

ANOMALY DETECTION METHOD AND APPARATUS, AND ELECTRONIC DEVICE AND STORAGE MEDIUM

[0001] This application claims the priority to Chinese Patent Application No.
5 202311814369.1, titled "ANOMALY DETECTION METHOD AND APPARATUS, AND
ELECTRONIC DEVICE AND STORAGE MEDIUM", filed on December 27, 2023 with the
China National Intellectual Property Administration, which is incorporated herein by reference
in its entirety.

10 FIELD

[0002] The present disclosure relates to the technical field of image processing, and in
particular to a method for anomaly detection, an apparatus for anomaly detection, an electronic
device, and a storage medium.

15 BACKGROUND

[0003] Anomaly detection is widely applied in the industrial field. According to anomaly
detection methods in the prior art, only a small amount of labeled data is typically required to
detect unknown anomalies. For example, in feature comparison algorithms (such as PatchCore),
features of labeled data are extracted based on a network model pre-trained on ImageNet to
20 generate a feature bank. However, the feature bank is greatly influenced by the pre-trained
network model, and the retrieval speed of the feature bank is further affected by the feature
dimension. Additionally, the features extracted by open-source pre-trained network models may
not be suitable for detail textures. Due to the rich variety of texture patterns, the distance metric
of the features extracted from normal textures may be greater than that of features extracted
25 from abnormal textures, making it impossible to determine an appropriate threshold to
accurately distinguish between abnormal textures and normal textures.

SUMMARY

[0004] In view of the above issues, the present disclosure is provided. According to the
30 present disclosure, a method for anomaly detection, an apparatus for anomaly detection, an

electronic device, and a storage medium are provided.

[0005] A method for anomaly detection is provided according to an aspect of the present disclosure. The method comprises:

acquiring a to-be-detected image, where the to-be-detected image comprises a target object;

5 inputting the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature; and

inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, where the anomaly detection result is to indicate whether an abnormal region exists in the target object,
10 and where

the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing, the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with
15 a feature bank to determine the anomaly detection result, and the feature bank is a sample feature bank obtained by the feature extraction module performing feature extraction on a plurality of first sample images.

[0006] In a possible implementation, the feature extraction module comprises an initial feature extraction network, a stacked flow network, and a feature fusion network, and the
20 inputting the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature comprises:

inputting the to-be-detected image into the initial feature extraction network to obtain a first image feature, where the first image feature is of multi-scale;

inputting the first image feature into the stacked flow network and mapping the first image
25 feature to a latent space by using the stacked flow network to obtain a second image feature, where the second image feature is of the same feature scale as the first image feature; and

inputting the second image feature into the feature fusion network to obtain the target image feature,

where the stacked flow network and the feature fusion network are configured to perform

the normalizing flow processing.

[0007] In a possible implementation, the to-be-detected image comprises a plurality of image patches, the inputting the to-be-detected image into the initial feature extraction network to obtain a first image feature comprises:

5 inputting the plurality of image patches into the initial feature extraction network respectively to obtain respective first image patch features corresponding to the plurality of image patches, where the first image feature comprises the first image patch features corresponding to the plurality of image patches,

10 the inputting the first image feature into the stacked flow network and mapping the first image feature to a latent space by using the stacked flow network to obtain a second image feature comprises:

15 inputting the first image patch features corresponding to the plurality of image patches into the stacked flow network respectively and mapping the first image patch features corresponding to the plurality of image patches to the latent space by using the stacked flow network to obtain respective second image patch features corresponding to the plurality of image patches, where the second image feature comprises the second image patch features corresponding to the plurality of image patches, and

 the inputting the second image feature into the feature fusion network to obtain the target image feature comprises:

20 inputting the second image patch features corresponding to the plurality of image patches into the feature fusion network to obtain respective target image patch features corresponding to the plurality of image patches, where the target image feature comprises the target image patch features corresponding to the plurality of image patches.

[0008] In a possible implementation, the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object comprises:

 calculating, by using the anomaly determination module, a distance between the target image feature and a reference image feature in the feature bank; and

30 determining, in a case that the distance is greater than or equal to a first distance threshold, that an abnormal region exists in the target object.

[0009] In a possible implementation, the feature bank comprises respective reference image patch features corresponding to a plurality of image positions, and the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object comprises:

5 for each image patch of the plurality of image patches:

calculating, by using the anomaly determination module, a distance between the target image patch feature corresponding to the image patch and a reference image patch feature among the reference image patch features in the feature bank, where the reference image patch feature corresponds to the image position of the image patch; and

10 determining, in a case that the distance is greater than or equal to a second distance threshold, that an abnormal region exists within the image patch.

[0010] In a possible implementation, the trained anomaly detection model is obtained by:

acquiring a plurality of second sample images;

15 inputting the plurality of second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features corresponding to the plurality of second sample images;

calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images; and

20 optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model.

[0011] In a possible implementation, the trained anomaly detection model is obtained by:

acquiring a plurality of second sample images;

25 inputting the plurality of second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features which are outputted by the stacked flow network and correspond to the plurality of second sample images, where each of the sample image features comprises multi-scale sub-sample image features;

calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images; and

optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model,

where the calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images comprises:

5 calculating, for each feature scale among feature scales of the sample image features, a sub-prediction loss for the feature scale based on differences between respective sub-sample image features of the feature scale corresponding to the plurality of second sample images; and

summing or averaging the sub-prediction losses for all the feature scales of the sample image features to obtain the prediction loss.

10 **[0012]** In a possible implementation, in the process of summing or averaging the sub-prediction losses for all the feature scales of the sample image features, the sub-prediction losses for the feature scales are provided with respective weighting factors.

[0013] In a possible implementation, before the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly
15 detection result of the target object, the method further comprises:

acquiring the plurality of first sample images;

inputting the plurality of first sample images into the feature extraction module in the trained anomaly detection model respectively to obtain respective sample image features corresponding to the plurality of first sample images;

20 performing clustering on the sample image features corresponding to the plurality of first sample images; and

for each cluster of clusters obtained through the clustering, determining a representative sample image feature in the cluster as a reference image feature to obtain the feature bank.

