CLAIMS

1. A texture generation method, comprising:

acquiring a query feature map, a key feature map, and a value feature map of an image to be transformed, wherein the query feature map comprises a plurality of pixel points to be queried in a texture space; the key feature map is used for representing a semantic feature of the image to be transformed; the value feature map is used for determining, in combination with an attention feature map of each of the pixel points to be queried, a texture parameter of each of the pixel points to be queried; the attention feature map is determined by the query feature map and the key feature map; and

generating a target texture feature map from the query feature map, the key feature map, and the value feature map based on an attention mechanism, wherein the target texture feature map comprises the texture parameter corresponding to each of the pixel points to be queried;

wherein the generating a target texture feature map from the query feature map, the key feature map, and the value feature map based on an attention mechanism comprises:

performing feature extraction processing at different scales on an input feature map of the attention mechanism to obtain input features corresponding to a plurality of scales, wherein the input feature map comprises the query feature map, the key feature map, and the value feature map;

inputting an input feature corresponding to each of the scales into a corresponding transforming module to obtain an output feature corresponding to each of the scales; and

fusing the output features corresponding to all of the scales to obtain the target texture feature map;

wherein the performing feature extraction processing at different scales on an input feature map of the attention mechanism to obtain input features corresponding to a plurality of scales comprises:

performing downsampling processing on the input feature map for a plurality of times to obtain an input feature corresponding to a preset scale; and

performing upsampling processing on the input feature corresponding to the preset scale for a plurality of times to obtain input features corresponding to the plurality of scales, wherein a number of times of the downsampling processing is the same as a number of times of the upsampling processing.

2. The method of claim 1, wherein the inputting an input feature corresponding to each of the scales into a corresponding transforming module to obtain an output feature corresponding to each of the scales comprises:

for each scale, fusing the query feature map and a position code of the query feature map to obtain a first input feature;

fusing the key feature map and a position code of the key feature map to obtain a second input feature;

obtaining an intermediate feature map based on the first input feature, the second input feature, and the value feature map; and

determining the output feature corresponding to the scale based on the intermediate feature map and a residual network.

3. The method of claim 1, wherein the fusing the output features corresponding to all of the scales to obtain the target texture feature map comprises:

fusing the output features corresponding to all of the scales to obtain a multi-scale output feature; and

determining the target texture feature map based on the multi-scale output feature.

4. The method of claim 3, wherein the fusing the output features corresponding to all of the scales to obtain a multi-scale output feature comprises:

performing downsampling fusion processing on the output feature corresponding to each of the scales in turn according to an order from a largest scale to a smallest scale of the plurality of scales to obtain a downsampling fusion feature corresponding to each of the scales, wherein the downsampling fusion processing is used for fusing the output feature corresponding to the scale and a downsampling fusion feature obtained from a last downsampling fusion processing to obtain a downsampling intermediate feature, and performing downsampling processing on the downsampling intermediate feature to obtain a downsampling fusion feature corresponding to the scale; and

performing upsampling fusion processing on the output feature corresponding to each of the scales in turn according to an order from the smallest scale to the largest scale of the plurality of scales to obtain an upsampling fusion feature corresponding to each of the scales, wherein the upsampling fusion processing is used for fusing the output feature corresponding to the scale and an upsampling fusion feature obtained from a last upsampling fusion processing to obtain an upsampling intermediate feature, and performing upsampling processing on the upsampling intermediate feature to obtain an upsampling fusion feature corresponding to the scale,

wherein the upsampling fusion processing of the smallest scale is used for performing upsampling processing on a downsampling fusion feature corresponding to the smallest scale, and fusing the output feature corresponding to the smallest scale, to obtain an upsampling fusion feature corresponding to the smallest scale; and an upsampling fusion feature corresponding to the largest scale is a multi-scale output feature.

5. The method of claim 3, wherein in a case where the multi-scale output feature comprises a pixel feature map, the value feature map comprises an original pixel feature of the image to be transformed; the determining the target texture feature map based on the multi-scale output feature comprises:

taking the pixel feature map as the target texture feature map.

6. The method of claim 3, wherein in a case where the multi-scale output feature comprises a texture flow feature map, the value feature map comprises a pixel distribution feature of the image to be transformed; the determining the target texture feature map based on the multi-scale output feature comprises:

generating the target texture feature map based on the texture flow feature map and the image to be transformed.

7. The method of claim 3, wherein in a case where the multi-scale output feature comprises a pixel feature map, a texture flow feature map and a confidence feature map, the value feature map comprises an original pixel feature and a pixel distribution feature of the image to be transformed; the determining the target texture feature map based on the multi-scale output feature comprises:

determining confidence information of each of the pixel points to be queried in the texture space based on the confidence feature map, wherein the confidence information comprises a first confidence of the pixel point to be queried corresponding to the pixel feature map and a second confidence of the pixel point to be queried corresponding to the texture flow feature map; and

fusing the pixel feature map and the texture flow feature map based on the first confidence and the second confidence of each of the pixel points to be queried, to obtain the target texture feature map.

