statistical texture features by performing texture feature statistics based on the input image features comprises:

obtaining the target statistical texture features by performing two-dimensional quantization on the input image features, the target statistical texture features representing a feature distribution relationship between pixels; or

obtaining local image features corresponding to the input image features by performing size segmentation on the input image features, and obtaining the target statistical texture feature by performing two-dimensional quantization on the local image features; or

obtaining global quantization features corresponding to the input image features and local quantization features corresponding to the local image feature by performing two-dimensional quantization on the input image features and the local image features respectively; and

obtaining the target statistical texture features by performing feature fusion based on the global quantization features and the local quantization features.

16. The method of claim 15, wherein obtaining the target statistical texture features by performing two-dimensional quantization on the input image features comprises:

obtaining a second average feature corresponding to the input image features by pooling the input image features;

obtain a second similarity matrix by calculating a second similarity between the input image features and the second average feature;

obtaining a second quantization series matrix and a second coding feature by performing two-dimensional quantization on the second similarity matrix; and

obtaining a second initial quantization feature matrix by fusing the second coding features with the second quantization series matrix.

17. The method of claim 16, wherein obtaining the second quantization series matrix and the second coding features by performing two-dimensional quantization on the second similarity matrix comprises:

obtaining a second quantization eigenvalue by performing M-level quantization on two dimensions of the second similarity matrix, the second quantization eigenvalue forming the second quantization series matrix; wherein M is a positive integer greater than or equal to 1;

determining an intermediate coding feature matrix based on the second quantization eigenvalue and the similarity matrix; and

determining a coding feature corresponding to a current intermediate coding feature by multiplying the current intermediate coding feature by a transpose of adjacent intermediate coding feature of the current intermediate coding feature based on the intermediate coding feature matrix, to obtain the second coding features.

18. The method of claim 16 or 17, wherein obtaining the second initial quantization feature matrix by fusing the second coding features with the second quantization series matrix comprises:

obtaining a second average coding feature by performing spatial size averaging on the second coding features;

obtaining a second quantization statistical matrix by splicing the second average coding feature with the second quantization progression vector; and

performing spatial domain transformation on the second quantization statistical matrix, and fusing the second quantization statistical matrix with the second average feature to obtain the second initial quantization feature matrix.

19. The method of claim 15, wherein obtaining the target statistical texture features by

performing feature fusion based on the global quantization features and the local quantization features comprises:

- obtaining a global average feature by performing spatial transformation and averaging on the global quantization features;

- obtaining a local average feature by performing spatial transformation and averaging on the local quantization features; and

- obtaining the target statistical texture features by performing feature fusion on the global average feature and the local average feature.

20. An image segmentation apparatus, comprising:

- a feature extraction part, configured to extract low-level image features from an image to be processed;

- a semantic segmentation part, configured to obtain high-level semantic features by performing semantic segmentation on the low-level image features;

- a texture feature processing part, configured to obtain low-level texture features by performing at least one of texture enhancement and texture feature statistics on the low-level image features, the low-level texture features representing at least details of the texture enhancement or statistical distribution of the texture features of the image to be processed; and

- a feature fusion part, configured to obtain a semantic segmentation image by combining the high-level semantic features and the low-level texture features.

21. An electronic device, comprising;

- a memory, configured to store executable instructions;

a processor, configured to implement, when executing the executable instructions stored in the memory, the method of any one of claims 1 to 19.

22. A computer-readable storage medium, storing executable instructions that when executed by a processor, implement the method of any one of claims 1 to 19.

23. A computer program, comprising computer-readable codes, wherein when the computer-readable codes runs in an electronic device, a processor in the electronic device implements the method of any one of claims 1 to 19.