[0014] An apparatus for anomaly detection is provided according to another aspect of the
25 present disclosure. The apparatus comprises:

an acquisition module, configured to acquire a to-be-detected image, where the to-be-detected image comprises a target object;

a first input module, configured to input the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature; and

a second input module, configured to input the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, where the anomaly detection result is to indicate whether an abnormal region exists in the target object, and where

the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing, and the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with a feature bank to determine the anomaly detection result.

[0015] An electronic device is provided according to another aspect of the present disclosure. The electronic device comprises a processor and a memory, where the memory stores a computer program instruction, and the computer program instruction, when executed by the processor, performs the method for anomaly detection described above.

[0016] A storage medium is provided according to another aspect of the present disclosure. The storage medium stores a computer program/instruction, where the computer program/instruction, when executed, performs the method for anomaly detection described above.

[0017] According to the above technical solution, the acquired to-be-detected image is inputted into a feature extraction module in a trained anomaly detection model, so as to obtain a target image feature. Then, the target image feature is inputted into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of a target object. In this solution, the multi-scale image feature of the to-be-detected image is extracted by using the feature extraction module, and the multi-scale image feature is converted into a target image feature of a preset scale through normalizing flow processing. In this solution, target image features of different scales may be determined and outputted according to practical application scenarios, thereby improving inference speed while reducing interference from background information. This method is applicable to anomaly detection in scenarios with variable backgrounds, such as textures of materials like metal. Furthermore, through normalizing flow processing, anomaly detection at high resolutions is achieved, while further enabling anomaly detection results to be determined for different feature scales of the to-be-detected image.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] The above and other objectives, features, and advantages of the present disclosure become apparent by describing the embodiments of the present disclosure in detail in conjunction with the drawings. The drawings are used to provide a further understanding of the
5 embodiments of the present disclosure, constitute a part of the specification, explain the present disclosure together with the embodiments of the present disclosure, and do not constitute a limitation to the present disclosure. In the drawings, the same reference numerals generally represent the same components or steps.

[0019] FIG. 1 is a flowchart of a method for anomaly detection according to an embodiment
10 of the present disclosure;

[0020] FIG. 2 is a schematic diagram of a feature extraction module according to an embodiment of the present disclosure;

[0021] FIG. 3 is a schematic block diagram of an apparatus for anomaly detection according to an embodiment of the present disclosure; and

[0022] FIG. 4 is a schematic block diagram of an electronic device according to an
15 embodiment of the present disclosure.

DETAILED DESCRIPTION

[0023] To make the objectives, technical solutions, and advantages of the present disclosure
20 apparent, exemplary embodiments of the present disclosure are described in detail below with reference to the drawings. Apparently, the described embodiments are only a part of the embodiments of the present disclosure, rather than all of them. It should be understood that the present disclosure is not limited by the exemplary embodiments described herein. Based on the embodiments of the present disclosure described in the present disclosure, all other
25 embodiments obtained by those skilled in the art without creative efforts shall fall within the protection scope of the present disclosure.

[0024] To at least partially address the above issues, a method for anomaly detection is provided according to embodiments of the present disclosure. FIG.1 is a flowchart of a method for anomaly detection according to an embodiment of the present disclosure. As shown in FIG.1,
30 the method includes steps S110, S120, and S130.

[0025] In step S110, a to-be-detected image is acquired, where the to-be-detected image includes a target object.

[0026] For example, the to-be-detected image may be any type of image including a target object. The target object may be any type of item such as a wafer, a character, an electronic component. The to-be-detected image may be a static image or any video frame in a dynamic video. The to-be-detected image may be an original image captured by an image acquisition apparatus (e.g., an image sensor in a camera), or an image obtained after preprocessing (such as digitizing, normalizing, smoothing) the original image.

[0027] In step S120, the to-be-detected image is inputted into a feature extraction module in a trained anomaly detection model to obtain a target image feature, where the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing.

[0028] For example, the anomaly detection model may be any suitable neural network model, which is currently available or may emerge in the future, capable of implementing an anomaly detection function. Moreover, the anomaly detection model is an algorithm suitable for performing anomaly detection based on extracted features. Anomaly detection algorithms include, but are not limited to, PaDim algorithm, PatchCore algorithm, and the like.

[0029] Through inputting the to-be-detected image into the feature extraction module in the trained anomaly detection model, a target image feature is obtained. The to-be-detected image may be represented as I , and $I \in \mathbb{R}^{H_0 \times W_0 \times C_0}$, where, H_0 , W_0 , and C_0 represent the height, width, and the channel count of the to-be-detected image I , respectively. For example, when the to-be-detected image I is an RGB image, the channel count C_0 may be 3. The to-be-detected image I is inputted into the feature extraction module, and the feature extraction module performs initial feature extraction on the to-be-detected image to obtain multiple image features of different scales (hereinafter referred to as a multi-scale image feature). The process of initial feature extraction mentioned here includes: increasing the channel count of the to-be-detected image to a preset channel count to obtain a single-scale image feature with the preset channel count, and performing Laplacian pyramid expansion on the single-scale image feature. For example, the initial channel count of the to-be-detected image is C_0 , after convolution with a 1×1 convolution kernel, a to-be-detected image I' with C_1 channels is obtained. By performing

multi-scale expansion on the image feature of the to-be-detected image I' using the Laplacian pyramid algorithm, a multi-scale image feature is obtained. Normalizing flow processing and inverse Laplacian pyramid processing are performed on the obtained multi-scale image feature to obtain a target image feature of a preset scale. The channel count of the target image feature of the preset scale is the same as the channel count of the to-be-detected image I'.

[0030] In an embodiment, after the channel count of the to-be-detected image I is increased by using 1×1 convolution, the image features X1, X2, X3, and X4 are obtained using the Laplacian pyramid algorithm. Then, normalizing flow processing is performed on the image features X1, X2, X3, and X4 to obtain image features M1, M2, M3, and M4, and the inverse Laplacian pyramid processing is performed on the image features M1, M2, M3, and M4 to obtain a target image feature Q. The scales corresponding to the image features X1, X2, X3, and X4, or the scales corresponding to the image features M1, M2, M3, and M4, increase or decrease in sequence.