8. The method of claim 7, wherein the fusing the pixel feature map and the texture flow feature map based on the first confidence and the second confidence of each of the pixel points to be queried, to obtain the target texture feature map comprises:

determining a first texture parameter corresponding to each of the pixel points to be queried based on the pixel feature map;

determining a second texture parameter corresponding to each of the pixel points to be queried based on the texture flow feature map and the image to be transformed; and

for each of the pixel points to be queried, performing weighted fusion on the first texture parameter and the second texture parameter based on the first confidence and the second confidence respectively to obtain a texture parameter corresponding to the pixel point to be queried.

9. The method of any one of claims 1 to 8, wherein the method is implemented by a trained texture generation model; a training process of the texture generation model comprises:

acquiring an sample image, and inputting the sample image into a texture generation model to be trained to obtain a predicted texture image corresponding to the sample image;

rendering the predicted texture image corresponding to the sample image to obtain a predicted rendered image corresponding to the sample image;

determining a loss value of the texture generation model to be trained based on the sample image and the predicted rendered image; and

adjusting, according to the loss value, a parameter of the texture generation model to be trained to obtain the trained texture generation model.

10. The method of claim 9, wherein the loss value of the texture generation model to be trained comprises a first loss value;

wherein the determining a loss value of the texture generation model to be trained based on the sample image and the predicted rendered image comprises:

extracting a first character feature in the sample image and a second character feature in the predicted rendered image respectively based on a person re-identification model; and

determining the first loss value based on a feature distance between the first character feature and the second character feature;

wherein the adjusting, according to the loss value, a parameter of the texture generation model to be trained to obtain the trained texture generation model comprises:

adjusting, based on at least the first loss value, the parameter of the texture generation model to be trained to obtain the trained texture generation model.

11. The method of claim 9, wherein the loss value of the texture generation model to be trained comprises a second value;

wherein the determining a loss value of the texture generation model to be trained based on the sample image and the predicted rendered image comprises:

extracting a first character feature in the sample image and a second character feature in the predicted rendered image respectively based on a person re-identification model;

acquiring a first segmentation result corresponding to the sample image, wherein the first segmentation result is used for representing first distribution information of each body part in the sample image;

acquiring a second segmentation result corresponding to the predicted rendered image, wherein the second segmentation result is used for representing second distribution information of each body part in the predicted rendered image;

for each body part, determining a part loss value corresponding to the body part based on a feature distance between a first part feature and a second part feature of the body part, wherein the first part feature is determined based on the first character feature and first part information of the body part, and the second part feature is determined based on the second character feature and second part information of the body part; and

determining the second loss value based on the part loss value corresponding to the each body part;

wherein the adjusting, according to the loss value, a parameter of the texture generation model to be trained to obtain the trained texture generation model comprises:

adjusting, based on at least the second loss value, the parameter of the texture generation model to be trained to obtain the trained texture generation model.

12. The method of claim 9, wherein the loss value of the texture generation model to be trained comprises a third loss value, the method further comprises:

acquiring a part texture image corresponding to a target body part in the predicted texture image;

acquiring a standard texture image corresponding to the target body part, wherein the standard texture image is a fusion image of part texture images of different characters regarding the target body part; and

determining the third loss value based on a structural similarity index between the part texture image and the standard texture image;

wherein the adjusting, according to the loss value, a parameter of the texture generation model to be trained to obtain the trained texture generation model comprises:

adjusting, based on at least the third loss value, the parameter of the texture generation model to be trained to obtain the trained texture generation model.

13. A texture generation apparatus, comprising:

an acquisition module, configured to acquire a query feature map, a key feature map, and a value feature map of an image to be transformed, the query feature map comprising a plurality of pixel points to be queried in a texture space, the key feature map being used for representing a semantic feature of the image to be transformed, the value feature map being used for determining, in combination with an attention feature map of each of the pixel points

to be queried, a texture parameter of each of the pixel points to be queried, and the attention feature map being determined by the query feature map and the key feature map; and

a generation module, configured to generate a target texture feature map from the query feature map, the key feature map, and the value feature map based on an attention mechanism, and the target texture feature map comprising the texture parameter corresponding to each of the pixel points to be queried;

wherein the generation module is further configured to:

perform feature extraction processing at different scales on an input feature map of the attention mechanism to obtain input features corresponding to a plurality of scales, wherein the input feature map comprises the query feature map, the key feature map, and the value feature map;

input an input feature corresponding to each of the scales into a corresponding transforming module to obtain an output feature corresponding to each of the scales; and

fuse the output features of corresponding to all of the scales to obtain a target texture feature map;

wherein the generation module is configured to:

perform downsampling processing on the input feature map for a plurality of times to obtain an input feature corresponding to a preset scale; and

perform upsampling processing on the input feature corresponding to the preset scale for a plurality of times to obtain input features corresponding to the plurality of scales, wherein a number of times of the downsampling processing is the same as a number of times of the upsampling processing.

