CLAIMS

1. A method of training a neural network model using at least one processor, the neural network model comprising a first neural network portion configured to perform image classification and a second neural network portion configured to perform image localization and/or segmentation, the method comprising:

training, at a first training round, the first neural network portion based on a training dataset to obtain first backbone parameters of the first neural network portion corresponding to the first training round;

applying, at the first training round, the first backbone parameters of the first neural network portion obtained corresponding to the first training round to the second neural network portion;

training, at the first training round, the second neural network portion based on the training dataset to obtain second backbone parameters of the second neural network portion corresponding to the first training round; and

performing, after the first training round, a plurality of additional training rounds, each additional training round comprising:

updating, at the additional training round, the first backbone parameters of the first neural network portion based on the second backbone parameters obtained corresponding to the first training round or an immediately previous additional training round;

training, at the additional training round, the first neural network portion based on the training dataset to obtain the first backbone parameters of the first neural network portion corresponding to the additional training round;

updating, at the additional training round, the second backbone parameters of the second neural network portion based on the first backbone parameters obtained corresponding to the additional training round; and

training, at the additional training round, the second neural network portion based on the training dataset to obtain the second backbone parameters of the second neural network portion corresponding to the additional training round.

2. The method according to claim 1, wherein

said updating, at the additional training round, the first backbone parameters of the first neural network portion is further based on a first backbone sharing parameter; and

said updating, at the additional training round, the second backbone parameters of the second neural network portion is further based on a second backbone sharing parameter.

3. The method according to claim 2, wherein

the first backbone sharing parameter is configured to control an amount of the second backbone parameters obtained for updating the first backbone parameters of the first neural network portion, and

the second backbone sharing parameter is configured to control an amount of the first backbone parameters obtained for updating the second backbone parameters of the second neural network portion.

4. The method according to any one of claims 1 to 3, wherein said performing the plurality of additional training rounds comprises performing a number of additional training rounds until each of the first and second neural network portions converges.

5. The method according to any one of claims 1 to 4, wherein

the first and second neural network portions each comprises a Swin Transformer as a backbone thereof,

the first backbone parameters comprise first weight parameters; and

the second backbone parameters comprise second weight parameters.

6. The method according to any one of claims 1 to 5, wherein

the first neural network portion comprises a plurality of task classifiers configured to perform a plurality of classification tasks, respectively,

the training dataset comprises labelled images comprising multi-attribute labelled images, and

the first neural network portion is trained for the plurality of classification tasks simultaneously.

7. The method according to any one of claims 1 to 6, further comprising performing fine-tuning of the first and second neural network portions comprising:

performing, for each labelled image of a subset of labelled images of the training dataset, a low-frequency spectral alignment of the labelled image with respect to an image of a subset

of images of a second dataset to obtain a low-frequency aligned labelled image, thereby obtaining a low-frequency aligned training data subset.

8. The method according to claim 7, wherein said performing, for each labelled image of the subset of labelled images of the training dataset, the low-frequency spectral alignment of the labelled image with respect to the image of the subset of images of the second dataset comprises:

obtaining a Fourier transform of the labelled image and a Fourier transform of the image;

obtaining a magnitude spectrum and a phase spectrum of the Fourier transform of the labelled image and obtaining a magnitude spectrum of the Fourier transform of the image; and

performing the low-frequency spectral alignment of the labelled image based on the magnitude spectrum and the phase spectrum of the Fourier transform of the labelled image, the magnitude spectrum of Fourier transform of the image and a first alignment parameter.

9. The method according to claim 8, wherein said performing the low-frequency spectral alignment of the labelled image comprises:

modifying the magnitude spectrum of the Fourier transform of the labelled image based on the magnitude spectrum of the Fourier transform of the image and the first alignment parameter to obtain a modified magnitude spectrum of the Fourier transform of the labelled image; and

performing an inverse Fourier transform based on the modified magnitude spectrum and the phase spectrum of the Fourier transform of the labelled image to obtain the low-frequency aligned labelled image.

10. The method according to claim 8 or 9, wherein

the first alignment parameter is configured to control an amount of the magnitude spectrum of the Fourier transform of the image obtained for modifying the magnitude spectrum of the Fourier transform of the labelled image, and

the first alignment parameter is defined as a function of a frequency point in a Fourier space and based on whether the frequency point corresponds to a low-frequency point.

11. The method according to any one of claims 7 to 10, wherein said performing fine-tuning of the first and second neural network portions further comprises:

performing, for each image of a plurality of images of the subset of images of the second dataset, self-supervised labelling of the image to obtain a pseudo-labelled image, thereby obtaining a subset of pseudo-labelled images; and

performing, for each pseudo-labelled image of the subset of pseudo-labelled images, a low-frequency spectral alignment of the pseudo-labelled image with respect to a labelled image of the subset of labelled images of the training dataset to obtain a low-frequency aligned pseudo-labelled image, thereby obtaining a low-frequency aligned second data subset.