[0031] In an embodiment, the feature extraction module includes an initial feature extraction network, a stacked flow network, and a feature fusion network, and the inputting the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature includes: inputting the to-be-detected image into the initial feature extraction network to obtain a first image feature, where the first image feature is of multi-scale; inputting the first image feature into the stacked flow network and mapping the first image feature to a latent space by using the stacked flow network to obtain a second image feature, where the second image feature is of the same feature scale as the first image feature; inputting the second image feature into the feature fusion network to obtain the target image feature, where the stacked flow network and the feature fusion network are configured to perform the normalizing flow processing.

[0032] In an embodiment, the to-be-detected image is inputted into the initial feature extraction network to obtain a first image feature. For example, the initial feature extraction network is implemented using a 1×1 convolution kernel, a convolutional neural network (CNN), or a vision transformer (ViT). By way of example and not limitation, the stacked flow network may be implemented using flow networks such as stacked normalization flow network (NF). FIG. 2 is a schematic diagram of a feature extraction module according to an embodiment of the present disclosure. After the initial channel count C_0 of the to-be-detected image

$I \in \mathbb{R}^{H_0 \times W_0 \times C_0}$ is increased, a to-be-detected image $I' \in \mathbb{R}^{H_0 \times W_0 \times C_1}$ is obtained. The Laplacian pyramid expansion is performed on the initial image feature of the to-be-detected image I' to obtain first image features X of 4 scales, namely first image features X_1, X_2, X_3 , and X_4 . The first image features X of 4 scales are inputted into the stacked flow network (NF), and the first
5 image features X are mapped to the latent space by using the stacked flow network to obtain second image features M of 4 scales, namely second image features M_1, M_2, M_3 , and M_4 . The feature scales of the second image features M are the same as the feature scales of the first image features X . For example, if the scale of the first image feature X_1 is 4×4 , the scale of the second image feature M_1 corresponding to the first image feature X_1 is 4×4 . Corresponding
10 elements across the second image features M of 4 scales are fused by using the inverse Laplacian pyramid algorithm to obtain a target image feature Q . The normalizing flow processing in the foregoing embodiments may include the operation performed on the first image features by the stacked flow network and the fusing operation performed on the second image features by the feature fusion network. The stacked flow network shown in FIG.2 is
15 merely illustrative, and the number of stacked flow networks may be set according to practical application scenarios, which is not limited in the present disclosure.

[0033] According to the above technical solution, the to-be-detected image is inputted into the initial feature extraction network to obtain a first image feature. The first image feature is inputted into the stacked flow network and the first image feature is mapped to the latent space
20 by using the stacked flow network to obtain a second image feature. Then, the second image feature is inputted into the feature fusion network to obtain the target image feature. Based on the solution, a dual mapping relationship between the image space and the latent space is established, thereby enabling a better fit to the true distribution of the data based on the obtained target image feature. In addition, by performing normalizing processing using the stacked flow
25 network, image features at different scales and multiple frequencies are extracted, thus improving the accuracy of the anomaly detection result.

[0034] In step S130, the target image feature is inputted into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, where the anomaly detection result is to indicate whether an abnormal region
30 exists in the target object, the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with a feature bank to determine the anomaly detection result, and the feature bank is a sample feature bank obtained by the feature

extraction module performing feature extraction on multiple first sample images.

[0035] In an embodiment, the trained anomaly detection model further includes an anomaly determination module. By inputting the target image feature into the anomaly determination module, an anomaly detection result of the target object is obtained. In an embodiment of the present disclosure, the anomaly detection result is indicated by an anomaly detection bounding box. The anomaly detection bounding box may be of any shape, and a rectangular box is preferable. In a case that the anomaly detection bounding box is a rectangular box, the position information of an abnormal region in the to-be-detected image may be represented by the position of the corresponding rectangular box in the to-be-detected image. The similarity between any reference image feature in the feature bank and the target image feature is determined through comparison by using the anomaly determination module. When the similarity does not exceed a target similarity threshold, it is determined that an abnormal region exists in the target object. A distance between any reference image feature in the feature bank and the target image feature is further determined by using the anomaly determination module. By comparing the distance with a preset distance threshold (e.g., a first distance threshold or a second distance threshold in the following embodiments), it is determined whether an abnormal region exists in the target object. The feature bank may be a sample feature bank obtained by performing clustering on multiple reference image features. The multiple reference image features may be obtained by the feature extraction model processing multiple first sample images according to the above method. For example, the first sample images are positive sample images (OK images).

[0036] According to the above technical solution, the acquired to-be-detected image is inputted into a feature extraction module in a trained anomaly detection model, so as to obtain a target image feature. Then, the target image feature is inputted into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of a target object. In this solution, the multi-scale image feature of the to-be-detected image is extracted by using the feature extraction module, and the multi-scale image feature is converted into a target image feature of a preset scale through normalizing flow processing. In this solution, target image features with different channel counts may be determined and outputted according to practical application scenarios to meet the requirements of different application scenarios. In addition, the outputted target features experience reduced interference from background information with this solution. This method is applicable to anomaly detection in scenarios with

variable backgrounds, such as textures of materials like metal. Furthermore, through normalizing flow processing, anomaly detection at high resolutions is achieved, while further enabling anomaly detection results to be determined for different feature scales of the to-be-detected image.

5 **[0037]** In an embodiment, the to-be-detected image includes multiple image patches, the inputting the to-be-detected image into the initial feature extraction network to obtain a first image feature includes: inputting the multiple image patches into the initial feature extraction network respectively to obtain respective first image patch features corresponding to the multiple image patches, where the first image feature includes the first image patch features
10 corresponding to the multiple image patches, the inputting the first image feature into the stacked flow network and mapping the first image feature to a latent space by using the stacked flow network to obtain a second image feature includes: inputting the first image patch features corresponding to the multiple image patches into the stacked flow network respectively and mapping the first image patch features corresponding to the multiple image patches to the latent
15 space by using the stacked flow network to obtain respective second image patch features corresponding to the multiple image patches, where the second image feature includes the second image patch features corresponding to the multiple image patches, and the inputting the second image feature into the feature fusion network to obtain the target image feature includes: inputting the second image patch features corresponding to the multiple image patches into the
20 feature fusion network to obtain respective target image patch features corresponding to the multiple image patches, where the target image feature includes the target image patch features corresponding to the multiple image patches.