14. The apparatus of claim 13, further comprising a training module,

wherein the training module is configured to:

acquire a sample image, and input the sample image into a texture generation model to be trained to obtain a predicted texture image corresponding to the sample image;

render the predicted texture image corresponding to the sample image to obtain a predicted rendered image corresponding to the sample image;

determine a loss value of the texture generation model to be trained based on the sample image and the predicted rendered image; and

adjust, according to the loss value, a parameter of the texture generation model to be trained to obtain the trained texture generation model.

15. A texture generation device, comprising:

a memory, configured to store an executable instruction; and

a processor, configured to execute the executable instruction stored in the memory to:

acquire a query feature map, a key feature map, and a value feature map of an image to be transformed, wherein the query feature map comprises a plurality of pixel points to be queried in a texture space; the key feature map is used for representing a semantic feature of the image to be transformed; the value feature map is used for determining, in combination with an attention feature map of each of the pixel points to be queried, a texture parameter of each of the pixel points to be queried; the attention feature map is determined by the query feature map and the key feature map; and

generate a target texture feature map from the query feature map, the key feature map, and the value feature map based on an attention mechanism, wherein the target texture feature map comprises the texture parameter corresponding to each of the pixel points to be queried;

wherein the processor is configured to:

perform feature extraction processing at different scales on an input feature map of the attention mechanism to obtain input features corresponding to a plurality of scales, wherein

the input feature map comprises the query feature map, the key feature map, and the value feature map;

input an input feature corresponding to each of the scales into a corresponding transforming module to obtain an output feature corresponding to each of the scales; and

fuse the output features corresponding to all of the scales to obtain the target texture feature map;

wherein the processor is configured to:

perform downsampling processing on the input feature map for a plurality of times to obtain an input feature corresponding to a preset scale; and

perform upsampling processing on the input feature corresponding to the preset scale for a plurality of times to obtain input features corresponding to the plurality of scales, wherein a number of times of the downsampling processing is the same as a number of times of the upsampling processing.

16. A computer readable storage medium, storing an executable instruction, wherein the executable instruction, when executed by a processor, is configured to:

acquire a query feature map, a key feature map, and a value feature map of an image to be transformed, wherein the query feature map comprises a plurality of pixel points to be queried in a texture space; the key feature map is used for representing a semantic feature of the image to be transformed; the value feature map is used for determining, in combination with an attention feature map of each of the pixel points to be queried, a texture parameter of each of the pixel points to be queried; the attention feature map is determined by the query feature map and the key feature map; and

generate a target texture feature map from the query feature map, the key feature map, and the value feature map based on an attention mechanism, wherein the target texture feature map comprises the texture parameter corresponding to each of the pixel points to be queried;

wherein the executable instruction is configured to:

perform feature extraction processing at different scales on an input feature map of the attention mechanism to obtain input features corresponding to a plurality of scales, wherein the input feature map comprises the query feature map, the key feature map, and the value feature map;

input an input feature corresponding to each of the scales into a corresponding transforming module to obtain an output feature corresponding to each of the scales; and

fuse the output features corresponding to all of the scales to obtain the target texture feature map;

wherein the executable instruction is configured to:

perform downsampling processing on the input feature map for a plurality of times to obtain an input feature corresponding to a preset scale; and

perform upsampling processing on the input feature corresponding to the preset scale for a plurality of times to obtain input features corresponding to the plurality of scales, wherein a number of times of the downsampling processing is the same as a number of times of the upsampling processing.

17. A computer program, comprising a computer program code, wherein when the computer program code is executed by an electronic device, a processor in the electronic device is configured to:

acquire a query feature map, a key feature map, and a value feature map of an image to be transformed, wherein the query feature map comprises a plurality of pixel points to be queried in a texture space; the key feature map is used for representing a semantic feature of the image to be transformed; the value feature map is used for determining, in combination with an attention feature map of each of the pixel points to be queried, a texture parameter of each of the pixel points to be queried; the attention feature map is determined by the query feature map and the key feature map; and

generate a target texture feature map from the query feature map, the key feature map, and the value feature map based on an attention mechanism, wherein the target texture feature map comprises the texture parameter corresponding to each of the pixel points to be queried;

wherein the processor is configured to:

perform feature extraction processing at different scales on an input feature map of the attention mechanism to obtain input features corresponding to a plurality of scales, wherein the input feature map comprises the query feature map, the key feature map, and the value feature map;

input an input feature corresponding to each of the scales into a corresponding transforming module to obtain an output feature corresponding to each of the scales; and

fuse the output features corresponding to all of the scales to obtain the target texture feature map;

wherein the processor is configured to:

perform downsampling processing on the input feature map for a plurality of times to obtain an input feature corresponding to a preset scale; and

perform upsampling processing on the input feature corresponding to the preset scale for a plurality of times to obtain input features corresponding to the plurality of scales, wherein a number of times of the downsampling processing is the same as a number of times of the upsampling processing.