12. The method according to claim 11, wherein said performing, for each pseudo-labelled image of the subset of pseudo-labelled images, the low-frequency spectral alignment of the pseudo-labelled image with respect to the labelled image of the subset of labelled images of the training dataset comprises:

obtaining a Fourier transform of the pseudo-labelled image and a Fourier transform of the labelled image;

obtaining a magnitude spectrum and a phase spectrum of the Fourier transform of the pseudo-labelled image and obtaining a magnitude spectrum of the Fourier transform of the labelled image; and

performing the low-frequency spectral alignment of the pseudo-labelled image based on the magnitude spectrum and the phase spectrum of the Fourier transform of the pseudo-labelled image, the magnitude spectrum of Fourier transform of the labelled image and a second alignment parameter.

13. The method according to claim 12, wherein said performing the low-frequency spectral alignment of the pseudo-labelled image comprises:

modifying the magnitude spectrum of the Fourier transform of the pseudo-labelled image based on the magnitude spectrum of the Fourier transform of the labelled image and the second alignment parameter to obtain a modified magnitude spectrum of the Fourier transform of the pseudo-labelled image; and

performing an inverse Fourier transform based on the modified magnitude spectrum and the phase spectrum of the Fourier transform of the pseudo-labelled image to obtain the low-frequency aligned pseudo-labelled image.

14. The method according to claim 12 or 13, wherein

the second alignment parameter is configured to control an amount of the magnitude spectrum of the Fourier transform of the labelled image obtained for modifying the magnitude spectrum of the Fourier transform of the pseudo-labelled image, and

the second alignment parameter is defined as a function of a frequency point in a Fourier space and based on whether the frequency point corresponds to a low-frequency point.

15. The method according to any one of claims 11 to 14, wherein said performing fine-tuning of the first and second neural network portions comprises performing a plurality of fine-tuning rounds, each of the plurality of fine-tuning rounds comprising:

shuffling the labelled images in the training dataset and the images in the second dataset;

extracting a plurality of subsets of images from the second dataset, and for each subset of images, extracting a subset of labelled images from the training dataset to form a subset pair of the subset of images and the subset of labelled images, thereby forming a plurality of subset pairs; and

for each subset pair of the plurality of subset pairs:

performing, for each labelled image of the subset of labelled images of the subset pair, said low-frequency spectral alignment of the labelled image with respect to an image of the subset of images of the subset pair to obtain the low-frequency aligned labelled image, thereby obtaining the low-frequency aligned training data subset;

performing, for each image of a plurality of images of the subset of images of the subset pair, said self-supervised labelling of the image to obtain the pseudo-labelled image, thereby obtaining the subset of pseudo-labelled images;

performing, for each pseudo-labelled image of the subset of pseudo-labelled images, said low-frequency spectral alignment of the pseudo-labelled image with respect to a labelled image of the subset of labelled images of the subset pair to obtain the low-frequency aligned pseudo-labelled image, thereby obtaining the low-frequency aligned second data subset; and

training the first and second neural network portions based on the low-frequency aligned training data subset and the low-frequency aligned second data subset.

16. The method according to claim 10 or 14, wherein the frequency point is determined to correspond to a low-frequency point if the frequency point corresponds to a frequency component in a lowest 1% to a lowest 10% of frequency components of an image being subjected to the low-frequency spectral alignment.

17. The method according to any one of claims 7 to 16, wherein the training dataset and the second dataset are obtained from different sources.

18. The method according to any one of claims 1 to 17, wherein

the second neural network portion is configured to perform building defect localization and/or building defect segmentation, and

the training dataset comprises labelled structural images for training the first neural network portion to perform image classification and for training the second neural network portion to perform building defect localization and/or building defect segmentation.

19. A system for training a neural network model, the neural network model comprising a first neural network portion configured to perform image classification and a second neural network portion configured to perform image localization and/or segmentation, the system comprising:

at least one memory; and

at least one processor communicatively coupled to the at least one memory and configured to perform the method of training a neural network model according to any one of claims 1 to 18.

20. A computer program product, embodied in one or more non-transitory computer-readable storage mediums, comprising instructions executable by at least one processor to perform the method of training a neural network model according to any one of claims 1 to 18.

21. A method of using the neural network model trained according to any one of claims 1 to 18, the method comprising:

receiving, by the neural network model, an input image;

performing, by the neural network model, image classification of the input image; and

performing, by the neural network model, image localization and/or segmentation of the input image,

wherein said performing image classification of the input image and said performing image localization and/or segmentation of the input image are performed simultaneously.