[0038] In an embodiment, the to-be-detected image is divided into multiple image patches. For example, for a to-be-detected image with an image size of 16×16 , the to-be-detected image
25 is divided into four image patches, each with a size of 4×4 . Referring to the method for acquiring the first image feature, the second image feature, and the target image feature in the foregoing embodiments, a first image patch feature, a second image patch feature, and a target image patch feature which correspond to each image patch may be obtained. For brevity, details are not repeated here. The target image feature may include respective target image patch
30 features corresponding to the multiple image patches.

[0039] According to the above technical solution, the to-be-detected image is divided into

multiple image patches, and the first image patch feature and the second image patch feature which correspond to each image patch are acquired to obtain the target image patch feature. In this way, feature extraction is performed for individual portions of the region in the to-be-detected image to detect whether an abnormal region exists in the multiple image patches, further improving the reliability of the anomaly detection result.

[0040] In an embodiment, the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object includes: calculating, by using the anomaly determination module, a distance between the target image feature and a reference image feature in the feature bank; and determining, in a case that the distance is greater than or equal to a first distance threshold, that an abnormal region exists in the target object.

[0041] In an embodiment, the target image feature is inputted into the anomaly determination module, and the distance between the target image feature and the reference image feature in the feature bank is calculated by using the anomaly determination module. The distance may be any distance such as Euclidean distance, and Mahalanobis distance. The feature bank may be acquired and stored in a storage apparatus of a host computer in advance, or may be obtained after training the anomaly detection model. The reference image feature may be any image feature in the feature bank. In an embodiment of the present disclosure, the Euclidean distance between the target image feature and the reference image feature in the feature bank is calculated. By way of example but not limitation, methods such as the distance-based K-nearest neighbor algorithm (Nested-loop Method), and the local distance-based outlier factor (LDOF) method may be used to calculate the distance between the target image feature and the reference image feature in the feature bank. When the calculated distance is greater than or equal to the first distance threshold, it is determined that an abnormal region exists in the target object.

[0042] According to the above technical solution, the distance between the target image feature and the reference image feature in the feature bank is calculated by using the anomaly determination module. The calculated distance is compared with the first distance threshold. When the distance is greater than or equal to the first distance threshold, it is determined that an abnormal region exists in the target object. This method does not require a complex determination process. Instead, by comparing the calculated distance between the target image feature and the reference image feature in the feature bank with the first distance threshold, it

is determined whether an abnormal region exists in the target object. Therefore, the method offers strong reliability and high efficiency.

[0043] In an embodiment, the feature bank includes respective reference image patch features corresponding to multiple image positions, and the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object includes: for each image patch of the multiple image patches: calculating, by using the anomaly determination module, a distance between the target image patch feature corresponding to the image patch and a reference image patch feature among the reference image patch features in the feature bank, where the reference image patch feature corresponds to the image position of the image patch; and determining, in a case that the distance is greater than or equal to a second distance threshold, that an abnormal region exists within the image patch.

[0044] In an embodiment, the to-be-detected image is divided into multiple image patches. For example, for a to-be-detected image with an image size of 16×16 , the to-be-detected image is divided into four image patches, each with a size of 4×4 . Each image patch corresponds to a respective image position in the to-be-detected image. The feature bank may include respective reference image patch features corresponding to multiple image positions. That is, the feature bank includes respective reference image patch features for the four image patches. For each image patch of the multiple image patches, the distance between the target image patch feature corresponding to the image patch and the corresponding reference image patch feature in the feature bank is calculated by using the anomaly determination module. When the distance is greater than or equal to the second distance threshold, it is determined that an abnormal region exists in the image patch. For specific implementation, reference may be made to the relevant description of determining whether an abnormal region exists in the target object based on the target image feature in the foregoing embodiments. For brevity, details are not repeated here. The second distance threshold may be the same as or different from the first distance threshold, which is not limited in the present disclosure.

[0045] According to the above technical solution, for each image patch of the multiple image patches, the distance between the target image patch feature corresponding to the image patch and the reference image patch feature in the feature bank is calculated by using the anomaly determination module. The calculated distance is compared with the second distance threshold,

and when the distance is greater than or equal to the second distance threshold, it is determined that an abnormal region exists in the image patch. This method does not require a complex process to determine whether an abnormal region exists in the target object, and thus is efficient. In addition, this method can be applied to anomaly determination on respective target features
5 corresponding to all image patches included in the to-be-detected image to detect whether an abnormal region exists in each image patch, thereby further improving the reliability of the anomaly detection result.

[0046] In an embodiment, the trained anomaly detection model is obtained by: acquiring multiple second sample images; inputting the multiple second sample images into a feature
10 extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features corresponding to the multiple second sample images; calculating a prediction loss based on differences between the sample image features corresponding to the multiple second sample images; and optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained
15 anomaly detection model.

[0047] In an embodiment, the method for acquiring the second sample images is similar to the method for acquiring the to-be-detected image. The method for acquiring the to-be-detected image is described in detail in step S110, and for brevity, details are not repeated here. Similar to the to-be-detected image, by inputting the multiple second sample images into the feature
20 extraction module in a to-be-trained anomaly detection model respectively, respective sample image features corresponding to the multiple second sample images are obtained. Based on the differences between the sample image features corresponding to the multiple second sample images, a prediction loss is calculated. For example, a difference Δz between a sample image feature Z_1 and a sample image feature Z_2 may be calculated by: $\Delta z = (Z_1 - Z_2)^2$. For any two
25 sample image features, the difference between these two sample image features may be substituted into a preset loss function for loss calculation to obtain a prediction loss. By way of example but not limitation, the preset loss function may be a Fourier loss function. Then, based on the prediction loss, a parameter of the initial anomaly detection model may be optimized using backpropagation and gradient descent algorithms. The optimization operation may be
30 repeatedly performed until the anomaly detection model reaches a converged state. After training is complete, the obtained anomaly detection model is used for subsequent anomaly detection of the to-be-detected image.

