

CLAIMS

1. An image style transform method, comprising:

acquiring an initial image to be subjected to style transform;

5 inputting a gradient of the initial image to an image style transform model, and obtaining a feature map of the initial image in a gradient domain from the image style transform model, wherein the image style transform model is obtained by being trained in the gradient domain based on a pixel-wise loss and a perceptual loss; and

10 performing image reconstruction according to the feature map of the initial image in the gradient domain to obtain a style image.

2. The method according to claim 1, further comprising: training the image style transform model, wherein a training target of the image style transform model is that a total loss L_{total} is minimum, wherein L_{total} is represented by the following equation:

15 $L_{total} = \alpha L_{feat} + \beta L_{pixel}$, where L_{feat} represents the perceptual loss, L_{pixel} represents the pixel-wise loss, and values of α and β are real numbers.

3. The method according to claim 1, wherein the image style transform model comprises a pixel-wise loss model and a perceptual loss model, wherein the pixel-wise loss model is obtained by taking minimization of the pixel-wise loss as a training target when being trained in the gradient domain, and the perceptual loss model is obtained by taking minimization of the perceptual loss as a training target when being trained in the gradient domain training.

25 4. The method according to claim 3, wherein the training process of the pixel-wise loss model and the perceptual loss model comprises: inputting a gradient of a training sample to the pixel-wise loss model, and obtaining a sample output result of the training sample from the pixel-wise loss model;

30 determining a gradient of a stylized reference image corresponding to the training sample; and

training the perceptual loss model according to a first output feature map of the gradient of the reference image in a j-th convolutional layer of the perceptual loss model and a second output feature map of the sample output result in the j-th

convolutional layer of the perceptual loss model.

5. The method according to claim 4, wherein the training the perceptual loss model according to a first output feature map of the gradient of the reference image in a j-th convolutional layer of the perceptual loss model and a second output feature map of the sample output result in the j-th convolutional layer of the perceptual loss model comprises: training the perceptual loss model by using the following equation:

$$L_{feat} = \frac{1}{C_j H_j W_j} \|\psi_j(\partial \mathcal{L}(I_i)) - \psi_j(F_W(\partial I_i))\|,$$

where ∂I_i represents a gradient of an i-th training sample, F_W represents the pixel-wise loss model, $F_W(\partial I_i)$ represents an output result of the gradient of the i-th training sample through the pixel-wise loss model, $\partial \mathcal{L}(I_i)$ represents a gradient of a stylized reference image of the i-th training sample; $\psi_j(\cdot)$ represents the output feature map of the j-th convolutional layer when the perceptual loss model adopts a convolutional neural network model, and $C_j H_j W_j$ respectively represent the number of channels, the height, and the width of the feature map corresponding to the j-th convolutional layer.

6. The method according to claim 3, wherein the training process of the pixel-wise loss model comprises: using the gradient of the training sample as an input of the pixel-wise loss model, and obtaining a sample output result from the pixel-wise loss model;

determining the gradient of the stylized reference image corresponding to the training sample; and

training the pixel-wise loss model according to the gradient of the reference image and the sample output result.

7. The method according to claim 4, wherein the pixel-wise loss model comprises a first convolutional layer set, an up-sampling layer and a second convolutional layer set, and wherein the training the pixel-wise loss model according to the gradient of the reference image and the sample output result comprises:

inputting the gradient of the training sample to the first convolutional layer set to obtain the sample feature map;

inputting the sample feature map to the up-sampling layer, and up-sampling the sample feature map to a pixel size of the initial image; and

inputting the up-sampled sample feature map to the second convolutional layer set to obtain the sample output result.

5

8. The method according to claim 7, wherein the training the pixel-wise loss model according to the gradient of the reference image and the sample output result comprises: training the pixel-wise loss model according to an absolute value of a difference between $F_W(\partial I_i)$ and corresponding $\partial \mathcal{L}(I_i)$ of each training sample;

10

where ∂I_i represents the gradient of the i-th training sample, F_W represents the pixel-wise loss model, $F_W(\partial I_i)$ represents an output result of the gradient of the i-th training sample through the pixel-wise loss model F_W , and $\partial \mathcal{L}(I_i)$ represents the gradient of the stylized reference image of the i-th training sample.

15

9. The method according to claim 8, wherein the training the pixel-wise loss model according to an absolute value of a difference between $F_W(\partial I_i)$ and corresponding $\partial \mathcal{L}(I_i)$ of each training sample comprises: training the pixel-wise loss model by using the following equation: $L_{\text{pixel}} = \frac{1}{D} \sum_{i=0}^{D-1} \left\{ \frac{1}{2} \|F_W(\partial I_i) - \partial \mathcal{L}(I_i)\|^2 \right\}$,

20

where ∂I_i represents the gradient of the i-th training sample, F_W represents the pixel-wise loss model, $F_W(\partial I_i)$ represents an output result of the gradient of the i-th training sample through the pixel-wise loss model F_W , $\partial \mathcal{L}(I_i)$ represents the gradient of the stylized reference image of the i-th training sample, and D represents a number of samples in a training sample set.

