Claims

1. A method for encoding a picture using a neural network, NN, wherein the NN comprises a plurality of layers comprising at least one downsampling layer (2110, 2120, 2130, 2140) that applies a downsampling to an input (2101, 2102, 2103, 2104) representing a matrix having a size S in at least one dimension, the method comprising, applying (2321), before processing (2311) the input with the at least one downsampling layer, a rescaling to the input, wherein the rescaling comprises changing the size S in the at least one dimension so that $\bar{S} = n \cdot r$, with $n \in \mathbb{N}$ and $r > 1, r \in \mathbb{N}$, where r is the downsampling ratio of the layer; after the rescaling, processing (2311) the input by the downsampling layer (2110, 2120, 2130, 2140) and providing (2312) an output with the size $S_1 < S$; and providing, after processing the picture using the NN, a bitstream as output of the NN.

2. The method according to claim 1, wherein the NN comprises a number of $M \in \mathbb{N}$ downsampling layers, wherein each downsampling layer m (2110, 2120, 2130, 2140), $m \leq M, m \in \mathbb{N}$ has an associated downsampling ratio $r_m > 1, r_m \in \mathbb{N}$ and the method further comprises: before processing (2311) an input (2101, 2102, 2103, 2104) representing a matrix having a size S_m in at least one dimension with a downsampling layer m, applying, if the size S_m is not an integer multiple of the downsampling ratio r_m of the downsampling layer m, a rescaling to the input, wherein the rescaling comprises changing the size S_m in the at least one dimension so that $\overline{S_m} = n \cdot r_m, n \in \mathbb{N}$; and/or

   wherein, before applying (2321) the rescaling to the input (2101, 2102, 2103, 2104) with the size S_m a determination is made whether S_m is an integer multiple of the downsampling ratio r_m and, if it is determined that S_m is not an integer multiple of the downsampling ratio r_m, the rescaling is applied to the input so that the size S_m is changed in the at least one dimension so that $\overline{S_m} = n \cdot r_m, n \in \mathbb{N}$.

3. The method according to claim 1 or 2, wherein the size $\overline{S_m}$ is determined using a function comprising at least one of ceil, int, floor.

4. The method according to claim 3, wherein:

   - the size $\overline{S_m}$ is determined using $floor\left(\frac{S_m}{r_m}\right) \cdot r_m = \overline{S_m}$; or

   - the size $\overline{S_m}$ is determined using $ceil\left(\frac{S_m}{r_m}\right) \cdot r_m = \overline{S_m}$; or

- the size $\overline{S_m}$ is determined using $int\left(\frac{S_m}{r_m}\right) \cdot r_m = \overline{S_m}$.

5. The method according to any of claims 1 to 4, wherein the input (2101, 2102, 2103, 2104) to a downsampling layer m has a size S_m in the at least one dimension that has a value that is between a closest smaller integer multiple of the downsampling ratio r_m of the downsampling layer m (2110, 2120, 2130, 2140) and a closest larger integer multiple of the downsampling ratio r_m of the downsampling layer m (2110, 2120, 2130, 2140) and wherein, depending on a condition, the size S_m of the input is changed during the rescaling to either match the closest smaller integer multiple of the downsampling ratio r_m or to match the closest larger integer multiple of the downsampling ratio r_m.

6. The method according to any of claims 1 to 5, wherein the input (2101, 2102, 2103, 2104) to a downsampling layer m has a size S_m in the at least one dimension that has a value that is not an integer multiple of the downsampling ratio r_m of the downsampling layer m (2110, 2120, 2130, 2140), wherein the size S_m of the input is changed during the rescaling to either match the closest smaller integer multiple of the downsampling ratio r_m or to match the closest larger integer multiple of the downsampling ratio r_m.

7. The method according to any of claims 1 to 6, wherein, if the size S_m of the input (2101, 2102, 2103, 2104) is closer to the closest larger integer multiple of the downsampling ratio r_m of the downsampling layer m (2110, 2120, 2130, 2140) than to the closest smaller integer multiple of the downsampling ratio r_m, the size S_m of the input is increased to a size $\overline{S_m}$ that matches the closest larger integer multiple of the downsampling ratio r_m, wherein increasing the size S_m of the input (2101, 2102, 2103, 2104) to the size $\overline{S_m}$ comprises padding the input with the size S_m with zeros or with padding information obtained from the input with the size S_m..