[0048] According to the above technical solution, the anomaly detection model is trained based on the differences between respective sample image features corresponding to the multiple second sample images. The anomaly detection model trained in this way is applicable to anomaly detection in scenarios with variable backgrounds, such as textures of materials like metal. In addition, through normalizing flow processing, anomaly detection at high resolutions is achieved, while further enabling anomaly detection results to be determined for different feature scales of the to-be-detected image.

[0049] In an embodiment, the trained anomaly detection model is obtained by: acquiring multiple second sample images; inputting the multiple second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features which are outputted by the stacked flow network and correspond to the multiple second sample images, where each of the sample image features includes multi-scale sub-sample image features; calculating a prediction loss based on differences between the sample image features corresponding to the multiple second sample images; and optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model, where the calculating a prediction loss based on differences between the sample image features corresponding to the multiple second sample images includes: calculating, for each feature scale among feature scales of the sample image features, a sub-prediction loss for the feature scale based on differences between respective sub-sample image features of the feature scale corresponding to the multiple second sample images; and summing or averaging the sub-prediction losses for all the feature scales of the sample image features to obtain the prediction loss.

[0050] In an embodiment, the implementation of the process which includes: acquiring multiple second sample images; inputting the multiple second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features which are outputted by the stacked flow network and correspond to the multiple second sample images, where each of the sample image features includes multi-scale sub-sample image features; calculating a prediction loss based on differences between the sample image features corresponding to the multiple second sample images; and optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model, is described in detail in the foregoing embodiments, and for brevity, details are not repeated here. For each feature scale

among feature scales of the sample image features, a sub-prediction loss for the feature scale is calculated based on differences between respective sub-sample image features of the feature scale corresponding to the multiple second sample images. For example, a sample image feature Z_1 may include a sub-sample image feature Z_{11} , a sub-sample image feature Z_{12} , a sub-sample image feature Z_{13} , and a sub-sample image feature Z_{14} , and a sample image feature Z_2 may include a sub-sample image feature Z_{21} , a sub-sample image feature Z_{22} , a sub-sample image feature Z_{23} , and a sub-sample image feature Z_{24} . The sub-sample image feature Z_{11} and the sub-sample image feature Z_{21} are of the same feature scale, the sub-sample image feature Z_{12} and the sub-sample image feature Z_{22} are of the same feature scale, and so on. For sub-sample image features at each feature scale, a respective sub-prediction loss is calculated. For example, the difference ΔZ_1 between the sub-sample image feature Z_{11} and the sub-sample image feature Z_{21} is calculated by $\Delta Z_1 = (Z_{11} - Z_{21})^2$. Based on the calculated difference and a preset loss function, a sub-prediction loss for the feature scale is calculated. By using a similar method, for each feature scale, the difference between two sub-sample image features is calculated, and the sub-prediction loss is calculated. For brevity, details are not repeated here. The prediction loss may be obtained by summing the sub-prediction losses for all feature scales of the sample image features. In addition, the prediction loss may further be obtained by averaging the sub-prediction losses for all feature scales of the sample image features.

[0051] According to the above technical solution, sub-prediction losses are calculated based on the sub-sample image features, that is, the sample image features divided across different feature scales. By summing or averaging the sub-prediction losses for all feature scales of the sample image features, a prediction loss is obtained. The anomaly detection model trained in this way further improves the accuracy of anomaly detection for the target object at different feature scales.

[0052] In an embodiment, in the process of summing or averaging the sub-prediction losses for all the feature scales of the sample image features, the sub-prediction losses for the feature scales are provided with respective weighting factors.

[0053] In an embodiment, the sub-prediction losses for different feature scales are provided with respective weighting factors. Thus, the importance of each feature scale during the training process of the anomaly detection model is adjusted by using these weighting factors to improve the training efficiency of the anomaly detection model.

[0054] In an embodiment, before the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, the method further includes: acquiring the multiple first sample images; inputting the multiple first sample images into the feature extraction module in the trained anomaly detection model respectively to obtain respective sample image features corresponding to the multiple first sample images; performing clustering on the sample image features corresponding to the multiple first sample images; and for each cluster of clusters obtained through the clustering, determining a representative sample image feature in the cluster as a reference image feature to obtain the feature bank.

[0055] In an embodiment, the method for acquiring the first sample images is similar to the method for acquiring the to-be-detected image. The method for acquiring the to-be-detected image is described in detail in step S110, and for brevity, details are not repeated here. Multiple first sample images are inputted into the feature extraction module in the trained anomaly detection model respectively to obtain respective sample image features corresponding to the multiple first sample images. Then, clustering is performed on the sample image features corresponding to the multiple first sample images. Representative sample image features in all clusters obtained through clustering are determined as reference image features to obtain the feature bank. Clustering algorithms may include but are not limited to K-Means algorithm, K-Means++ algorithm, Minibatch K-Means algorithm, and any other type of clustering algorithm. For example, after feature extraction is performed on any one of the multiple first sample images, the sample image feature corresponding to the first sample image may be obtained as $10 \times 10 \times C_3$. If the number of the multiple first sample images is 100, then the total number of sample image features of all 100 first sample images is 10,000 (that is, $100 \times 10 \times 10$), and the channel count of each sample image feature is C_3 . By using the clustering algorithm, clustering is performed on these 10,000 sample image features, and the top 10% of sample image features in the clustering result are selected as reference image features. That is, the number of reference image features is 1,000. Multiple reference image features constitute a feature bank (Memory Bank).

[0056] According to the above technical solution, the multiple first sample images are inputted into the feature extraction module in the trained anomaly detection model respectively, so as to obtain respective sample image features corresponding to the multiple first sample images. Clustering is performed on the sample image features corresponding to the multiple

first sample images, and representative sample image features in all clusters obtained through clustering are determined as reference image features to obtain the feature bank. The reference image features in the feature bank obtained in this way indicate anomalies in respective clusters. The obtained target image features are compared with the feature bank. Therefore, the accuracy of the anomaly detection result is improved.

[0057] According to another aspect of the present disclosure, an apparatus for anomaly detection is further provided. FIG. 3 is a schematic block diagram of an apparatus for anomaly detection according to an embodiment of the present disclosure. As shown in FIG. 3, the apparatus 300 includes an acquisition module 310, a first input module 320, and a second input module 330.

