

CLAIMS

1. A method for training a neural network, comprising:
Forming an autoencoder comprising the neural network as encoder and comprising a decoder;
For each training image of multiple training images
 Generating a latent representation of the training image by the encoder;
 Transforming the training image;
 Masking the latent representation;
 Supplying information about the transformation, the masked latent representation and the unmasked latent representation to the decoder to generate a decoder output for the training image; and
Adjusting the encoder and the decoder to reduce a loss between the transformed training images and the decoder outputs.
2. The method of claim 1, comprising subdividing the training image into a plurality of training image patches, wherein the latent representation comprises an encoding for each training image patch and wherein masking the latent representation comprises replacing at least some of the training image patches by mask tokens.
3. The method of claim 2, comprising randomly selecting the training image patches replaced by mask tokens.
4. The method of claim 3, comprising adjusting the mask tokens and the encoder and the decoder to reduce the loss between the transformed training images and the decoder outputs.
5. The method of any one of claims 1 to 4, wherein the loss between the transformed training images and the decoder outputs comprises a mean-square-error loss, a cosine distance or a Kullback-Leibler divergence of the transformed training images and the decoder outputs.

6. The method of any one of claims 1 to 5, wherein the transformation comprises a feature extraction of the training image followed by a homography transformation.
7. The method of any one of claims 1 to 5, wherein the transformation is a homography transformation of the training image.
8. The method of any one of claims 1 to 7, wherein the information about the transformation is an encoding of hyper parameters of the transformation.
9. The method of claim 8, comprising generating the encoding of hyper parameters of the transformation by a further neural network.
10. The method of claim 9, comprising adjusting the further neural network and the encoder and the decoder to reduce the loss between the transformed training images and the decoder outputs.
11. The method of any one of claims 1 to 10, wherein the neural network is a convolutional neural network, a vision transformer network or a multi-layer perceptron-based neural network.
12. A training device configured to perform the method of any one of claims 1 to 11.
13. A computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 11.
14. A computer-readable medium comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 11.