8. The method according to claim 7, wherein the padding information obtained from the input with the size S_m is applied as redundant padding information to increase the size S_m of the input (2101, 2102, 2103, 2104) to the size $\overline{S_m}$ and wherein the padding with redundant padding information comprises at least one of reflection padding and repetition padding.

9. A method for decoding a bitstream representing a picture using a neural network, NN, wherein the NN comprises a plurality of layers comprising at least a first upsampling layer and a second upsampling layer that apply an upsampling to an input (2201, 2202, 2203, 2204) representing a matrix having a size T in at least one dimension, the method

comprising processing the input by the first upsampling layer and providing an output of the first upsampling layer, wherein the output has a size $\overline{T_1} = T \cdot u_1, u_1 > 1, u_1 \in \mathbb{N}$, wherein u_1 is the upsampling ratio of the first upsampling layer; applying, before processing the output of the first upsampling layer by the second upsampling layer, a rescaling (2421) to the output of the first upsampling layer, wherein the rescaling comprises changing the size $\overline{T_1}$ of the output in the at least one dimension to a size $\widehat{T_1}$ in the at least one dimension based on information obtained; processing the rescaled output by the second upsampling layer and providing an output of the second upsampling layer, wherein the output has a size $\overline{T_2} = \widehat{T_1} \cdot u_2, u_2 > 1, u_2 \in \mathbb{N}$ in the at least one dimension, wherein u_2 is the upsampling ratio of the second upsampling layer; and providing, after processing the bitstream using the NN, a decoded picture as output of the NN.

10. The method according to claim 9, wherein the upsampling layer comprises a transposed convolution or convolution layer.

11. The method according to claim 9 or 10, wherein the information comprises at least one of a target size of the decoded picture comprising at least one of a height H of the decoded picture and a width W of the decoded picture, the upsampling ratio u_1, the upsampling ratio u_2, a target output size $\overline{T_2}$ of the second upsampling layer, the size $\widehat{T_1}$.

12. The method according to any of claims 9 to 11, wherein the information is obtained from at least one of: the bitstream, a second bitstream, information available at the decoder.

13. The method according to any one of claims 9 to 12, wherein the formula is given by $\widehat{T_1} = ceil\left(\frac{T_{output}}{u^N}\right)$, wherein T_{output} is the target size of the output of the NN and u is an upsampling ratio.

14. The method according to any one of claims 9 to 13, wherein the method further comprises determining whether $\overline{T_1}$ is larger than $\widehat{T_1}$ or whether $\overline{T_1}$ is smaller than $\widehat{T_1}$, and wherein if it is determined that $\overline{T_1}$ is larger than $\widehat{T_1}$, the rescaling comprises applying a cropping to the output with the size $\overline{T_1}$ such that the size $\overline{T_1}$ is reduced to the size $\widehat{T_1}$; and

    if it is determined that $\overline{T_1}$ is smaller than $\widehat{T_1}$, the rescaling comprises applying a padding to the output with the size $\overline{T_1}$ such that the size $\overline{T_1}$ is increased to the size $\widehat{T_1}$, wherein

the padding comprises padding the output with the size $\overline{T_1}$ with zeros or with padding information obtained from the output with the size $\overline{T_1}$, wherein the padding information obtained from the output with the size $\overline{T_1}$ is applied as redundant padding information to increase the size $\overline{T_1}$ of the output to the size $\widehat{T_1}$ and wherein the padding comprises reflection padding or repetition padding.

15. The method according to any of claims 1 to 14, wherein the NN comprises, in the processing order of the bitstream through the NN, a further unit that applies a transformation to the input that does not change the size of the input in the at least one dimension, wherein the method comprises applying the rescaling after the processing of the input by the further unit and before processing the input by the following layer, if the rescaling results in an increase of the size of the input in the at least one dimension, and/or wherein the method comprises applying the rescaling before the processing of the input by the further unit, if the rescaling comprises a decrease of the size of the input in the at least one dimension, wherein the further unit is or comprises a batch normalizer and/or a rectified linear unit, ReLU.

16. An encoder (2100) for encoding a picture, wherein the encoder comprises one or more processors for implementing a neural network, NN, wherein the one or more processors are adapted to perform a method according to any of claims 1 to 8 or 15 in combination with any of claims 1 to 8.

17. A decoder (2200) for decoding a bitstream representing a picture, wherein the decoder comprises one or more processors for implementing a neural network, NN, wherein the one or more processors are adapted to perform a method according to any of claims 9 to 14 or 15 in combination with any of claims 9 to 14.