[0058] The acquisition module 310 is configured to acquire a to-be-detected image, where the to-be-detected image includes a target object.

[0059] The first input module 320 is configured to input the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature, where the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing.

[0060] The second input module 330 is configured to input the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, where the anomaly detection result is to indicate whether an abnormal region exists in the target object, the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with a feature bank to determine the anomaly detection result, and the feature bank is a sample feature bank obtained by the feature extraction module performing feature extraction on multiple first sample images.

[0061] Those skilled in the art can understand the specific implementation and beneficial effects of the above apparatus for anomaly detection by reading the above related description of the method for anomaly detection. For brevity, details are not repeated here.

[0062] According to another aspect of the present disclosure, an electronic device is further provided. FIG. 4 is a schematic block diagram of an electronic device according to an

embodiment of the present disclosure. As shown in FIG. 4, the electronic device includes a processor 410 and a memory 420. The memory 420 stores a computer program instruction, and the computer program instruction, when executed by the processor 410, performs the above method for anomaly detection.

5 **[0063]** According to another aspect of the present disclosure, a storage medium storing a computer program/instruction is further provided. The storage medium may include, for example, a storage component of a tablet computer, a hard disk of a personal computer, an erasable programmable read-only memory (EPROM), a compact disc read-only memory (CD-ROM), a USB memory, or any combination of the above storage media. The storage medium
10 may be any combination of one or more computer-readable storage media. The computer program/instruction, when executed by a processor, performs the above method for anomaly detection.

[0064] Those skilled in the art can understand the specific implementation of the electronic device and the storage medium described above by reading the above related description of the
15 method for anomaly detection. For brevity, details are not repeated here.

[0065] Embodiments

[0066] Embodiment 1

A method for anomaly detection is provided, and the method comprises:

acquiring a to-be-detected image, where the to-be-detected image comprises a target object;

20 inputting the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature; and

inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, where the anomaly detection result is to indicate whether an abnormal region exists in the target object,
25 and where

the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing, the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with
30 a feature bank to determine the anomaly detection result, and the feature bank is a sample

feature bank obtained by the feature extraction module performing feature extraction on a plurality of first sample images.

[0067] Embodiment 2

The method described in embodiment 1 is provided, where the feature extraction module
5 comprises an initial feature extraction network, a stacked flow network, and a feature fusion network, and the inputting the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature comprises:

inputting the to-be-detected image into the initial feature extraction network to obtain a first image feature, where the first image feature is of multi-scale;

10 inputting the first image feature into the stacked flow network and mapping the first image feature to a latent space by using the stacked flow network to obtain a second image feature, where the second image feature is of the same feature scale as the first image feature; and

inputting the second image feature into the feature fusion network to obtain the target image feature,

15 where the stacked flow network and the feature fusion network are configured to perform the normalizing flow processing.

[0068] Embodiment 3

The method described in embodiment 1 or 2 is provided, where the to-be-detected image comprises a plurality of image patches, the inputting the to-be-detected image into the initial
20 feature extraction network to obtain a first image feature comprises:

inputting the plurality of image patches into the initial feature extraction network respectively to obtain respective first image patch features corresponding to the plurality of image patches, where the first image feature comprises the first image patch features corresponding to the plurality of image patches,

25 the inputting the first image feature into the stacked flow network and mapping the first image feature to a latent space by using the stacked flow network to obtain a second image feature comprises:

inputting the first image patch features corresponding to the plurality of image patches into the stacked flow network respectively and mapping the first image patch features corresponding

to the plurality of image patches to the latent space by using the stacked flow network to obtain respective second image patch features corresponding to the plurality of image patches, where the second image feature comprises the second image patch features corresponding to the plurality of image patches, and

5 the inputting the second image feature into the feature fusion network to obtain the target image feature comprises:

 inputting the second image patch features corresponding to the plurality of image patches into the feature fusion network to obtain respective target image patch features corresponding to the plurality of image patches, where the target image feature comprises the target image patch features corresponding to the plurality of image patches.

10

[0069] Embodiment 4

The method described in any one of embodiments 1 to 3 is provided, where the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object comprises:

15 calculating, by using the anomaly determination module, a distance between the target image feature and a reference image feature in the feature bank; and

 determining, in a case that the distance is greater than or equal to a first distance threshold, that an abnormal region exists in the target object.

[0070] Embodiment 5

20 The method described in any one of embodiment 1 to 4 is provided, where the feature bank comprises respective reference image patch features corresponding to a plurality of image positions, and the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object comprises:

25 for each image patch of the plurality of image patches:

 calculating, by using the anomaly determination module, a distance between the target image patch feature corresponding to the image patch and a reference image patch feature among the reference image patch features in the feature bank, where the reference image patch feature corresponds to the image position of the image patch; and

determining, in a case that the distance is greater than or equal to a second distance threshold, that an abnormal region exists within the image patch.

[0071] Embodiment 6

5 The method described in any one of embodiments 1 to 5 is provided, where the trained anomaly detection model is obtained by:

acquiring a plurality of second sample images;

inputting the plurality of second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features corresponding to the plurality of second sample images;

10 calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images; and

optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model.

[0072] Embodiment 7

15 The method described in any one of embodiments 1 to 6 is provided, where the trained anomaly detection model is obtained by:

acquiring a plurality of second sample images;

20 inputting the plurality of second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features which are outputted by the stacked flow network and correspond to the plurality of second sample images, where each of the sample image features comprises multi-scale sub-sample image features;

calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images; and

25 optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model,

where the calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images comprises:

calculating, for each feature scale among feature scales of the sample image features, a sub-prediction loss for the feature scale based on differences between respective sub-sample image features of the feature scale corresponding to the plurality of second sample images; and

5 summing or averaging the sub-prediction losses for all the feature scales of the sample image features to obtain the prediction loss.

[0073] Embodiment 8

The method described in any one of embodiments 1 to 7 is provided, where in the process of summing or averaging the sub-prediction losses for all the feature scales of the sample image features, the sub-prediction losses for the feature scales are provided with respective weighting
10 factors.