25

10. The method according to any one of claims 1 to 9, wherein the performing image reconstruction according to the feature map of the initial image in the gradient domain to obtain a style image comprises: using an image that satisfies a structural similarity condition to the feature map of the initial image in the gradient domain as the style image;

30

wherein satisfying the structural similarity condition to the feature map of the initial image in the gradient domain satisfies comprises:

a degree of structural difference between the style image and the initial image

is less than a similarity threshold, or the degree of structural difference between the style image and the initial image is the smallest, wherein the degree of structural difference is a variation trend of the style image in the gradient domain and the feature map of the initial image in the gradient domain in at least one reference direction.

5

11. The method according to claim 10, wherein the performing image reconstruction according to the feature map of the initial image in the gradient domain to obtain a style image comprises: performing image reconstruction according to $\|\partial_x S - F_W(\partial_x I)\|^2 + \|\partial_y S - F_W(\partial_y I)\|^2$ to obtain the style image, where $\partial_x I$ represents the gradient of the initial image in an x direction, $F_W(\partial_x I)$ represents a feature map of the gradient of the initial image in the x direction in the gradient domain through the image style transform model, $\partial_y I$ represents the gradient of the initial image in a y direction, $F_W(\partial_y I)$ represents a feature map of the gradient of the initial image in the y direction in the gradient domain through the image style transform model, $\partial_x S$ represents the gradient of the style image in the x direction, and $\partial_y S$ represents the gradient of the style image in the y direction.

12. The method according to any one of claims 1 to 9, wherein the performing image reconstruction according to the feature map of the initial image in the gradient domain to obtain a style image comprises: performing image reconstruction according to color information of the initial image and the feature map of the initial image in the gradient domain to obtain the style image.

13. The method according to claim 12, wherein the performing image reconstruction according to color information of the initial image and the feature map of the initial image in the gradient domain to obtain a style image comprises: using an image that satisfies a structural similarity condition to the feature map of the initial image in the gradient domain and an image that satisfies a color similarity condition to the initial image as the style image.

30

14. The method according to claim 12, further comprising: performing feature extraction on the initial image to obtain a face region in the initial image; and

correspondingly, the performing image reconstruction according to color information of the initial image and the feature map of the initial image in the gradient domain to obtain a style image comprises: using an image that satisfies a structural similarity condition to the feature map of the initial image in the gradient domain, and an image that satisfies a color similarity condition to a face region in the initial image as the style image.

15. The method according to claim 13, wherein the using an image that satisfies a structural similarity condition to the feature map of the initial image in the gradient domain and an image that satisfies a color similarity condition to the initial image as the style image comprises: performing image reconstruction according to $\|S - I\| + \lambda \left\{ \|\partial_x S - F_W(\partial_x I)\|^2 + \|\partial_y S - F_W(\partial_y I)\|^2 \right\}$ to obtain the style image, where I represents the initial image, S represents the style image, $\partial_x I$ represents the gradient of the initial image in the x direction, $F_W(\partial_x I)$ represents a feature map of the gradient of the initial image in the x direction in the gradient domain through the image style transform model, $\partial_y I$ represents the gradient of the initial image in the y direction, $F_W(\partial_y I)$ represents a feature map of the gradient of the initial image in the y direction in the gradient domain through the image style transform model, $\partial_x S$ represents the gradient of the style image in the x direction, and $\partial_y S$ represents the gradient of the style image in the y direction.

16. The method according to any one of claims 1 to 9, wherein the inputting a gradient of the initial image to an image style transform model, and obtaining a feature map of the initial image in a gradient domain from the image style transform model comprises: determining a gradient of the initial image in at least one reference direction;

inputting the gradient in the at least one reference direction to the image style transform model, and correspondingly obtaining a feature map of the initial image in the at least one reference direction in the gradient domain from the image style transform model; and

correspondingly performing image reconstruction according to the feature map in the at least one reference direction in the gradient domain to obtain the style image.

17. The method according to claim 16, wherein the at least one reference direction comprises x and y directions in a plane reference coordinate system, and correspondingly determining the gradients of the initial image in the x and y directions, respectively;

respectively inputting the gradients in the x and y directions to the image style transform model, and correspondingly obtaining a feature map of the initial image in the x and y directions in the gradient domain from the image style transform model; and

correspondingly performing image reconstruction according to the feature map in the x and y directions in the gradient domain to obtain the style image.