[0074] Embodiment 9

The method described in any one of embodiments 1 to 8 is provided, where before the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, the method further
15 comprises:

acquiring the plurality of first sample images;

inputting the plurality of first sample images into the feature extraction module in the trained anomaly detection model respectively to obtain respective sample image features corresponding to the plurality of first sample images;

20 performing clustering on the sample image features corresponding to the plurality of first sample images; and

for each cluster of clusters obtained through the clustering, determining a representative sample image feature in the cluster as a reference image feature to obtain the feature bank.

[0075] Embodiment 10

25 An apparatus for anomaly detection is provided, and the apparatus comprises:

an acquisition module, configured to acquire a to-be-detected image, where the to-be-detected image comprises a target object;

a first input module, configured to input the to-be-detected image into a feature extraction

module in a trained anomaly detection model to obtain a target image feature; and

a second input module, configured to input the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, where the anomaly detection result is to indicate whether an abnormal region exists in the target object, and where

the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing, and the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with a feature bank to determine the anomaly detection result.

[0076] Embodiment 11

An electronic device, comprising a processor and a memory, is provided. The memory stores a computer program instruction, and the computer program instruction, when executed by the processor, performs the method for anomaly detection described in any one of embodiments 1 to 9.

[0077] Embodiment 12

A storage medium, storing a computer program/instruction, is provided. The computer program/instruction, when executed, performs the method for anomaly detection described in any one of embodiments 1 to 9.

[0078] Although the embodiments have been described herein with reference to the drawings, it should be understood that the above embodiments are merely exemplary and are not intended to limit the scope of the present disclosure. Various changes and modifications may be made by those skilled in the art without departing from the scope and spirit of the present disclosure. All such changes and modifications are deemed to fall into the scope claimed by the claims of the present disclosure.

[0079] Those skilled in the art may realize that the units and algorithm steps of each example described in conjunction with the embodiments herein may be implemented by electronic hardware, or a combination of computer software and electronic hardware. Whether these functions are performed in hardware or software depends on the specific application and design constraints of the technical solution. Those skilled in the art may use different methods to

implement the described functions for each specific application, but such implementation should not be regarded as exceeding the scope of the present disclosure.

[0080] It should be understood that the devices and methods disclosed in the embodiments of the present disclosure may be implemented in other manners. For example, the device
5 embodiments described hereinabove are only schematic. For example, the division of the units is only based on logical functions, and there may be other division manners in practical implementation, such as multiple units or components may be combined or integrated into another device, or some features may be neglected or not executed.

[0081] In the specification provided herein, a lot of specific details are explained. However,
10 it may be understood that the embodiments of the present disclosure may be practiced without these specific details. In some embodiments, well-known methods, structures and techniques are not shown in detail in order to avoid obscuring understanding of the specification.

[0082] Similarly, it should be understood that in order to simplify the present disclosure and facilitate understanding one or more of the various aspects of the disclosure, in the description
15 of the exemplary embodiments of the present disclosure, the various features of the present disclosure are sometimes grouped together into a single embodiment, figure, or description thereof. However, the method according to the present disclosure should not be interpreted as reflecting the following intention: the claimed disclosure requires more features than those explicitly stated in each claim. More specifically, as reflected in the corresponding claims, the
20 inventive point is that the corresponding technical problem can be solved with features fewer than all the features of a single embodiment being disclosed. Therefore, the claims following the detailed description are hereby expressly incorporated into the detailed description, where each claim itself serves as an independent embodiment of the present disclosure.

[0083] It can be understood by those skilled in the art that, except for mutually exclusive
25 features, all features disclosed in this specification (including the claims, abstract and drawings) and all processes or units of any method or device disclosed in the specification may be combined in any way. Unless otherwise expressly stated, each feature disclosed in this specification (including the claims, abstract and drawings) may be replaced by an alternative feature that provides the same, equivalent or similar purpose.

[0084] In addition, it can be understood by those skilled in the art that, although some
30 embodiments described herein include certain features included in other embodiments instead

of other features, the combination of features of different embodiments is meant to be within the scope of the present disclosure and form different embodiments. For example, in the claims, any one of the claimed embodiments may be used in any combination.

[0085] The various component embodiments of the present disclosure may be implemented in hardware, or in software modules running on one or more processors, or in a combination thereof. It can be understood by those skilled in the art that a microprocessor or a digital signal processor (DSP) may be adopted in practice to realize some or all of the functions of some of the modules of the apparatus for anomaly detection according to the embodiments of the present disclosure. The present disclosure may alternatively be implemented as a device program (such as a computer program and computer program product) for performing part or all of the methods described herein. Such a program implementing the present disclosure may be stored on a computer-readable medium, or may be in the form of one or more signals. Such a signal may be downloaded from an Internet website, or provided on a carrier signal, or provided in any other form.

[0086] It should be noted that the above embodiments illustrate rather than limit the present disclosure, and alternative embodiments may be designed by those skilled in the art without departing from the scope of the claims. In the claims, any reference symbol between parentheses shall not be construed as a limitation of the claims. The word "comprising/including" does not exclude the presence of elements or steps not listed in the claims. The word "a" or "an" preceding an element does not exclude the presence of multiple of such elements. The present disclosure may be implemented by means of hardware including several different elements and by means of a suitably programmed computer. In a unit claim enumerating several apparatuses, several of these apparatuses may be embodied by a same hardware item. The use of the words "first", "second", "third" and the like does not indicate any order. These words may be interpreted as names.

[0087] Described above are only specific embodiments of the present disclosure or explanation for the specific embodiments of the present disclosure, and the protection scope of the present disclosure is not limited thereto. Any changes or alternations that can be easily envisaged by those familiar with the art within the technical scope disclosed in the present disclosure shall fall within the protection scope of the present disclosure. The protection scope of the present disclosure shall be in accordance with the protection scope of the claims.

CLAIMS

1. A method for anomaly detection, comprising:

acquiring a to-be-detected image, wherein the to-be-detected image comprises a target
5 object;

inputting the to-be-detected image into a feature extraction module in a trained anomaly
detection model to obtain a target image feature; and

inputting the target image feature into an anomaly determination module in the trained
anomaly detection model to obtain an anomaly detection result of the target object, wherein the
10 anomaly detection result is to indicate whether an abnormal region exists in the target object,
and wherein

the feature extraction module is configured to extract a multi-scale image feature of the to-
be-detected image and convert the multi-scale image feature into the target image feature of a
preset scale through normalizing flow processing, the anomaly determination module is
15 configured to compare the target image feature outputted by the feature extraction module with
a feature bank to determine the anomaly detection result, and the feature bank is a sample
feature bank obtained by the feature extraction module performing feature extraction on a
plurality of first sample images.