18. An image style transform apparatus, comprising:

an acquisition unit, configured to acquire an initial image to be subjected to style transform;

an obtaining unit, configured to input a gradient of the initial image to an image style transform model, and obtain a feature map of the initial image in a gradient domain from the image style transform model, wherein the image style transform model is obtained by being trained in the gradient domain based on a pixel-wise loss and a perceptual loss; and

a reconstruction unit, configured to perform image reconstruction according to the feature map of the initial image in the gradient domain to obtain a style image.

19. The apparatus according to claim 18, further comprising:

a training unit, configured to train the image style transform model, wherein a training target of the image style transform model is that a total loss L_{total} is minimum, wherein L_{total} is represented by the following equation: $L_{total} = \alpha L_{feat} + \beta L_{pixel}$, where L_{feat} represents the perceptual loss, L_{pixel} represents the pixel-wise loss, and values of α and β are real numbers.

20. The apparatus according to claim 18, wherein the image style transform model comprises a pixel-wise loss model and a perceptual loss model, wherein the pixel-wise loss model is obtained by taking minimization of the pixel-wise loss as a training

target when being trained in the gradient domain, and the perceptual loss model is obtained by taking minimization of the perceptual loss as a training target when being trained in the gradient domain training.

5 21. The apparatus according to claim 20, wherein the training unit comprises:

a first input module, configured to input a gradient of a training sample to the pixel-wise loss model, and obtain a sample output result of the training sample from the pixel-wise loss model;

10 a first determining module, configured to determine a gradient of a stylized reference image corresponding to the training sample; and

a first training module, configured to train the perceptual loss model according to a first output feature map of the gradient of the reference image in a j-th convolutional layer of the perceptual loss model and a second output feature map of the sample output result in the j-th convolutional layer of the perceptual loss model.

15

22. The apparatus according to claim 21, wherein the first training module trains the perceptual loss model by using the following equation: $L_{feat} = \frac{1}{C_j H_j W_j} \|\psi_j(\partial \mathcal{L}(I_i)) - \psi_j(F_W(\partial I_i))\|$,

20 where ∂I_i represents a gradient of an i-th training sample, F_W represents the pixel-wise loss model, $F_W(\partial I_i)$ represents an output result of the gradient of the i-th training sample through the pixel-wise loss model, $\partial \mathcal{L}(I_i)$ represents a gradient of a stylized reference image of the i-th training sample; $\psi_j(\cdot)$ represents the output feature map of the j-th convolutional layer when the perceptual loss model adopts a convolutional neural network model, and $C_j H_j W_j$ respectively represent the number of
25 channels, the height, and the width of the feature map corresponding to the j-th convolutional layer.

23. The apparatus according to claim 20, wherein the training unit comprises:

30 a second determining module, configured to determine a gradient of a training sample;

a second input module, configured to use the gradient of the training sample as an input of the pixel-wise loss model, and obtain a sample output result from the

pixel-wise loss model;

a third determining module, configured to determine the gradient of the stylized reference image corresponding to the training sample; and

5 a second training module, configured to train the pixel-wise loss model according to the gradient of the reference image and the sample output result.

24. The apparatus according to claim 21, wherein the pixel-wise loss model comprises a first convolutional layer set, an up-sampling layer, and a second convolutional layer set; and the second training module comprises: a first input sub-module, configured to
10 input the gradient of the training sample to the first convolutional layer set to obtain a sample feature map;

an up-sampling sub-module, configured to input the sample feature map to the up-sampling layer, and up-sample the sample feature map to a pixel size of the initial image; and

15 a second input sub-module, configured to input the up-sampled sample feature map to the second convolutional layer set to obtain the sample output result.

25. The apparatus according to claim 24, wherein the second training module is configured to train the pixel-wise loss model according to an absolute value of a
20 difference between $F_W(\partial I_i)$ and corresponding $\partial \mathcal{L}(I_i)$ of each training sample, where ∂I_i represents the gradient of the i-th training sample, F_W represents the pixel-wise loss model, $F_W(\partial I_i)$ represents an output result of the gradient of the i-th training sample through the pixel-wise loss model F_W , and $\partial \mathcal{L}(I_i)$ represents the gradient of the stylized reference image of the i-th training sample.

25

26. The apparatus according to claim 25, wherein the second training module is configured to train the pixel-wise loss model by using the following equation:

$$L_{\text{pixel}} = \frac{1}{D} \sum_{i=0}^{D-1} \left\{ \frac{1}{2} \|F_W(\partial I_i) - \partial \mathcal{L}(I_i)\|^2 \right\},$$

where ∂I_i represents the gradient of the i-th training sample, F_W represents the
30 pixel-wise loss model, $F_W(\partial I_i)$ represents an output result of the gradient of the i-th training sample through the pixel-wise loss model F_W , $\partial \mathcal{L}(I_i)$ represents the gradient of the stylized reference image of the i-th training sample, and D represents a number

of samples in a training sample set.