2. The method according to claim 1, wherein the feature extraction module comprises an
20 initial feature extraction network, a stacked flow network, and a feature fusion network, and the
inputting the to-be-detected image into a feature extraction module in a trained anomaly
detection model to obtain a target image feature comprises:

inputting the to-be-detected image into the initial feature extraction network to obtain a
first image feature, wherein the first image feature is of multi-scale;

25 inputting the first image feature into the stacked flow network and mapping the first image
feature to a latent space by using the stacked flow network to obtain a second image feature,
wherein the second image feature is of the same feature scale as the first image feature; and

inputting the second image feature into the feature fusion network to obtain the target
image feature,

wherein the stacked flow network and the feature fusion network are configured to perform the normalizing flow processing.

3. The method according to claim 2, wherein the to-be-detected image comprises a plurality of image patches, the inputting the to-be-detected image into the initial feature extraction network to obtain a first image feature comprises:

inputting the plurality of image patches into the initial feature extraction network respectively to obtain respective first image patch features corresponding to the plurality of image patches, wherein the first image feature comprises the first image patch features corresponding to the plurality of image patches,

the inputting the first image feature into the stacked flow network and mapping the first image feature to a latent space by using the stacked flow network to obtain a second image feature comprises:

inputting the first image patch features corresponding to the plurality of image patches into the stacked flow network respectively and mapping the first image patch features corresponding to the plurality of image patches to the latent space by using the stacked flow network to obtain respective second image patch features corresponding to the plurality of image patches, wherein the second image feature comprises the second image patch features corresponding to the plurality of image patches, and

the inputting the second image feature into the feature fusion network to obtain the target image feature comprises:

inputting the second image patch features corresponding to the plurality of image patches into the feature fusion network to obtain respective target image patch features corresponding to the plurality of image patches, wherein the target image feature comprises the target image patch features corresponding to the plurality of image patches.

4. The method according to claim 1 or 2, wherein the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object comprises:

calculating, by using the anomaly determination module, a distance between the target image feature and a reference image feature in the feature bank; and

determining, in a case that the distance is greater than or equal to a first distance threshold,

that an abnormal region exists in the target object.

5 5. The method according to claim 3, wherein the feature bank comprises respective reference image patch features corresponding to a plurality of image positions, and the inputting the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object comprises:

for each image patch of the plurality of image patches:

10 calculating, by using the anomaly determination module, a distance between the target image patch feature corresponding to the image patch and a reference image patch feature among the reference image patch features in the feature bank, wherein the reference image patch feature corresponds to the image position of the image patch; and

determining, in a case that the distance is greater than or equal to a second distance threshold, that an abnormal region exists within the image patch.

6. The method according to claim 1 or 2, wherein the trained anomaly detection model is obtained by:

15 acquiring a plurality of second sample images;

inputting the plurality of second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features corresponding to the plurality of second sample images;

20 calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images; and

optimizing a parameter of the feature extraction module in the to-be-trained anomaly detection model based on the prediction loss to obtain the trained anomaly detection model.

7. The method according to claim 3, wherein the trained anomaly detection model is obtained by:

25 acquiring a plurality of second sample images;

inputting the plurality of second sample images into a feature extraction module in a to-be-trained anomaly detection model respectively to obtain respective sample image features which are outputted by the stacked flow network and correspond to the plurality of second sample images, wherein each of the sample image features comprises multi-scale sub-sample

image features;

calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images; and

optimizing a parameter of the feature extraction module in the to-be-trained anomaly
5 detection model based on the prediction loss to obtain the trained anomaly detection model,

wherein the calculating a prediction loss based on differences between the sample image features corresponding to the plurality of second sample images comprises:

calculating, for each feature scale among feature scales of the sample image features, a
sub-prediction loss for the feature scale based on differences between respective sub-sample
10 image features of the feature scale corresponding to the plurality of second sample images; and

summing or averaging the sub-prediction losses for all the feature scales of the sample image features to obtain the prediction loss.

8. The method according to claim 7, wherein in the process of summing or averaging the
sub-prediction losses for all the feature scales of the sample image features, the sub-prediction
15 losses for the feature scales are provided with respective weighting factors.

9. The method according to any one of claims 1 to 3, wherein before the inputting the
target image feature into an anomaly determination module in the trained anomaly detection
model to obtain an anomaly detection result of the target object, the method further comprises:

acquiring the plurality of first sample images;

20 inputting the plurality of first sample images into the feature extraction module in the
trained anomaly detection model respectively to obtain respective sample image features
corresponding to the plurality of first sample images;

performing clustering on the sample image features corresponding to the plurality of first
sample images; and

25 for each cluster of clusters obtained through the clustering, determining a representative
sample image feature in the cluster as a reference image feature to obtain the feature bank.

10. An apparatus for anomaly detection, comprising:

an acquisition module, configured to acquire a to-be-detected image, wherein the to-be-

detected image comprises a target object;

a first input module, configured to input the to-be-detected image into a feature extraction module in a trained anomaly detection model to obtain a target image feature; and

5 a second input module, configured to input the target image feature into an anomaly determination module in the trained anomaly detection model to obtain an anomaly detection result of the target object, wherein the anomaly detection result is to indicate whether an abnormal region exists in the target object, and wherein

10 the feature extraction module is configured to extract a multi-scale image feature of the to-be-detected image and convert the multi-scale image feature into the target image feature of a preset scale through normalizing flow processing, and the anomaly determination module is configured to compare the target image feature outputted by the feature extraction module with a feature bank to determine the anomaly detection result.

11. An electronic device, comprising: a processor and a memory, wherein the memory stores a computer program instruction, and the computer program instruction, when executed
15 by the processor, performs the method according to any one of claims 1 to 9.

12. A storage medium, storing a computer program/instruction, wherein the computer program/instruction, when executed, performs the method according to any one of claims 1 to 9.