27. The apparatus according to any one of claims 18 to 26, wherein the reconstruction unit is configured to use an image that satisfies a structural similarity condition to the feature map of the initial image in the gradient domain as the style image, wherein satisfying the structural similarity condition to the feature map of the initial image in the gradient domain satisfies comprises: a degree of structural difference between the style image and the initial image is less than a similarity threshold, or the degree of structural difference between the style image and the initial image is the smallest, wherein the degree of structural difference is a variation trend of the style image in the gradient domain and the feature map of the initial image in the gradient domain in at least one reference direction.

28. The apparatus according to claim 27, wherein the reconstruction unit is configured to perform image reconstruction according to $\|\partial_x S - F_W(\partial_x I)\|^2 + \|\partial_y S - F_W(\partial_y I)\|^2$ to obtain the style image, where $\partial_x I$ represents the gradient of the initial image in an x direction, $F_W(\partial_x I)$ represents a feature map of the gradient of the initial image in the x direction in the gradient domain through the image style transform model, $\partial_y I$ represents the gradient of the initial image in a y direction, $F_W(\partial_y I)$ represents a feature map of the gradient of the initial image in the y direction in the gradient domain through the image style transform model, $\partial_x S$ represents the gradient of the style image in the x direction, and $\partial_y S$ represents the gradient of the style image in the y direction.

29. The apparatus according to any one of claims 18 to 26, wherein the reconstruction unit is configured to perform image reconstruction according to color information of the initial image and the feature map of the initial image in the gradient domain to obtain a style image.

30. The apparatus according to claim 29, wherein the reconstruction unit is configured to use an image that satisfies a structural similarity condition to a feature map of the initial image in the gradient domain and an image that satisfies a color similarity

condition to the initial image as the style image.

31. The apparatus according to claim 29, further comprising: an extraction unit, configured to perform feature extraction on the initial image to obtain a face region in the initial image; and correspondingly, the reconstruction unit is configured to use an image that satisfies a structural similarity condition to the feature map of the initial image in the gradient domain, and an image that satisfies a color similarity condition to a face region in the initial image as the style image.

32. The apparatus according to claim 30, wherein the reconstruction unit is configured to perform image reconstruction according to $\|S - I\| + \lambda \left\{ \|\partial_x S - F_W(\partial_x I)\|^2 + \|\partial_y S - F_W(\partial_y I)\|^2 \right\}$ to obtain the style image,

where I represents the initial image, S represents the style image, $\partial_x I$ represents the gradient of the initial image in the x direction, $F_W(\partial_x I)$ represents a feature map of the gradient of the initial image in the x direction in the gradient domain through the image style transform model, $\partial_y I$ represents the gradient of the initial image in the y direction, $F_W(\partial_y I)$ represents a feature map of the gradient of the initial image in the y direction in the gradient domain through the image style transform model, $\partial_x S$ represents the gradient of the style image in the x direction, and $\partial_y S$ represents the gradient of the style image in the y direction.

33. The apparatus according to any one of claims 18 to 26, wherein the obtaining unit comprises:

a fourth determining module, configured to determine a gradient of the initial image in at least one reference direction; and

an obtaining module, configured to input the gradient in the at least one reference direction to the image style transform model, and correspondingly obtain a feature map of the initial image in the at least one reference direction in the gradient domain from the image style transform model;

correspondingly, the reconstruction unit is configured to perform image reconstruction according to the feature map in the at least one reference direction in the gradient domain to obtain the style image.

34. The apparatus according to claim 33, wherein the at least one reference direction comprises x and y directions in a plane reference coordinate system, and correspondingly, the determining unit is configured to determine the gradients of the initial image in the x and y directions, respectively; the obtaining unit is configured to respectively input the gradients in the x and y directions to the image style transform model, and correspondingly obtain a feature map of the initial image in the x and y directions in the gradient domain from the image style transform model; and correspondingly, the reconstruction unit is configured to perform image reconstruction according to the feature map in the x and y directions in the gradient domain to obtain the style image.

35. A computer program product, comprising a computer executable instruction, wherein the computer executable instruction is executed to implement the operations of the image style transform method according to any one of claims 1 to 17.

36. A computer storage medium, having a computer executable instruction stored thereon, wherein the computer executable instruction is executed to implement the operations of the image style transform method according to any one of claims 1 to 17.

37. A computer device, comprising a memory and a processor, wherein the memory stores a computer program that, when executed by the processor, cause the processor to implement the operations of the image style transform method according to any one of claims 1 to 17.